Characterisation of the Murine Gene and the Porcine cDNA for Lung Surfactant Protein D and Genetic Mapping of Murine Complement Components using a Novel Approach

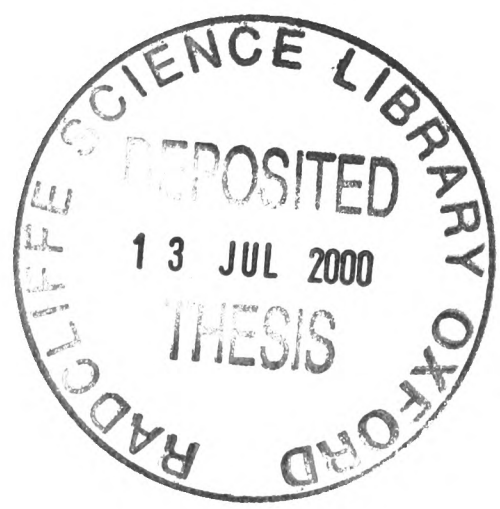
A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

By

Peter R. Lawson

New College, Oxford.
Hilary 2000

MRC Immunochemistry Unit,
Department of Biochemistry,
University of Oxford



ACKNOWLEDGEMENTS

This thesis is dedicated to my loving parents, sister and my wonderful girlfriend, Karen Browne.

Firstly, I'd like to thank my supervisor Ken Reid for his support, guidance, generosity and relaxing nature that allowed me to conduct and design my own research plans within his laboratory. For this I am eternally grateful.

I want to thank Viv Perkins and Nick Kent, who are soon to be married and wish them the happiest life together. Without them and my girlfriend Karen, I would never have such a beautiful thesis. Their proof reading and help with the figures has once again make me realise what wonderful friends I have made in Oxford. I owe them tremendous thanks, praise and admiration.

I would also like to thank my other dear friends I have made within the unit that have had the arduous task of proof reading. Martin van Eijk for his keen eyes, wonderful friendship and hospitality in the Netherlands with Karen. Christian Specht for the days and nights in the laboratory, our quick dashes for last orders and terrible pranks that we should be grateful for being alive. Jon Higgins for his help, guidance, friendship and quick proof reading. Alvaro Diaz for superb friendship, conversations and a wonderful time in sharing a house. Alister Dodds for his constant nagging to finish papers and the thesis plus his wealth of knowledge that has both informed and entertained me.

Other people who either proof read papers or the thesis. Bob Sim for his enlightening advice and refining my writing. Alex Law for cutting remarks and precise criticism. Henk Haagsman for his enthusiasm on the collaborative work, his hospitable nature and of course introducing me to Martin van Eijk.

People who made the laboratory an exciting and wonderfully open, richly academic and supportive environment. Jin Hua Lu for his kindness and teaching. Kazuo Miyamura who I hold in high esteem and is a truly wonderful gentleman. Uffe Holmskov for his guidance and later the opportunity to work within his laboratory in Denmark - with him and the wonderful people at the department in Odense that made those four months an enjoyable part of my life. Choi Seong-Hyun who is an extra-ordinary kind and thoughtful man for his truly amazing friendship and entertaining stories. Hans-Jürgen Hoppe for above all, his mind bending advice. Leonora Leigh for encouragement, beaming smiles and friendship. Tony Willis for cracking advice, care and kindness.

Of the people not connected with my work I'd like to thank are Kitty Hauser for listening to my woes and warm friendship, Neil Hanlon for his wit and introducing me to some amazing friends; Marc Deller, Paul Saleh, Deepan Shah and Kath and Chris Bower for all keeping me sane through holidays, nights out and passing out.

Without Karen I would never have finished this thesis; calming me, soothing me and loving me. She encouraged and drove me on through the worse patches by flattering me with her charming words and grace. I cannot believe how lucky I am to meet and fall in love with such an amazing woman. I would also like to thank her parents, brother and sisters for their loving generosity, support, plus tolerating my writing during the Christmas and new year period.

I want to thank my parents and my beautiful sister, Sarah, for their enduring love and support, in addition to their immense pride, encouragement and faith in my success. They are truly amazing people and I hope I can do them justice and return their love.

Abbreviations

Nomenclature of the complement components is that recommended by the World Health Organisation, (1986). Nucleotides containing the bases adenine, cytosine, guanine and thymine are abbreviated to A, C, G and T respectively. One and three letter codes are used to refer to amino acids as recommended by the IUPAC-IUB Commission on Biochemical Nomenclature, (1969). The nomenclature used for restriction enzymes is that of Smith and Nathans (1973). Other abbreviations are defined in the text on their first occurrence or else are listed below.

ARDS	acute respiratory distress syndrome
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
C1	complement component 1
C1q, r or s	complement component 1, subcomponent q, r or s
C3aR	complement component 3a receptor
cd	cluster determination
cDNA	complementary DNA
Cdx	transcription factor binding site
CGN1	bovine conglutinin gene
CL-43	collectin 43
cM	centimorgan
CRD	carbohydrate recognition region
CUB	C1r/C1s/sea urchin uefg/bone morphogenic protein domain
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanidine triphosphate
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DPPC	dipalmitoylphosphatidylcholine
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetra-acetic acid, disodium salt
ELISA	enzyme-linked immunosorbent assay
EST	expressed sequence tag
EUCIB	European collaborative interspecific backcross
FISH	flourescent *in situ* hybridisation

Fkh	forkhead family
Ig	Immunoglobulin
kb	kilobase pair
kDa	kilo dalton
LINE	long interspersed repetitive element
LPS	lipopolysaccharide
MASP	mannan binding lectin assosciated serine protease
MBL	mannan binding lectin
mRNA	messenger ribonucleic acid
PACS	P1 artificial chromosome
PC	phosphatidylcholine
PCR	polymerase chain reaction
PE	phosphatidylethanolamine
pfu	plaque forming units
PG	phosphatidylglycerol
PI	phosphatidylinositol
PKC	protein kinase C
PS	phosphatidylserine
RACE	5' rapid amplification of cDNA ends
RaRF	Ra-reactive factor
RDS	respiratory distress syndrome
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RT-PCR	reverse transcriptase polymerase chain reaction
SDS	sodium dodecyl sulphate
Sftpa	murine gene for SP-A
Sftpd	murine gene for SP-D
SINE	short interspersed repetitive element
SP	surfactant protein
SSLP	simple sequence length polymorphism
STS	sequence tag site
TTF-1	thyroid transcription factor 1
U	units
UTR	untranslated region
YAC	yeast artificial chromosome

ABSTRACT

The work in this thesis describes the isolation and characterisation of the mouse surfactant protein D (SP-D) gene (*Sftpd*). This gene has been show to span 8 exons across 14 kb of DNA sequence and displays an overall organisation similar to the other collectin genes. The complete 5' untranslated region of the mRNA was also cloned which allowed the identification of the transcription start site for SP-D. Subsequently, the analysis of the promoter region of *Sftpd* revealed positional conservation of a number of transcription factor binding sites between the murine and human SP-D genes and the bovine conglutinin gene. A single copy SP-D-like gene has been shown to be present in mammals, birds and amphibians. Three polymorphic microsatellites were identified for the SP-D gene. In addition, the murine genes for surfactant protein A, SP-D and mannose binding lectin (MBL-A) were shown to co-segregate within a 5.64 cM area on chromosome 14. PCR cloning identified the presence of a CL-43-like gene in sheep, this represent the second member of the bovidea to contain CL-43. Specific antibodies were raised against mouse SP-A and SP-D by cDNA cloning fragments of these molecules and using them for protein expression.

The complete cDNA sequence of the porcine cDNA was cloned. Three unique features were revealed in comparison to SP-D from other species. The collagen region contains an extra cysteine residue, which may have important structural consequences. Two other differences lie within loop regions of the carbohydrate recognition domain, a potential glycosylation site and an insertion of three amino acids, which may have functional implications by influencing carbohydrate binding. The porcine genes for SP-D and SP-A were also shown to co-localise to porcine chromosome 14, which is syntenic with murine chromosome 14.

A novel and rapid PCR restriction fragment length polymorphism method was developed to exploit the murine expressed sequence tag (EST) database. This technique circumvents the laborious cDNA or genomic cloning steps of other mapping methods by relying on EST data and the prediction of exon-intron boundaries. This approach revealed that the C3a receptor, C1r and C1s genes form an unexpected complement gene cluster while the second MBL associated serine protease gene does not co-segregate with other complement activating serine protease genes. The gene localisations performed here help to explain the evolution of this group of proteins. This method can easily be applied to the genes of other systems, ranging from the interests of the individual researcher to large scale gene localisation projects.

TABLE OF CONTENTS

Acknowledgements
Abbreviations
Abstact
Table of Contents

CHAPTER 1: INTRODUCTION 1
1.1. Lung Surfactants 1
1.2. The Surfactant Lipids 1
1.2.1. The Lipid Composition of Surfactant 1
1.3. The Proteins of Lung Surfactant 3
1.3.1. The Hydrophobic Surfactant Protein 3
1.3.1.1. Surfactant Protein B 4
1.3.1.2. Surfactant Protein C 6
1.3.2. The Surfactant Collectins 7
1.3.2.1. The Primary Structure of the Collectins 8
1.3.2.2. Higher Order Oligomers 9
1.3.2.3. Collagen Kink 10
1.3.2.4. The CRD: Structure and Function 10
1.3.2.5. The Neck Region and Carbohydrate Binding 11
1.3.2.6. The Gene Organisation 12
1.3.2.7. The Expression of the Surfactant Collectins 15
1.3.2.8. Roles of the Surfactant Collectins Proteins 16
1.3.2.9. Effects Mediated by SP-A and SP-D Binding 19
1.3.2.10. Functional Sites in SP-A and SP-D 24
1.3.2.11. Levels of Collectins Within the Human Population 26
1.3.2.12. Association of Alleles with Disease and Protein Levels 27
1.3.2.13. Varying Levels of the Surfactant Collectins in Disease 28
1.3.2.14. Transgenic Animals Lacking the Collectins 29
1.4. The Complement System 32
1.4.1. The Classical Pathway 33
1.4.2. The MBL activating pathway 34
1.4.2.1. Activators of the MBL Activation Pathway of Complement 36
1.4.2.2. The Modular Organisation of the MASPs 36
1.4.2.3. The MBL-MASP complex 36
1.4.2.4. The Chromosomal Localisation of the MASPs, C1r and Cls 37
1.4.2.5. The Activities of the MASPs 37
1.4.2.6. Evolution and Protein Families 39
1.4.3. Anaphylatoxin Receptors 40
1.4.4. Gene Clusters 40
1.5. Gene Mapping 41
1.5.1. Chromosomal Mapping 41
1.5.2. Physical Mapping 42
1.5.3. The Genetic Approach for Gene Mapping 42
1.5.4. The Use of ESTs in Gene Mapping 43
1.6. Aims of the project 43

CHAPTER 2: MATERIALS AND METHODS **46**
2.1. Materials 46
2.1.1. Enzyme and Chemicals Suppliers 46
2.1.2. Buffers and Solutions 46
2.1.3. Media and Antibiotics 47
2.1.4. Bacterial Strains 48
2.1.5. Vectors 48
2.2. Recombinant DNA and Cloning Methods 48
2.2.1. Bacterial Manipulation and Storage 48

2.2.1.1. Growth of Bacteria in Liquid Media 48
2.2.1.2. Growth of Bacteria on Solid Media 49
2.2.1.3. Storage of Bacterial Stocks 49
2.2.2. Plasmid DNA Preparation 49
2.2.2.1. Minipreps of Plasmid DNA 49
2.2.2.2. Large Scale Preparation of Plasmid DNA 50
2.2.3. Introducing Plasmid DNA into *E. coli* 51
2.2.3.1. Preparation of Competent Cells for Transformation 51
2.2.3.2. Transforming Competent *E. coli* with Plasmid DNA 52
2.2.4. Preparation of Lambda Bacteriophage DNA 52
2.2.4.1. Making Bacteriophage Lysates 52
2.2.4.2 . Plate Lysates 52
2.2.4.3. Liquid Lysates 52
2.2.4.4. Storage of Phage Lysates 53
2.2.4.5. Checking Phage Titre 53
2.2.4.6. Preparation of Lambda DNA from Phage Lysates 54
2.2.5. Extraction and Concentration of DNA 55
2.2.5.1. Organic Extraction of Protein from Nucleic Acid: Phenol Extraction 55
2.2.5.2. Ethanol Precipitation 55
2.2.5.3. Preparation of Genomic DNA from Mammalian Tissues 55
2.2.5.4. Isolation and Purification of DNA from Agarose Gels 56
2.2.6. Analysis of DNA 56
2.2.6.1. Agarose Gel Electrophoresis 56
2.2.6.2. High-Resolution Denaturing Gel Electrophoresis 57
2.2.7. Southern Blotting 58
2.2.7.1. Blotting DNA Fragments onto Nylon Supports 58
2.2.7.2. Hybridisation of a DNA Blot with a Radiolabelled Probe 58
2.2.7.3. Removal of Probes from the Hybridisation Membrane 59
2.2.8. Measuring DNA Concentration 59
2.2.8.1. Optical Density 59
2.2.8.2. Rapid DNA Concentration Estimation Using Small Volumes 60
2.2.9. Enzymatic Manipulation of DNA 60
2.2.9.1. Digestion of DNA with a Single Endonuclease 60
2.2.9.2. Digestion of DNA with Multiple Endonucleases 60
2.2.9.3. Restriction Mapping with Multiple Endonucleases 61
2.3. DNA Modification and Labelling 61
2.3.1. Labelling of DNA by Random Oligonucleotide-Primed Synthesis 61
2.3.2. Terminal Deoxynucleotidyltransferases 61
2.3.3. Alkaline Phosphatase 62
2.3.4. Ligation with T4 DNA Ligases 62
2.4. Construction of Hybrid DNA Molecules 63
2.5. Preparation and Analysis of RNA 63
2.5.1. Purification of Total RNA from Tissue 63
2.5.2. Reverse Transcriptase for Single-Stranded DNA Synthesis 64
2.6. Screening of Recombinant DNA Libraries 64
2.6.1. Plating and Transferring Bacteriophage Libraries to Filters 64
2.6.2. Purification of Bacteriophage Clones 65
2.7. DNA Sequencing 65
2.7.1. Manual Sequencing by Dideoxy Method 65
2.7.2. Automated Sequencing 66
2.8. Polymerase Chain Reaction (PCR) 66
2.8.1. Enzymatic Amplification of DNA by PCR 66
2.8.2. Preparing PCR Products for Direct Sequencing 68
2.8.3. Anchor PCR 68
2.8.4. Cloning of PCR Products by TA Overhangs 69
2.8.5. Reverse Transcriptase PCR 69
2.8.6. PCR Screening for Positive Colonies 70
2.9. Bioinformatics 70

2.9.1. Submission of Sequences to Databases 70
2.9.2. Submitting Sequences for Homology Searches using BLAST 71
2.9.3. Sequence Analysis 71
2.10. Protein Methods 71
2.10.1. Electrophoretic Separation of Proteins 71
2.10.1.1. Detection of Protein 72
2.10.1.2. Coomassie Blue Staining 72
2.10.1.3. Silver Staining 73
2.10.2. Protein Blotting and Detection 73
2.10.2.1. Transfer of Protein to Nitrocellulose Support 73
2.10.2.2. Qualitating Protein with Ponceau S 74
2.10.2.3. Dotblot 74
2.10.3. Antibody Methods 75
2.10.3.1. Preparation of Polyclonal Antisera 75
2.10.3.2. Preparation of Antisera from Blood 75
2.10.3.3. Indirect ELISA to Detect Specific Antibodies 75
2.10.3.4. Probing with Antibody 76

CHAPTER 3: THE MOUSE SP-D GENE 78
3.1. Introduction and Aims 78
3.2. Generating a Probe 80
3.3. Cloning and Characterisation of the Murine Gene for SP-D 81
3.4. Sequencing the Murine Gene for SP-D 82
3.5. Determining the 5' Extremity of the Mouse SP-D Gene 83
3.6. The Presence of the SP-D Gene in Other Species 85
3.6.1. Zooblot 85
3.6.2. Low stringency PCR on Different Species and Direct Sequencing 86
3.7. Dispersed Repetitive Elements in the Murine SP-D Gene 87
3.8. Polymorphisms in the Murine SP-D gene 88
3.8.1. Detection of polymorphisms by PCR 88
3.8.2. The CA Repeats 89
3.8.3. The Polyadenosine Repeat 90
3.8.4. The Implications of Repeats 90
3.9. Chromosomal Localisation Of Murine Gene for SP-D 90
3.10. Genetic Mapping of The Collectin Locus 91
3.11. Analysis of the promoter sequence of murine SP-D 94
3.11.1. The Promoter Region Contains LINEs and SINEs 94
3.11.2. Conservation in the Promoter Region 94
3.11.3. Transcription Factor Binding Sites 95
3.12. Generation of Specific Antibodies for Mouse SP-A and SP-D 97
3.12.1. Cloning of the cDNA Fragment for the Neck-CRD of Murine SP-A 97
3.12.2. Cloning of the cDNA Fragment for the Neck-CRD of murine SP-D 97
3.12.3. Confirming the Sequence of Murine SP-A 98
3.12.4. Murine SP-A and SP-D Used to Generate Antibodies 98
3.12.5. Raising Polyclonal Antibodies to Murine SP-A and SP-D 99
3.13. Discussion 100
3.13.1. The Mouse SP-D Gene Sequence 100
3.13.1.1. The Exons of the Mouse SP-D Gene 101
3.13.1.2. The Collagen Encoding Exons 102
3.13.2. Repeats 103
3.13.2.1. Microsatellites 103
3.13.2.1.1. Significance of the Repeats 103
3.13.2.2. Minisatellites and Repetitive Elements 103
3.13.3. Genetic Mapping of the Collectins 104
3.13.4. The Promoter Region 104
3.13.4.1. Homology Within the Promoter Region 105
3.13.4.2. Conservation of Transcription Factor Binding Sites 105
3.13.4.3. Lung-Specific Transcription Factor Binding Sites 105

3.13.5. Collectins in Other Species 106
3.13.5.1. SP-D-like Genes in Other Species 107
3.13.5.2. Sheep CL-43 107

CHAPTER 4: CHARACTERISATION OF PORCINE SP-D 109
4.1. Introduction and Aims 109
4.2. Construction of the Probe for Screening the Porcine Lung cDNA Library 111
4.3. Identifying the Lambda Bacteriophage Clones 111
4.4. PCR Sequencing of the Positive Clones 112
4.5. Characterisation of the Porcine SP-D cDNA Clones 112
4.5.1. Analysis of the 3' Clone for Porcine SP-D 113
4.5.2. Analysis of the 5' Clone for Porcine SP-D 113
4.6. 5' RACE to Clone the 5' Region of Porcine SP-D cDNA 113
4.7. Sequence Analysis of Porcine SP-D 115
4.7.1. Sequence Analysis of the Untranslated Region 115
4.7.2. Analysis of the Predicted Protein Sequence of Porcine SP-D 115
4.8. The Major Sequence Variations Seen in Porcine SP-D 116
4.9. Confirmation of the Major Sequence Variations in Porcine SP-D 117
4.9.1. Amplification of the Genomic Regions Containing Sequence Variation 117
4.9.2. Analysing the Different Pig Breeds For The Sequence Variations 118
4.10. Confirming the Sequence of the Other Porcine Surfactant Collectin: SP-A 118
4.11. Chromosomal Localisation of Porcine SP-A and SP-D 119
4.12. Discussion 120
4.12.1. The Cloning of Porcine SP-D cDNA 120
4.12.2. The Differences in SP-D from Porcine compared to other Species 122
4.12.2.1. The Extra Cysteine 122
4.12.2.2. The CRD Insertion 123
4.12.2.3. The Additional N-linked Glycosylation Site 124
4.12.3. The Porcine Surfactant Collectin Cluster 124
4.12.4. Porcine Surfactant Protein A 125
4.13. Perspective 126

CHAPTER 5: A NOVEL APPROACH TO GENETIC MAPPING: MURINE COMPLEMENT 128
5.1. Introduction and Aims 128
5.1.1. Testing the Approach: Genetic Mapping of Murine Complement Genes 129
5.2. Identification of the Murine Complement Homologues 130
5.3. Primer Design 131
5.4. Amplification of the Murine Complement Gene 132
5.5. Sequence Confirmation of the PCR Products used for Localisation 132
5.6. Screening for Restriction Fragment Length Polymorphisms 133
5.7. The Intron-Exon Boundary of the Serine Protease Domain of MASP2 135
5.8. Chromosomal Localisation 136
5.8.1. Chromosomal Localisation of C1r and C1s 137
5.8.2. Chromosomal Localisation of C3aR 137
5.8.3. Chromosomal Localisation of murine *Masp2* 138
5.9. Discussion 138
5.9.1. Development of a Novel Approach to Genetic Mapping 138
5.9.2. Discussion of the Data 139
5.9.2.1. The C1r, C1s and C3aR Gene Cluster 139
5.9.2.2. The Localisation of C3aR 140
5.9.2.3. Genetic Mapping of *Masp2* 141
5.9.2.4. The Non-Homologous Functions Between the MASPs and C1r/s 141
5.9.2.5. The Evolutionary Theory from the Protein and cDNA Level 142
5.9.2.6. The Evolutionary Theory from the Genetic Viewpoint. 143
5.10. Summary 145

Chapter 6: Summary 146

References 152

Appendices 1-9

Publication List

CHAPTER 1: INTRODUCTION

1.1. Lung Surfactant

Pulmonary surfactant is an essential constituent of lung physiology, comprising of a complex mixture of lipids and proteins that coat the alveoli. Surfactant has two roles within the lung. Firstly, as a true surfactant, it lowers the surface tension at the air-liquid interface of terminal alveoli to prevent alveolar collapse, important for normal respiratory function (Rooney *et al.*, 1994; Johansson and Curstedt, 1997; Creuwels *et al.*, 1997). Secondly, surfactant plays a role in local innate immune defence. This function is imparted by both lipid and protein constituents of surfactant (Wright, 1997; Lawson and Reid, 2000).

1.2. The Surfactant Lipids

Although proteins act in co-operation with lipids in surfactant, it is ultimately the latter that are responsible for the surface-tension lowering properties during respiration. However, there are other roles that surfactant lipids are involved in; the metabolic aspects of surfactant life cycle, assembly of the lamellar bodies and the release of lipids from storage structures (Figure 1.1).

1.2.1. The Lipid Composition of Surfactant

The major surfactant lipids are the phospholipids; phosphatidylcholine (PC), phosphatidylglycerol (PG) and phosphatidylethanolamine (PE) accounting for greater than 95% of the total extracellular surfactant (Goerke, 1998).

PC imparts the major contribution to the surface-tension reducing capacity of surfactant at the air-liquid interface. The lipid composition of extracellular surfactant is 80% PC, 50% of which is in the form of dipalmitoylphosphatidylcholine (DPPC) (Possmayer *et al.*, 1984). This contrasts with only 10-20% of the PC content of erythrocyte membranes being DPPC (Veldhuizen *et al.*, 1998). DPPC is the only surfactant lipid

capable of creating low surface tensions and maintaining them over several hours without exogenous influence. DPPC also plays a role in surfactant film respreading during respiration cycles (Schurch, 1982).

The acidic phospholipids, phosphatidylglycerol and phosphatidylinositol (PI), account for 8-15% of the total surfactant. The precise roles of these components are unknown, although they are clearly important as their ratios change with age and in lung diseases such as acute respiratory distress syndrome (ARDS).

The remaining 3-5% of phospholipids are PE and phosphatidylserine (PS). The role of these minor phospholipid components in the surfactant is unclear. DPPC has the most influential role in surface active role of surfactant, therefore other explanations have been sought for the function of the minor phospholipids. One possibility is that the minor phospholipids are vestigial from the lysosomal origins of the lamellar bodies: the intracellular storage structure for lipids in type II alveolar cells (Figure 1.1).

There is far less information about the neutral lipids of the surfactant, of which cholesterol is the major component (80-90%). Cholesterol has been shown to increase the absorption of DPPC into surfactant by increasing the fluidity of the bilayer through changes in lipid packing. Conversely, increasing the fluidity of the surfactant also increases the surface tension. However, one of the hydrophilic surfactant proteins, surfactant protein A (SP-A, see section 1.3.2), binds cholesterol (Yu and Possmayer, 1996) limiting the access of cholesterol to the air-liquid interface (Yu and Possmayer, 1998). A role of cholesterol in surfactant function is further suggested by the rapid alterations in the amount of cholesterol during exercise induced hyperventilation (Orgeig *et al.*, 1995).

The absorption of other phospholipids into a DPPC bilayer is poor. Two hydrophobic surfactant proteins (surfactant proteins B and C, see section 1.3.1) actively promote the

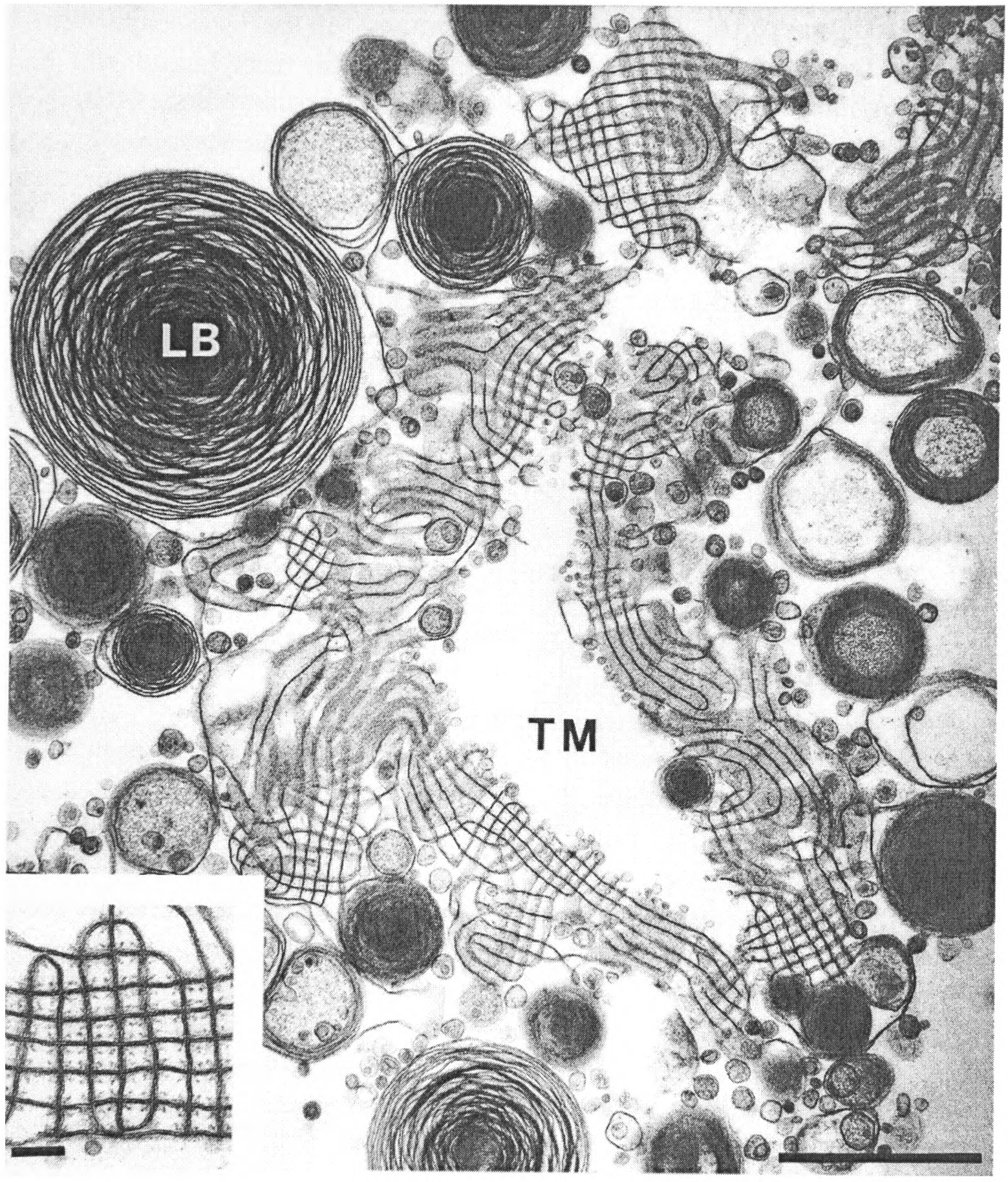


Figure 1.1. Particles in the alveolar subphase. An electron micrograph of a rat lung shows lamellar bodies (LB) forming tubular myelin (TM) (bar = 1.0 μm). Remaining vesicular structures may represent both used and rejected surfactant materials. The inset shows detail of tubular myelin, with small projections at the corners thought to represent SP-A (bar = 0.1 μm). From Goerke *et al.*, 1998.

absorption of other surfactant lipids into surfactant bilayers of pure DPPC (Wang *et al.*, 1996).

Extracted porcine or bovine surfactant, which is composed mainly of lipids and trace amounts of the two hydrophobic surfactant proteins, is used clinically as a successful replacement therapy for many pulmonary disorders including respiratory distress syndrome (RDS), meconium aspiration syndrome, pneumonia, pulmonary haemorrhage, and lung hypoplasia (Spragg, 1998; Sun, 1996). The recent use of native surfactant and extracted surfactant reconstituted with other surfactant proteins indicates that at least one of the more hydrophilic surfactant proteins (SP-A) may also play a beneficial role in surfactant replacement therapy (Sun *et al.*, 1997).

1.3. The Proteins of Lung Surfactant

Four proteins are classically thought to be associated with pulmonary surfactant. Two hydrophobic proteins, surfactant protein B (SP-B) and surfactant protein C (SP-C), are engaged in the organisation and rearrangement of surfactant lipid which influences the surface tension lowering activity of surfactant. The other two proteins are more hydrophilic; SP-A and surfactant protein D (SP-D) are considered to be involved in non-clonal innate immunity and lipid homeostasis in the lung. SP-A can also speed up the process of absorbing lipids into the air-liquid interface.

1.3.1. The Hydrophobic Surfactant Proteins

The hydrophobic surfactant proteins were first identified in the organic extractions of lamellar bodies. These proteins were subsequently termed SP-B and SP-C. The cDNA cloning of these proteins helped reveal their polypeptide structure (Hawgood *et al.*, 1987; Glasser *et al.*, 1987; Jacobs *et al.*, 1987; Warr *et al.*, 1987; Glasser *et al.*, 1988).

1.3.1.1. Surfactant Protein B

A. The Structure of SP-B

The mature protein of SP-B is a homodimer of a polypeptide of approximately 8.7 kDa or 79 amino acids (Hawgood *et al.*, 1987) linked by interchain disulphide bonding (Figure 1.2). The structure of SP-B has been proposed as five amphipathic helices oriented pairwise in an anti-parallel motif (Andersson *et al.*, 1991). SP-B belongs to the saposin family, a family that includes the lysosomal saposins A-D and whose members all are cleaved from a common precursor proprotein. The saposin family shares the same structural fold as SP-B but their functions vary from glycoceramide binding to anti-microbial activity. Other members of the family are the T cell cytolytic protein, NK-lysin and the pore-forming proteins from *Entamoeba histolytica* (Patthy, 1991; Liepinsh *et al.*, 1997)

B. The Function of SP-B

SP-B is involved in the packaging and construction of lipid storage structures. Lipid vesicles of the multivesicular body in type II cells are broken down into small bilayer discs surrounded by a rim of SP-B, in a bilayered wheel configuration. These discs fuse with the aid of SP-B to form highly ordered multilayered stacks of the lamellar body (Williams *et al.*, 1991). The lamellar body is an intracellular compartmentalisation of lipid that is the precursor to an extracellular lipid storage structure in the surfactant, tubular myelin (Figure 1.1). SP-B is essential in the formation of these intracellular lipid storage structure as mice deficient in SP-B lack lamellar bodies (Clark *et al.*, 1995). After excretion from type II cells, the lamellar body forms tubular myelin: a complex three-dimensional grid-like structure. Both SP-B and SP-A are necessary in forming this structure. The biological significance of tubular myelin has been called into question after mice lacking SP-A fail to form tubular myelin structures but have normal surfactant function (Korfhagen *et al.*, 1996; Ikegami *et al.*, 1997; Ikegami *et al.*, 1998) (see section 1.3.2.12).

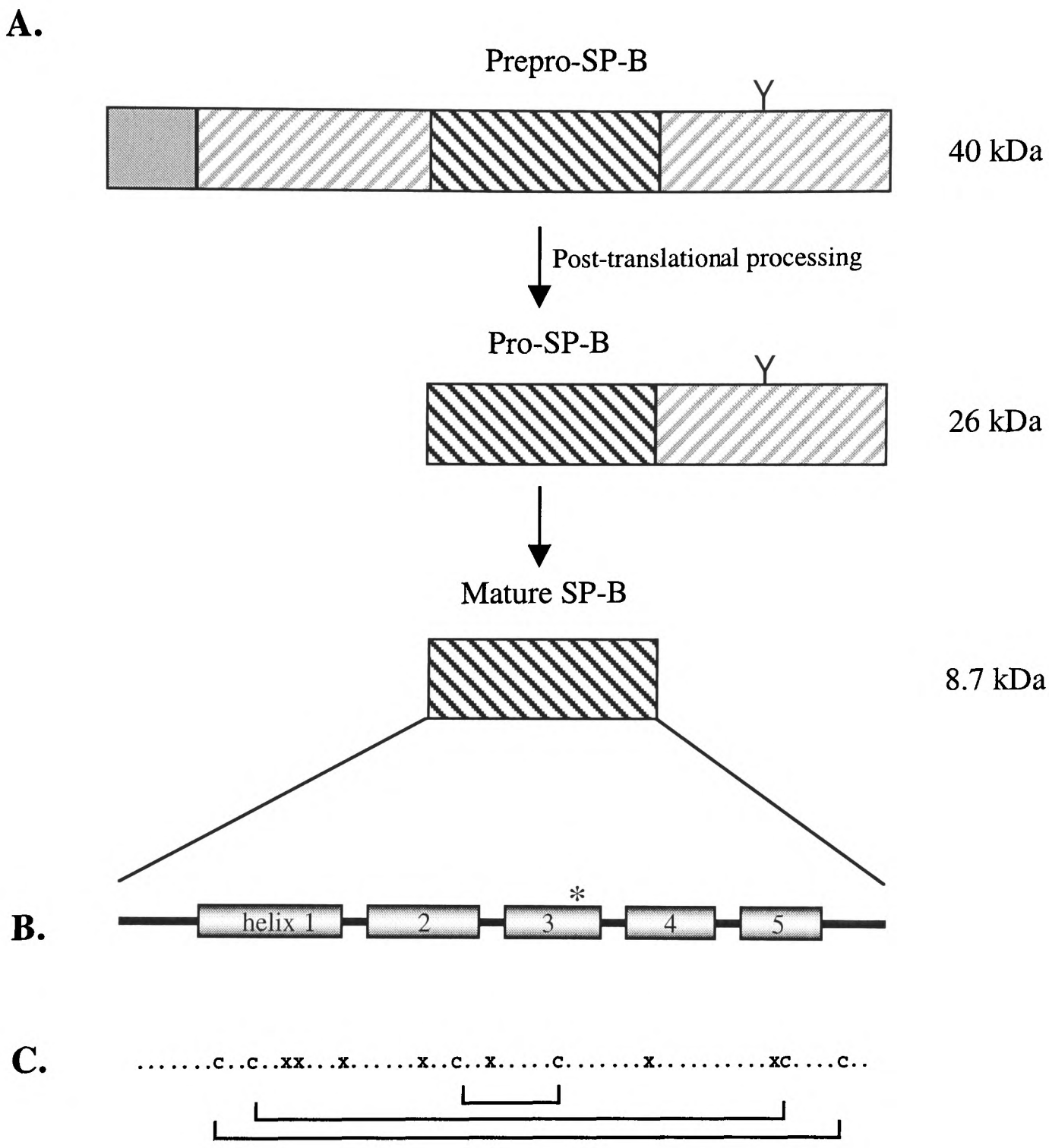


Figure 1.2. Schematic representation of the structure of preproprotein, proprotein and mature SP-B. **(A)** The preproprotein of SP-B (residues 1-381) comprises a signal peptide (shaded box) and three saposin-like domains (hatched boxes). The propeptide comprises two domains and the mature protein one saposin-like domain. **(B)** The mature form of SP-B (residues 201-279) is predicted to comprise five helices placed as the NK-lysin structure. **(C)** The consensus sequence of the saposin motif showing the six conserved cysteines, C, and seven hydrophobic residues, X. Intramolecular disulphide bonds are represented by brackets, the intermolecular disulphide is marked by an asterisk, *, and the glycosylation site indicated by a Y .

C. Translation of SP-B

Following expression, SP-B exists as a glycosylated prepro-protein of approximately 40 kDa (see Figure 1.2) (Weaver and Whitsett, 1989; Weaver *et al.*, 1992). There are 22 cysteines in the human prepro-protein, 10 forming intrachain disulphides and two free cysteines. SP-B prepro-protein is a monomer in type II cells and only forms the single interchain disulphide required for SP-B dimerisation after cleavage to the propeptide. The more hydrophilic flanking sequence is believed to function as a molecular chaperone, packaging the extremely hydrophobic mature protein and possibly protecting intracellular membranes from the lytic activity of SP-B before integration into lamellar bodies. Mature human SP-B contains three intrachain disulphide bonds and one interchain disulphide bond.

D. Transgenic Mice

Transgenic mice that lack SP-B (-/-) suffer a lethal respiratory distress soon after birth with an aberrant processing of the SP-C precursor (Clark *et al.*, 1995). SP-B transgenic mice can be rescued with administration of the proprotein of SP-B or ventilation (Akinbi *et al.*, 1997). Heterozygous mice, SP-B (+/-), have an intermediate phenotype, suffering from reduced lung compliance, air trapping and airway collapse (Clark *et al.*, 1997). In both homozygous and heterozygous transgenic mice the major phospholipid compositions of surfactant are unaffected.

E. Expression of SP-B

SP-B is expressed in alveolar type II cells and nonciliated bronchiolar epithelial (Clara) cells in the lung. However, SP-B expressed in Clara cells does not participate in surfactant function. Transgenic mice that only express SP-B in Clara cells, lacking type II cell mediated expression, suffer the same lethal neonatal respiratory distress as mice that lack SP-B (Lin *et al.*, 1999; Clark *et al.*, 1995). The function of SP-B expressed in Clara cells is unknown. SP-B, as well as SP-A and SP-D, have all also been detected in the gastrointestinal tract (Eliakim *et al.*, 1989; Fisher and Mason, 1995).

F. Disease and Alleles associated with SP-B

There have been a number of genetic variants described for SP-B (Nogee *et al.*, 1994; Nogee *et al.*, 1993; Nogee, 1998). One genetic variant of SP-B has been shown to be associated with RDS in infants, but not in adults (Floros *et al.*, 1995). Newborn RDS had always been associated with lung immaturity, but these results perhaps indicate that one route to the disease is due to SP-B deficiency. A deficiency in SP-B was identified as a cause of congenital alveolar proteinosis. The deficiency arose from two frameshift mutations in the SP-B mRNA which leads to a truncated prepro-protein and the absence of SP-B in surfactant. Homozygosity for these mutations are fatal, while heterozygotes are asymptomatic for an alveolar proteinosis-like disease (Nogee, 1998).

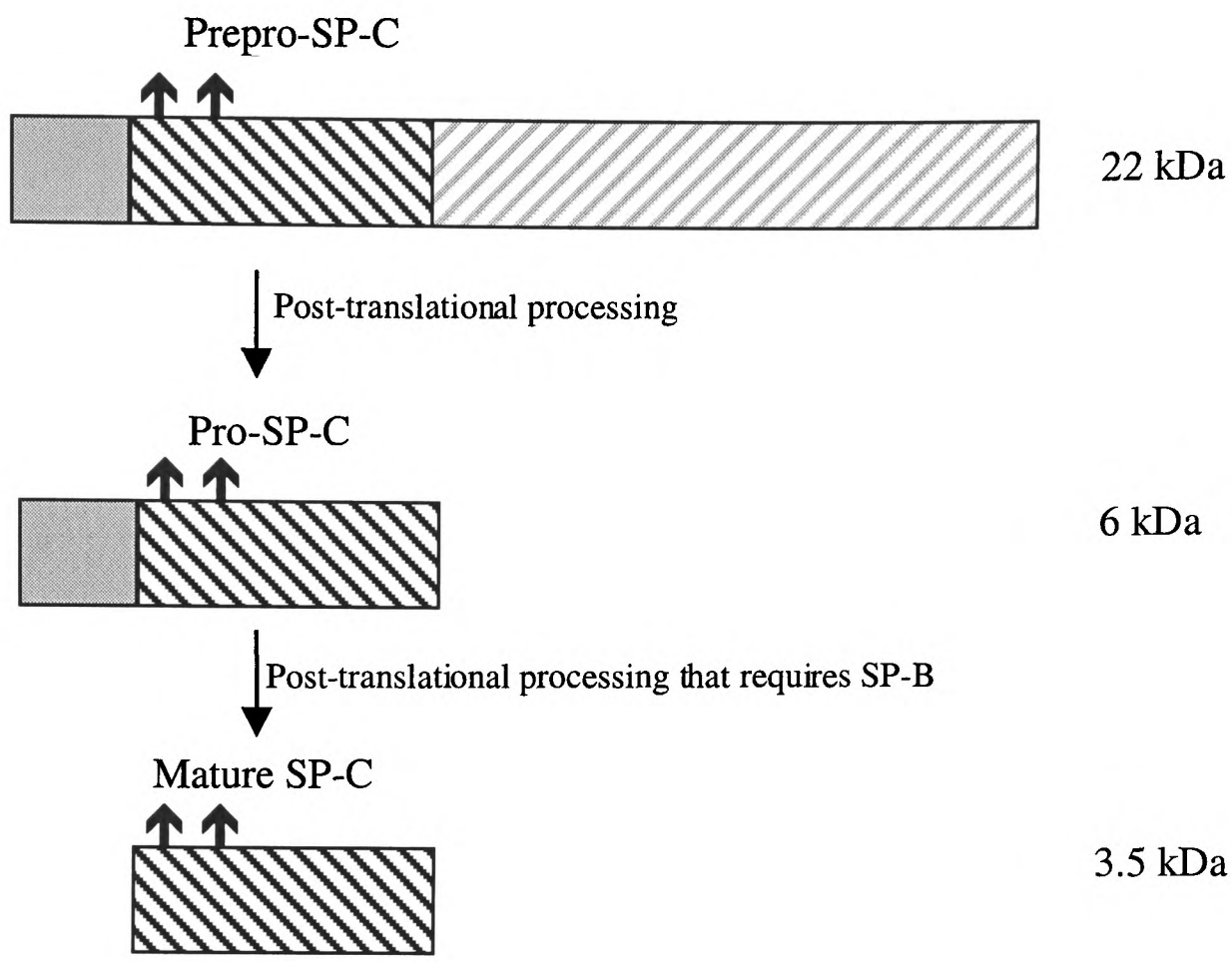
1.3.1.2. Surfactant Protein C

A. The Structure of SP-C

The 35 amino acid polypeptide chain of SP-C contains a unique posttranslational modification: covalently-linked palmitoyl chains (two in most species) via thioester bond formation with cysteine residues (see Figure 1.3). The polypeptide chain is extremely hydrophobic and is folded into a 3.7 nm alpha-helical membrane-spanning region. By virtue of the hydrophobic alpha-helical region, SP-C can interact with lipid bilayers composed of dipalmitoylphosphatidylglycerol, a negatively charged lipid constituent of surfactant, in a transmembrane fashion (Vandenbussche *et al.*, 1992). SP-C also integrates into lipid monolayers in a tilting fashion, depending upon the presence of palmitoyl chains (Creuwels *et al.*, 1993). Like SP-B, SP-C is derived from a prepro-protein. Neither mature SP-C nor the prepro-protein share significant sequence similarity with other known protein or cDNA sequences.

SP-C exists as mixture of isoforms that involve covalent changes to the polypeptide chain; absent palmitoyl chains, amino-terminal truncations, methionine oxidation,

A.



B.

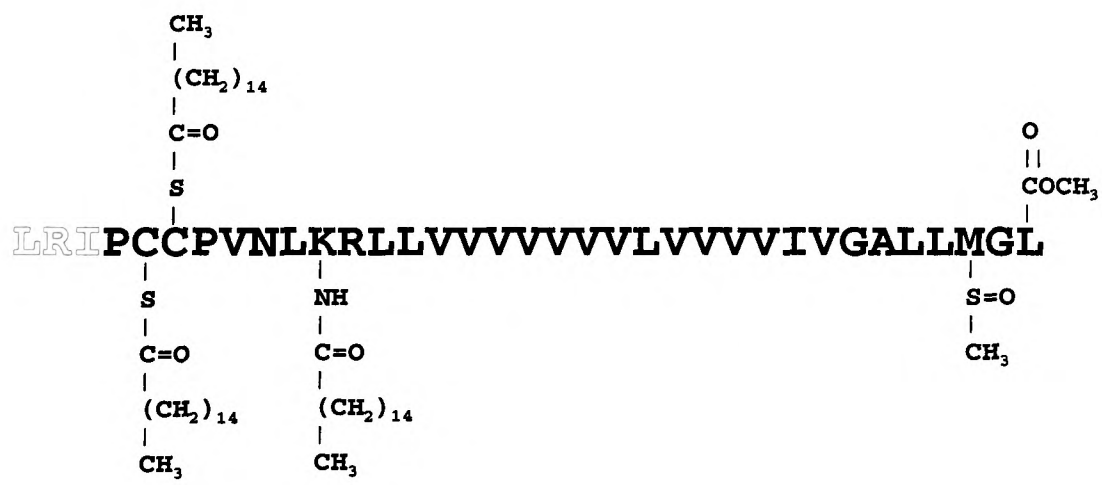


Figure 1.3. Schematic representation of the structure of preproprotein, proprotein and mature SP-C. **(A)** The preproprotein of SP-C is processed to the mature peptide only in the presence of SP-B. Palmitic acid modifications of SP-C are represented by arrows, ↑ . **(B)** The primary structure of SP-C, here the porcine sequence, in one-letter code. Known modifications of SP-C are represented; the N-terminal truncations (grey), the thioester-linked palmitoyl groups of C-5 and C-6, the amide-linked palmitoyl group of K-11, the sulphoxide of M-33 and the C-terminal methyl-ester.

esterification of the carboxyl-terminus and an extra palmitoyl chain covalently linked to a species conserved lysine residue (Figure 1.3).

B. The Expression and Function of SP-C

To date, SP-C is the only surfactant protein that has been found exclusively in the lung and is expressed only in alveolar type II cells. The majority of the investigations of the SP-C gene have centred on the use of the SP-C promoter at targeting lung specific expression of other genes in transgenic mice. The use of reporter genes fused to the SP-C promoter has clearly shown the restricted expression of SP-C in the lungs (Weaver, 1991).

SP-C disrupts lipid packing in the surfactant, but has no effect on fluid surfactant (Johansson *et al.*, 1995). SP-C also helps in the packing of lipid chains within lipid monolayers and to encourage lipid spreading (Gericke *et al.*, 1997). SP-C is possibly anchored into bilayers via the amino-terminal palmitoyl chains, while positive charges on the alpha helix interact with the negative charges on phospholipid headgroups. The palmitoyl group on SP-C has a second role; it is essential for maintaining a reservoir of the major surfactant lipid, DPPC, in lung surfactant (Qanbar *et al.*, 1996).

1.3.2. The Surfactant Collectins

The first protein found to be associated with lung surfactant lipids was identified in 1973 and was a hydrophilic surfactant protein, later named SP-A (Possmayer, 1988). There are now known to be two hydrophilic surfactant proteins, SP-A and SP-D. These belong to a sub-group of mammalian lectins that are calcium dependant (C-type) lectins that contain collagenous stalks and are termed the collectins (Bezouska *et al.*, 1991; Malhotra *et al.*, 1992). Other members of the family include the serum protein, mannan binding lectin (MBL) (Turner, 1996), which activates the serum complement system (Ikeda *et al.*, 1987), and two bovine serum collectins, collectin-43 (CL-43) and conglutinin. A sixth member of the collectins has been described recently was isolated

originally from human liver and is termed CL-L1 (Ohtani *et al.*, 1999). This protein appears to be expressed predominantly in the liver and to a limited extent in the placenta as indicated by northern blotting, while reverse transcriptase-polymerase chain reaction (RT-PCR) based assays imply an almost ubiquitous expression. Despite the presence of a leader peptide, a collagenous region, a CRD and paired cysteine bonds within the CRD, which strongly suggest CL-L1 as an extracellular protein. However, the authors conclude this protein is cytosolic due to the absence of the protein from liver microsomes. Phylogenetic analysis and sequence alignment by the authors indicate that this molecule forms a unique branch of the collectins. The distinct nature of CL-L1 is also consistent with the finding that the CL-L1 gene is localised on a different chromosome to the other human collectins. In addition, the initial genomic information by the authors suggests that the neck and CRD are unique amongst the collectin as they are encoded by a single hybrid exon (Ohtani *et al.*, 1999). The authors speculate that the function of the molecule resides outside the immune system acting as a cytoplasmic molecular scavenger or chaperone. This protein is not discussed further in this thesis.

1.3.2.1. The Primary Structure of the Collectins

The polypeptide chains of each of the collectins share a similar modular organisation, as illustrated in Figure 1.4. The primary structures are composed of an amino-terminal cysteine rich region, a collagen-like region, an alpha-helical 'neck' region and a carbohydrate recognition domain (CRD). The basic tertiary structural unit of the collectins is a trimer of this polypeptide chain, stabilised by non-covalent interactions in the collagen-like region and the alpha-helical neck region as well as disulphide bonds within the amino-terminal domain. Triple-helical formation is considered to be initiated by the neck region in the polypeptide chain that forms coiled coils of alpha helices (Hoppe *et al.*, 1994).

The polypeptide chains for each collectin have different molecular weights; MBL varies from 28-32 kDa, SP-A from 28-36 kDa and conglutinin, CL-43 and SP-D have

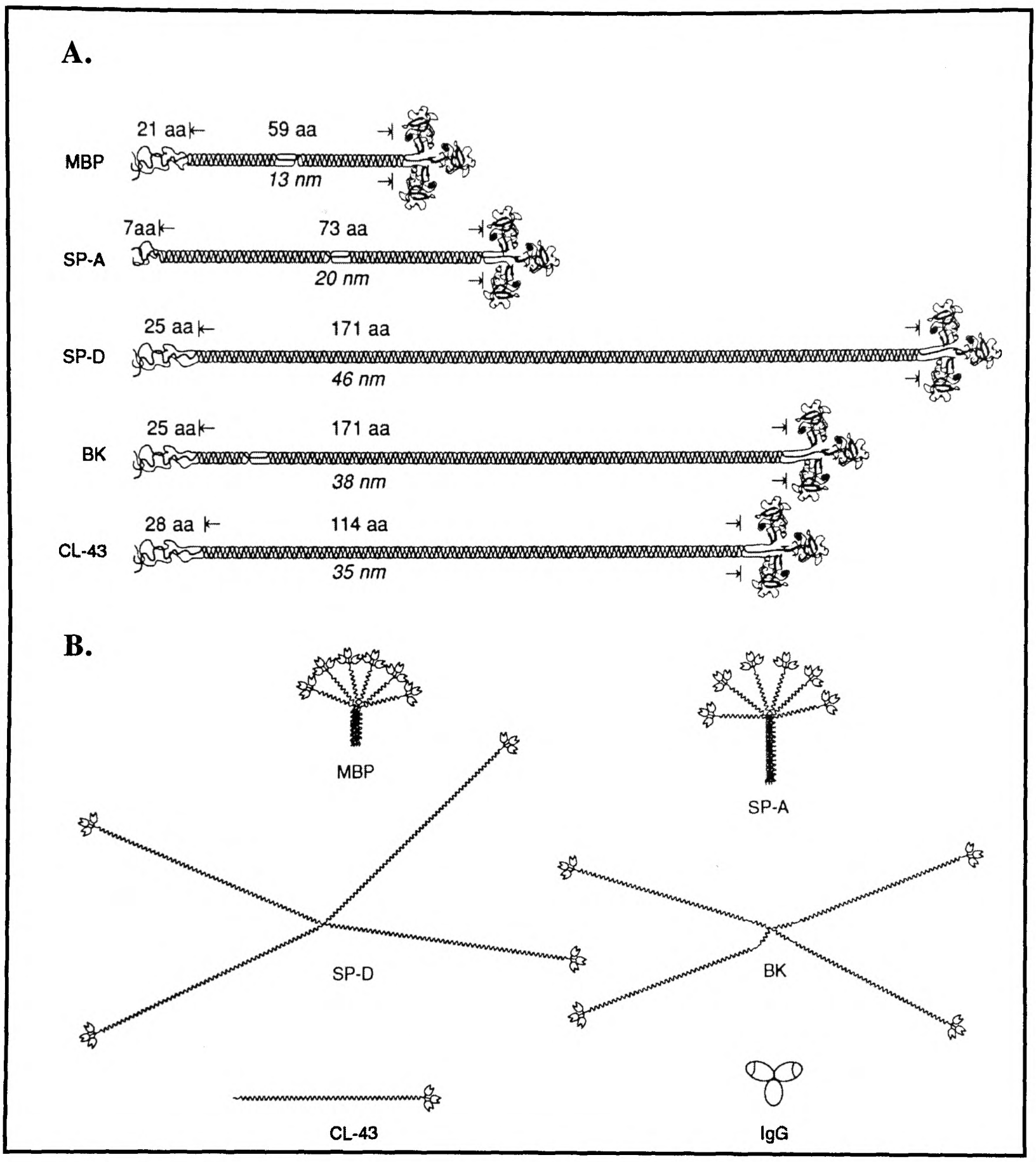


Figure 1.4. The polypeptide and multimeric structures of the collectins. **(A)** The tertiary structure of the collectins. The carbohydrate recognition domains (CRDs) are depicted by the structure determined using X-ray crystallography. The quaternary structures of the collectins are shown in **(B)** relative to IgG. From Holmskov *et al.*, 1994.

molecular weights between 42-44 kDa. The molecular weight of the collectins varies slightly between different species and is thought to be due primarily to different post-translation modifications (for example see Chapter 4, which contains information on the variation seen for porcine SP-D).

1.3.2.2. Higher Order Oligomers

Each member of the collectin family has a different degree of higher order oligomerisation. Higher order oligomerisation is mediated by cross linking between trimers, achieved through inter-trimer cysteine bond formation at the amino terminal cysteine rich domain or in the case of SP-A through a cysteine residue in the collagenous region that interrupts the three amino acid repeats in collagen (Gly-Xxx-Yyy repeats).

CL-43 is found as a monomer of the basic trimeric structural unit, SP-A and MBL are found as hexamers while SP-D and conglutinin forms tetramers. Lower order oligomers of SP-A, MBL, SP-D and conglutinin are also observed. SP-D has been observed to form even higher order oligomers containing 96 polypeptide chains that resemble 'fuzzy balls' under electron microscopy (Crouch *et al.*, 1994).

For human SP-A, there are two expressed genes, each differing in only 10 amino acid residues clustered in the amino terminal and collagenous region. The most significant difference is an extra cysteine residue after the interruption in the collagen region of the SP-A1 gene product (see section 1.3.2.3). Expression of SP-A from the individual cDNA chains in mammalian cells gave predominantly lower order oligomers (Voss *et al.*, 1991), suggesting that both genes are essential for the formation of higher order oligomers. The occurrence of irregular disulphide bonding when human SP-A is expressed from only one cDNA *in vitro* and the presence of both chains in sequenced preparations of native SP-A (Voss *et al.*, 1991), generally indicate that SP-A is a heterodimer of the two gene products, SP-A1 and SP-A2, in the ratio of 1:2 (Malhotra

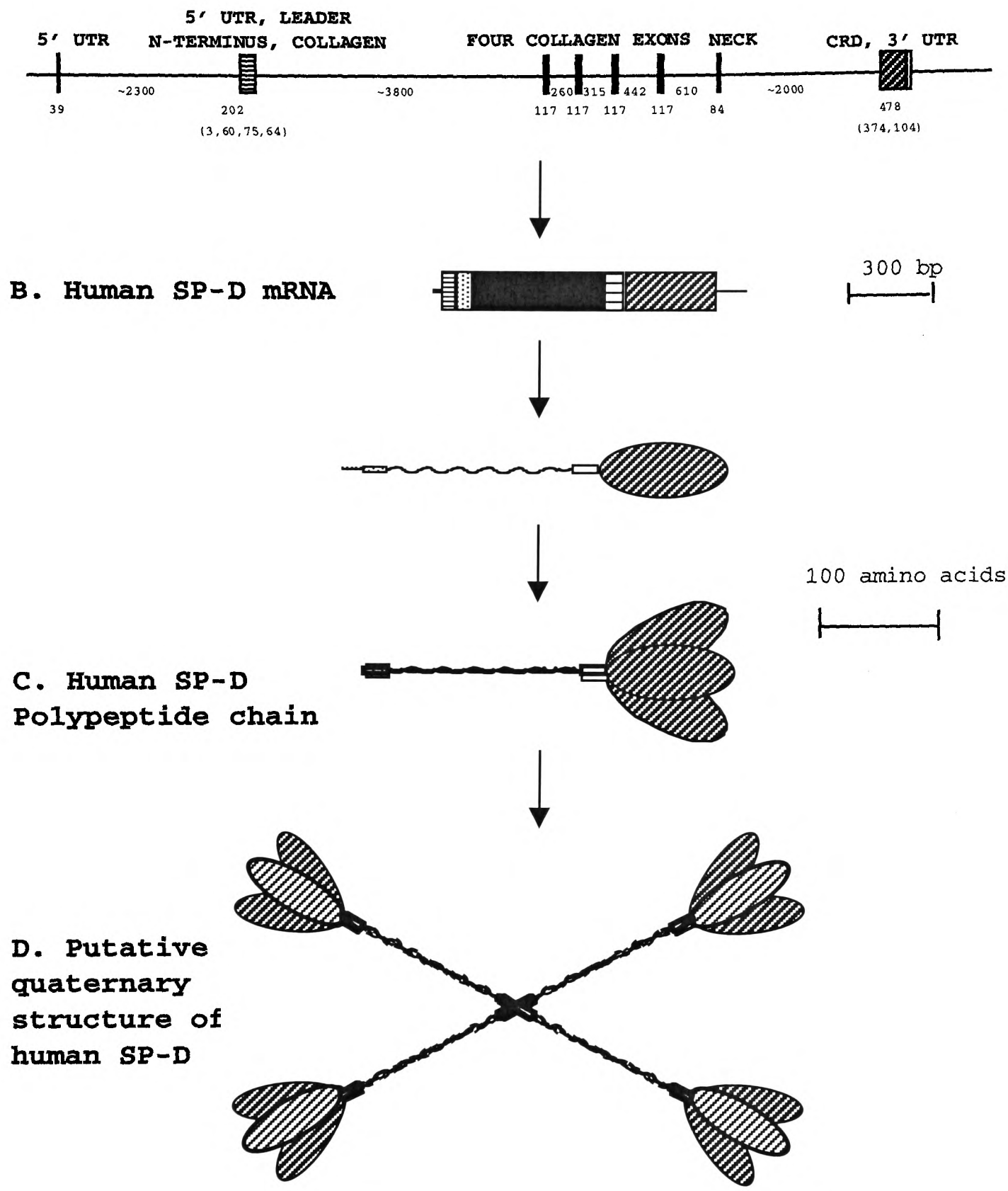


Figure 1.5. Human SP-D: From gene sequence to protein structure. Diagrammatic representation of the human SP-D gene, mRNA and protein. **(A, B)** Exons are shown as boxes and shaded according to the regions of the mRNA and protein they encode; 5' untranslated regions (black), signal peptide (dense horizontal lines), N-terminal region (dotted), collagen-like region (solid lines), neck region (horizontal lines) and CRD (dense diagonal lines). Lines represent introns. For **(B)**, boxes represent the protein coding regions, lines indicate the 5' and 3' untranslated regions. **(C, D)** The polypeptide forms a trimeric subunit, the leader peptide is indicated by a dotted line and the quaternary structure in **(D)** shows the association of four trimeric subunits via disulphide bonds at the N-terminal region, based on electon microscopy (Lu *et al.*, 1993) and biosynthesis (Brown Ausburger *et al.*, 1996a). Sizes of all elements are indicated.

et al., 1994). This contrasts with rodent and rhesus monkey MBL which also have two genes but these are believed to form two independent proteins assemblies (MBL-A and MBL-C). Humans have only one MBL gene that is analogous to MBL-C and an MBL pseudogene that expresses a truncated protein (Mogues *et al.*, 1996; Guo *et al.*, 1998).

1.3.2.3. Collagen Kink

The collagen-like sequence of the collectins is composed of a repeating Gly-Xxx-Yyy motif. SP-A and MBL resemble C1q (a subunit of the first component of complement, C1) in their tertiary structure by having a 'kink' in the collagenous stalks, possibly induced by an interruption of the collagen-like sequence (Sellar *et al.*, 1991; Lu *et al.*, 1993). Conglutinin also contains an interruption in the collagen repeat which is at the start of the collagenous region, so the kink in the collagenous stalk is less obvious and gives the appearance of an elongated central hub from which the collagen rods project. The generalised oligomeric forms of the collectins are shown in Figure 1.4. The formation of higher order oligomers from single polypeptide chains via trimers of the peptide chains is shown for SP-D as an example of higher order oligomerisation (Figure 1.5).

1.3.2.4. The CRD: Structure and Function

X-ray crystal structures have been determined for trimeric neck-CRD fragments of human MBL (Sheriff *et al.*, 1994), rat MBL-A (Weis and Drickamer, 1994) and human SP-D (Hakansson *et al.*, 1999) (see Figure 1.6). Three neck regions associate by non-covalent interaction to form an alpha-helical coiled coil. The neck regions interact with the CRD of neighbouring chains which is believed to generate specific spatial distances between the three CRDs. This may contribute to the types of complex carbohydrate structures that can be bound by each collectin. It is often speculated that the different complex carbohydrate structures found between self and non-self entities can be discriminated by the collectins, although there is no experimental evidence (Epstein *et al.*, 1996). Through gene segment recombination in B and T cells the number of



Figure 1.6. A model based on X-ray crystallography of the C-terminal portion of the rat serum mannose binding lectin. The three C-type carbohydrate recognition domains are shown complexed with a mannose-containing oligosaccharide. Only the alpha carbon backbone of the neck CRD is displayed and rendered in blue. Mannose-containing oligosaccharide is displayed as a ball and stick model in red and white. The white and green balls represent calcium ions. From Weis and Drickamer, 1994.

different antigen receptors is immense in contrast to the number of genes encoding the collectins and other lectins of the innate immune system. To compensate, the molecular recognition by the lectins in the innate immune system is broad in comparison to the highly specific recombinant antigen receptors of the adaptive immune system.

The C-type lectin domain is approximately 110-130 amino acids long and folds into a well-defined tertiary structure (see Figure 1.6). The CRD of each collectin has a distinct carbohydrate specificity (Table 1.1) which influences the targets that can be bound (Lu, 1997). By virtue of the CRD, receptors and, in the case of MBL, associated functional components, the collectins are all engaged to some degree in the innate immune system. Like the first component of the complement pathway (see section 1.4), MBL is the only collectin to have associated serine proteases, through which it mediates complement activation after binding carbohydrate targets (see section 1.4.2.3).

Collectin	**Carbohydrate specificity**
Human SP-A	ManNac > Fuc, Mal > Glu > Man >> Gal, Glc, GlcNAc
Human SP-D	Mal > Man, Glu > Lac, Gal > GlcNAc > Fuc
Human MBL	GlcNAc > Fuc, Man, ManNAc > Mal > Glu >> Gal, GalNAc
Conglutinin	GlcNAc > Fuc, Glu > Man, ManNAc >> Mal, Gal, GalNAc, Lac
CL-43	Man > ManNAc > Fuc, GluNAc > Glu, Mal > Gal, Lac >> GalNAc

Table 1.1. Carbohydrate selectivity of the collectins. As determined by saccharide inhibition of collectin binding to solid-phase mannan from several sources (Holmskov *et al.*, 1993; Lu *et al.*, 1992; Haurum *et al.*, 1993). GlcNAc, *N*-acetyl glucosamine; Fuc, L-fucose; Glu, glucose; Man, mannose; Gal, galactose; GalNAc, *N*-acetyl galactosamine; Lac, lactose; ManNAc, *N*-acetly mannosamine; Mal, maltose. Table was modified from a review article (Holmskov *et al.*, 1994).

1.3.2.5. The Neck Region and Carbohydrate Binding

The crystal structure of the alpha-helical coiled-coil 'neck' region and the CRD of MBL (Sheriff *et al.*, 1994; Weis and Drickamer, 1994) and later SP-D (Hakansson *et al.*, 1999) revealed that the neck region provides a second function other than trimerisation

(Figure 1.6). Hydrophobic interactions occur between the neck and the CRD of an adjacent chain to generate a distinct spatial distribution for the three CRDs. Conserved tyrosine residues at the top of the neck region in human SP-D that point into the centre of the alpha helical coiled-coil structure can also effect the arrangement of the CRDs (Hakansson *et al.*, 1999). This allows the possibility of a functional univalent subunit which, by virtue of each CRD, recognises different types of sugar but can also distinguish a pattern of carbohydrates ligands. The latter may be complex-branched, single carbohydrates with a defined distance between each of the terminal carbohydrate residues, or multiple glycoligands of unique spacing on the surface of a organism (Figure 1.6). Thus, carbohydrate specificity of a single CRD, combined with characteristic spatial discrimination, provides a system of recognising particular patterns of ligands for each collectin. This may be critical for non-self and self-discrimination by collectins in the innate immune system.

A notable feature seen in the structures for MBL and SP-D is a ring of positive charge created in the cleft formed by the three CRDs (Figure 1.7). It is speculated that this cluster of positive charge, sitting on top of the trimeric structural unit of SP-D, could interact with negatively charged ligands, such as lipopolysaachariде (Hakansson *et al.*, 1999).

1.3.2.6. The Gene Organisation

A. The General Genomic Arrangement of the Collectins

The intron-exon arrangement of the surfactant collectin genes is consistent with other members of collectin gene family (Figure 1.8). Essentially, the exon structure of the collectins can be described using five basic exon types. First, the 5' untranslated region (UTR) exon, of which the collectins contain one or more of these, for example SP-D has one, conglutinin has two. Second, a hybrid exon that encodes the remainder of the 5' UTR, the complete signal peptide, the amino terminal cysteine rich domain and the start of the collagen like region. The remainder of the collagen region is

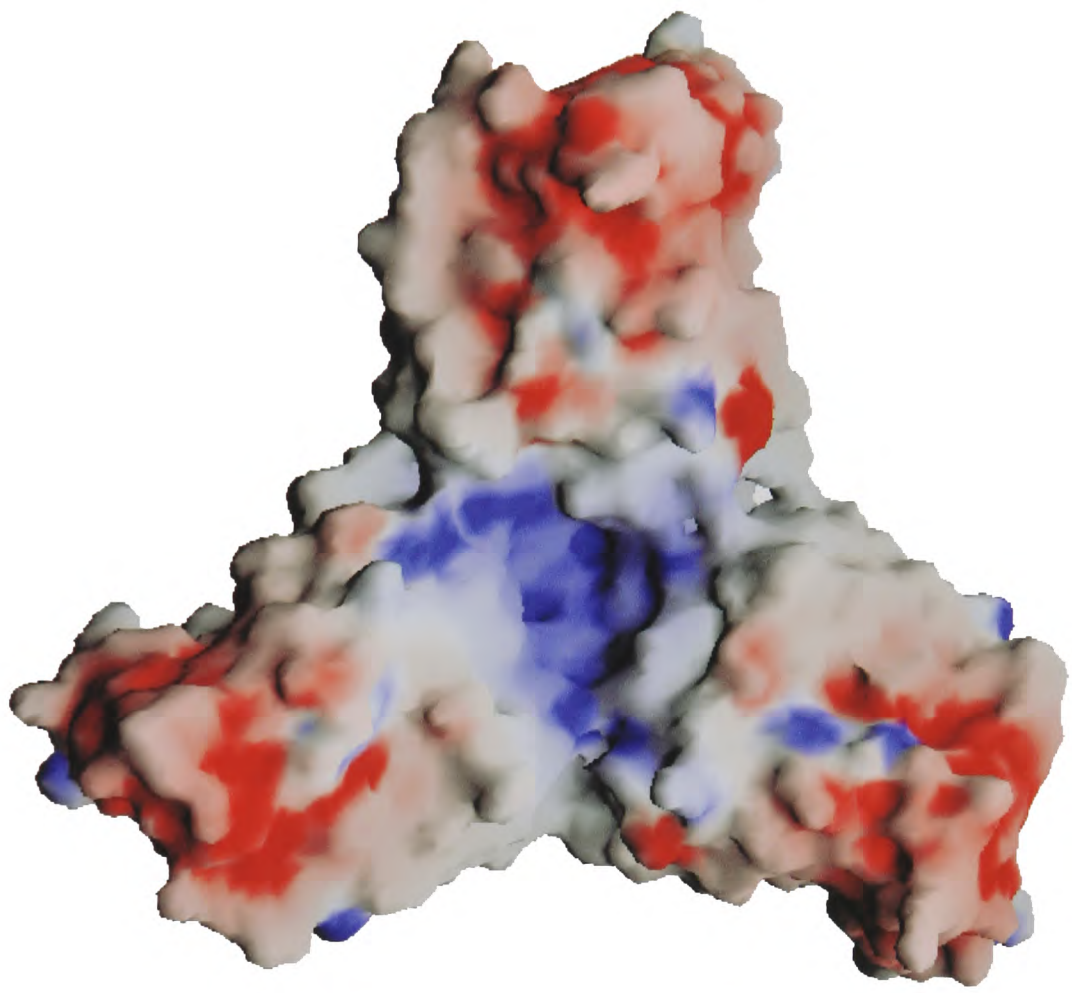


Figure 1.7. The electrostatic surface potential of the trimeric neck-CRD fragment of human SP-D. The molecule is shown from the top of the CRDs. Regions of positive potential are shown in blue and negative potential is shown in red. Adapted from Hakansson *et al.*, 1999.

encoded by the third type of exon of either 108, 117 or 120 bp in length. The number of these exons that exclusively encode collagen-like sequence, ranges from one to four among the collectins (see Figure 1.8). The organisation of the SP-D and conglutinin genes are strikingly similar, each having four collagen-encoding exons. SP-D collagen-like exons are each 117 bp in human (Crouch *et al.*, 1993) and mice (see Chapter 3) and in the case of bovine conglutinin, two exons are 117 bp long and the remaining two exons are 108 bp long (Kawasaki *et al.*, 1994; Liou *et al.*, 1994). The fourth exon type encodes the neck region that is responsible for initiating trimerisation of the polypeptide. The last exon encodes the entire CRD and the 3' UTR.

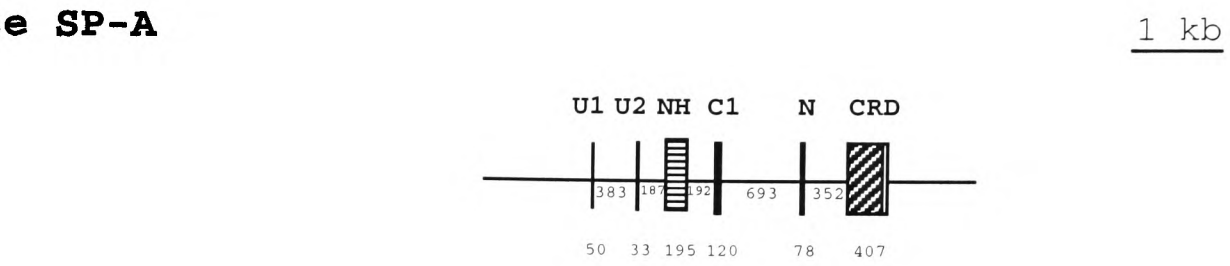
B. The Bovine Collectin Genes

Conglutinin and CL-43 have only been cloned in cattle. The cDNA and gene have been cloned for conglutinin (Liou *et al.*, 1994; Kawasaki *et al.*, 1994; Liou *et al.*, 1994) and only the cDNA has been cloned for CL-43 (Lim *et al.*, 1994), The conglutinin gene has been mapped to chromosomal 28 in cattle (Gallagher *et al.*, 1993) and by cross-hybridisation to a syntenic chromosome in river buffalo (Iannuzzi *et al.*, 1994). Conglutinin and CL-43 share the greatest sequence similarity to SP-D and are in fact believed to have arisen from gene duplication of the SP-D gene in bovidea. Chapter 3 of this thesis describes a partial gene sequence resembling CL-43 in a further member of the bovidea, namely sheep.

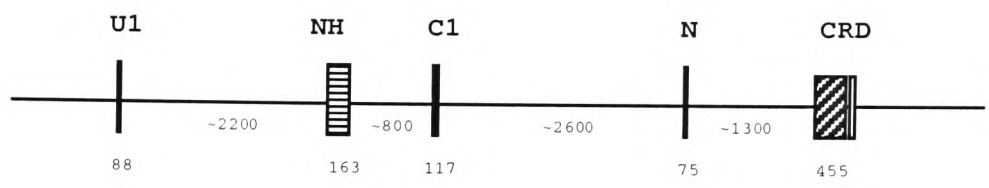
C. The Collagen Encoding Exons

The collagen region encoding exons are split by phase I introns occurring in the glycine codons of the Gly-Xxx-Yyy motif, thereby resembling the organisation of the non-fibrous vertebrate and invertebrate collagens (Exposito and Garrone, 1990). Split glycine codons have also been identified in the type IV collagens, the complement C1q genes (see section 1.4.4), the macrophage scavenger receptor gene (Emi *et al.*, 1993) and the human ficolin genes (Endo *et al.*, 1996). Phase I interrupted glycine codons are believed to be the most ancestral collagen exons (Exposito *et al.*, 1991). Phase I

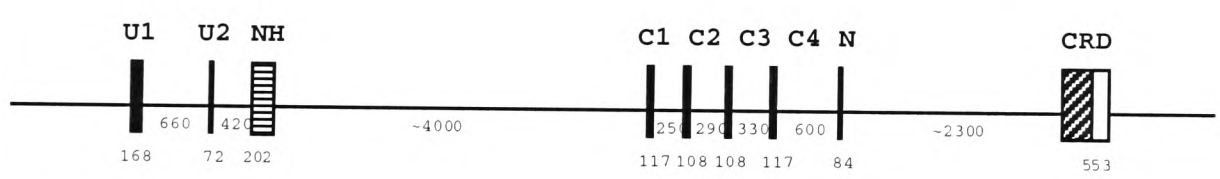
Mouse SP-A



Mouse MBL-A



Bovine Conglutinin



Human SP-D

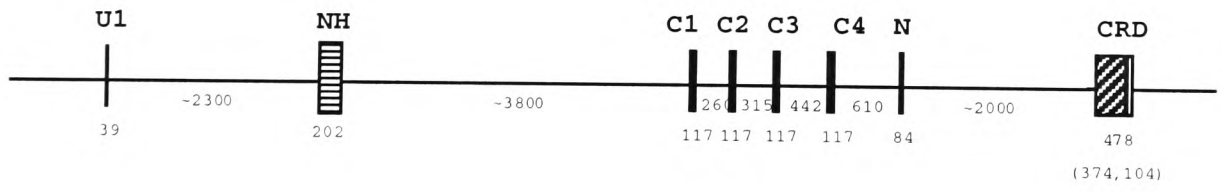


Figure 1.8. The organisation of the collectin genes. The boxes represent exons, with the name and number of the exon indicated above. Lines represent introns, with the sizes in base pairs shown below. The exons are abbreviated; **U**, untranslated; **NH**, amino terminal, cysteine rich and beginning of the collagen region; **C1-C4**, collagen encoding exons; **N**, neck exon; **CRD**, CRD exon.

introns are also the most prevalent type of intron and suitable for module shuffling (Patthy, 1987).

An outstanding feature of SP-D is that the collagen region is encoded by four identically sized 117 bp exons, clearly representative of collagen exon duplications. Despite the presence of these four genomic interruptions, the continuity of the Gly-Xxx-Yyy collagen-like motif of SP-D is not disrupted unlike similar collagen-containing proteins, such as C1q and MBL (Drickamer and McCreary, 1987).

D. The Collectin Gene Cluster

The gene arrangement of the human collectins has been established by radiation hybrid mapping (Hoover and Floros, 1998) and is shown in Figure 1.9. The surfactant collectins, the two SP-A genes (Bruns *et al.*, 1987; Fisher *et al.*, 1987), one SP-A pseudogene (Korfhagen *et al.*, 1991) and the SP-D gene (Crouch *et al.*, 1993) form a tight cluster on chromosome 10 (10q22.2-q23.1). The SP-A gene cluster lies telomeric of the SP-D gene, while the MBL gene lies further towards the centromere, at 10q21 (Sastry *et al.*, 1989). The human SP-A genes represent a recent gene duplication. A second gene for human MBL has also been mapped to the long arm of chromosome 10, however this is termed an expressed pseudogene as it is expressed as a truncated protein. It will be interesting to determine the exact location of this pseudogene in comparison to the surfactant collectins genes and to see if they form a chromosomal distribution similar to that of the murine genes.

Genetic mapping data indicates that a less dispersed collectin locus probably exists in mice, consisting of the single copies of the SP-A gene (*Sftpa*) and the SP-D gene (*Sftpd*) with one of the two MBL genes present in mice (*Mbl1*) co-segregating on chromosome 14. Restriction mapping and fluorescent *in situ* hybridisation (FISH) performed on a P1 Artificial chromosome clone (PAC) containing the complete murine collectin loci, has shown the murine genes are cluster in a 55 kb segment of DNA, with

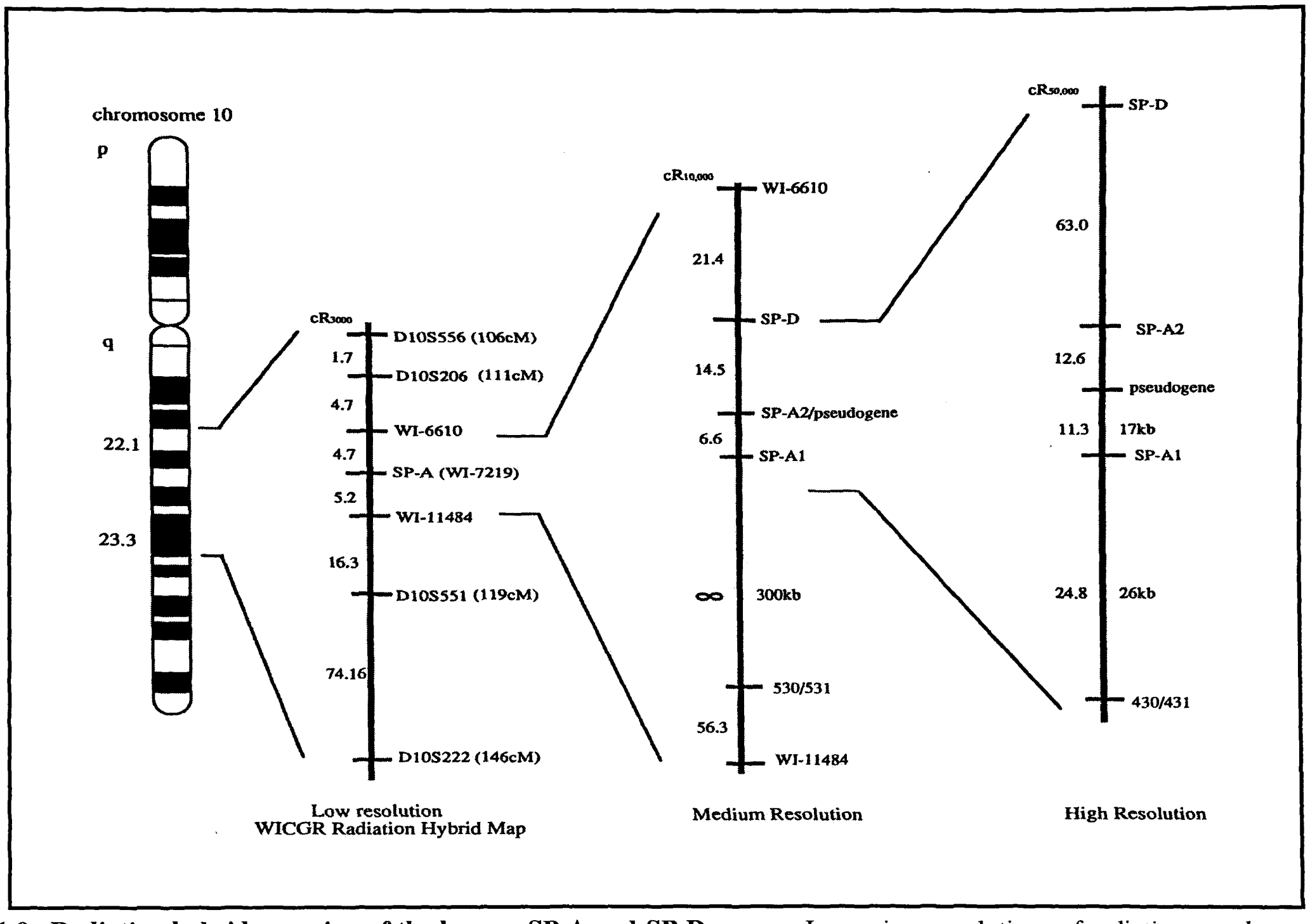


Figure 1.9. Radiation hybrid mapping of the human SP-A and SP-D genes. Increasing resolutions of radiation panels reveal the gene order of the human SP-A gene cluster (SP-A1, SP-A2 and the pseudogene) and the position of the SP-D gene. Physical distances between the genes were determined by restriction mapping the genomic clones and long-range PCR. From Hoover and Floros, 1998.

the gene order; *Sftpa-Mbl1-Sftpd.* The SP-A and MBL-A genes are in the same 5' to 3' arrangement and the SP-D gene is in the opposite orientation (Akiyama *et al.*, 1999). The distance separating these genes is consistent with the high-resolution mapping performed on the murine collectin gene described in this work in Chapter 3.

1.3.2.7. The Expression of the Surfactant Collectins

The surfactant collectins SP-A and SP-D is predominantly synthesised in the lung, being produced by both alveolar type II and Clara cells (Phelps and Floros, 1991; Crouch *et al.*, 1991). A number of reports have revealed that the surfactant collectins are expressed in a wide variety of non-pulmonary sites, abet at a far lower level. SP-D has been reported to be present in stomach (Fisher and Mason, 1995; Motwani *et al.*, 1995), trachea (Wong *et al.*, 1996), heart, pancreas, small intestines, ileal mucosa, colon (Lu, 1997; Van Eijk *et al.*, 2000), mesentery (Chailley-Heu B *et al.*, 1997), heart and kidney (Motwani *et al.*, 1995). SP-A along with SP-B has been shown to be associated with secreted surfactant-like particles in the intestinal lumen (Eliakim *et al.*, 1989; Eliakim *et al.*, 1997). In addition, SP-A has been detected immunochemically in the small and large intestines of the rat (Rubio *et al.*, 1995) and the lamellar body type structures of the serosal mesothelium (peritoneum, pleura, and pericardium) and synoviocytes (Dobbie *et al.*, 1994; Dobbie, 1996). MBL, conglutinin and CL-43 are predominantly expressed in hepatocytes (Wada *et al.*, 1990; Lim *et al.*, 1994; Lu *et al.*, 1993), however recently, rat MBL-A, but not MBL-C has shown to expressed in the kidney at a reduced level (Morio *et al.*, 1997).

The detection of the surfactant collectins in intestinal tissues and the ileal mucosa supports the potential function of these proteins in the host defence of mucosal surfaces other than lung. It must be noted, that extra-pulmonary expression of the surfactant collectins is probably species dependent. SP-D is expressed in the stomach of the mouse and rat (Fisher and Mason, 1995; Motwani *et al.*, 1995) but is absent in pigs.

Likewise, tracheal expression of SP-D is present in the mouse but absent in man (Wong *et al.*, 1996).

1.3.2.8. Roles of the Surfactant Collectins Proteins

Two of the chapters in this thesis (chapters 3 and 4) describe studies upon SP-D and to a lesser extent SP-A. The roles of these collectins will be detailed in this section with only special references made to the other members of the collectins. Further discussion of MBL can be found in the complement section of this introduction (section 1.4).

A. Investigating the Roles of SP-A and SP-D

Theories on the role of SP-D in surfactant have come full circle over the last decade. Initially, work concentrated upon the function of SP-D in lipid homeostasis (Persson *et al.*, 1990; Persson *et al.*, 1992). Later, *in vitro* evidence built upon a role for SP-D in the innate immune system. SP-D has been shown to bind directly to a variety of pathogens which can lead to phagocytosis of microorganisms (Hartshorn *et al.*, 1998). SP-D can modulate the immune cell and cytokine response (Borron *et al.*, 1998) and provide protective resistance to infection (Reading *et al.*, 1998). However, transgenic mice lacking SP-D have caused some reappraisal of the function of SP-D, since, unlike the immune compromised SP-A knockout mice (LeVine *et al.*, 1998), null SP-D mice exhibit a serious surfactant homeostatic imbalance (Botas *et al.*, 1998; Korfhagen *et al.*, 1998).

In the case of SP-A, initial investigations concentrated upon its role in lipid homeostasis and apparently showed clear roles in influencing lipid turnover and recycling. For SP-A, a wide variety of lipid ligands have been characterised and the amino acid residues responsible for lipid binding have been identified. Unlike SP-D, the roles of SP-A in both innate immunity and lipid binding have been described simultaneously. SP-A also binds to a wide and ever-expanding list of microorganisms, leading to a myriad of cellular responses. Complete deficiency of SP-A could have been predicted

to be fatal because of the strong role for SP-A in lipid homeostasis. However, the phenotype of transgenic mice lacking SP-A was found to be in stark contrast to this view. Mice null for the SP-A gene, show only minor disturbances in the surfactant lipid composition and dynamics (Korfhagen *et al.*, 1996). Moreover, mounting evidence clearly shows a role for SP-A in complementing the existing immune mechanisms and contributing to immune eradication of a wide range of microorganisms.

The functions of both SP-A and SP-D will be described starting with the types of targets that are bound, followed by the responses elicited.

B. Binding of SP-A and SP-D to Pathogens (see Figure 1.10)

One of the initial studies which demonstrated that surfactant proteins may participate in immune defence showed the beneficial effects of surfactant upon alveolar macrophage killing of *Staphylococcus aureus* (LaForce *et al.*, 1973). By 1990, two groups had reported the phenomenon of SP-A enhancing the uptake of pathogens through immunoglobulin receptors (Fc) and complement receptor 1 (CR1) on alveolar macrophages (Tenner *et al.*, 1989; van Iwaarden *et al.*, 1990). A similar finding was reported for C1q (Ohkuro *et al.*, 1994). There is now a large body of evidence indicating that SP-A and SP-D can bind or agglutinate a broad range of pathogens (viruses, bacteria, yeast and fungi), enhance the attachment of pathogens to phagocytic cells which can lead to killing and promote clearance through phagocytic cells. Table 1.2 attempts to summarise the pathogens and particulate matter that are bound, agglutinated and killed through the influence of SP-A and SP-D.

C. Glycoligands Recognised by SP-A and SP-D

One of the best-characterised native ligands for the surfactant collectins is lipopolysaccharide (LPS). In fact, the agglutination of *Escherichia coli* is mediated by SP-D binding rough LPS in a lectin dependent fashion (Kuan *et al.*, 1992). SP-D has

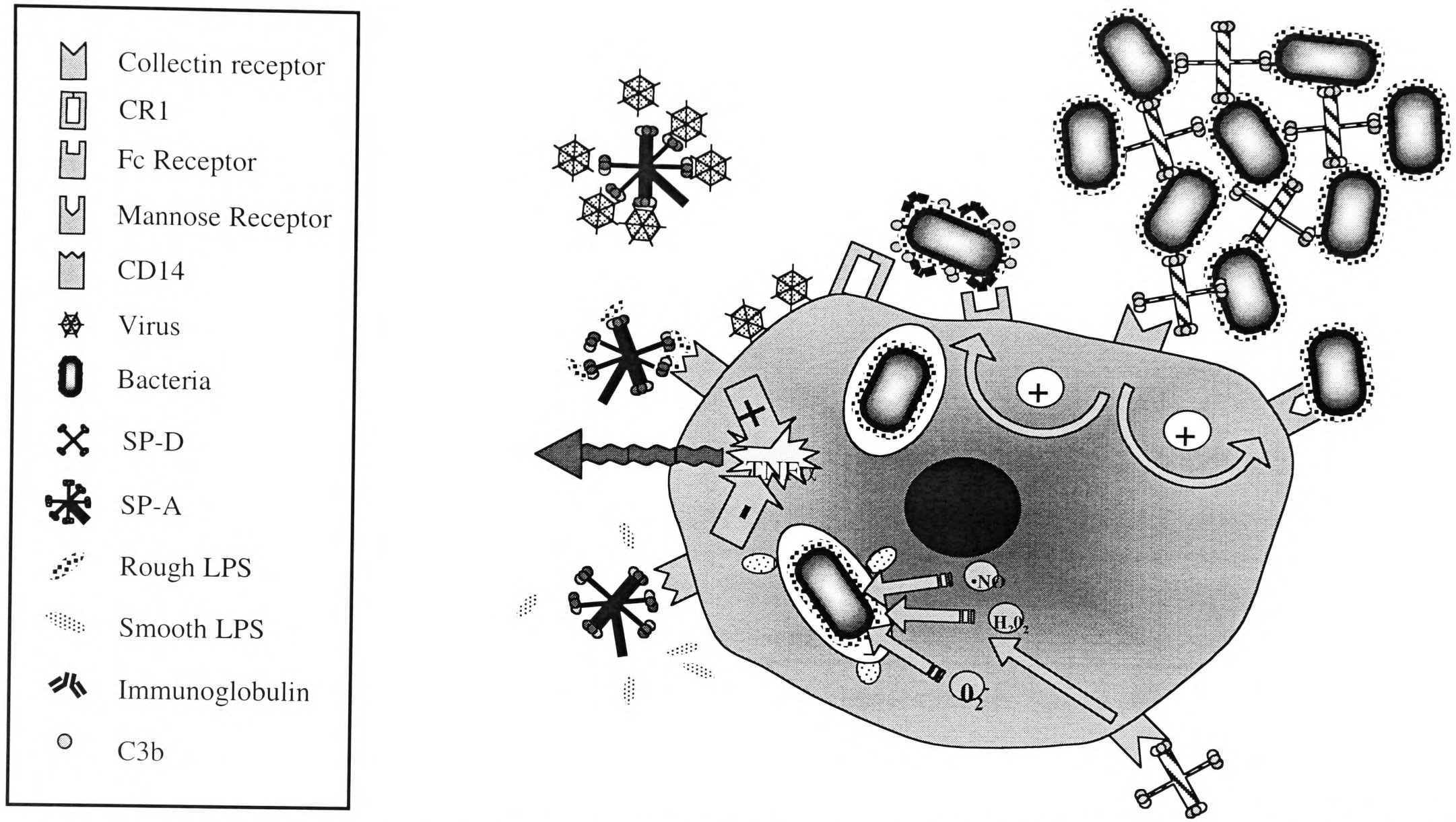


Figure 1.10. A summary diagram showing some of the functions of the collectins. (clockwise from the top): collectin-mediated agglutination of bacteria, plus enhanced mannose receptor uptake and Fc- and CR1-mediated uptake. Collectin-enhanced generation of reactive nitrogen and oxygen intermediates, as well as modulation of LPS-stimulated release of TNF-alpha. SP-A also binds viral particles and inhibits haemagglutinin activity. From Lawson and Reid, 2000.

	SP-D						SP-A					
Gram Negative	Binds	Agglu	Attach	Uptake	Killing	References	Binds	Agglu	Attach	Uptake	Killing	References
E. coli J5	✓	✓				(Kuan *et al.*, 1992)	✓		✓	✓	✓	(Van Iwaarden *et al.*, 1994; Pikaar *et al.*, 1995)
O111							✗		✗	✗	✗	(Van Iwaarden *et al.*, 1994; Pikaar *et al.*, 1995)
K12							✗			✓		(Manz Keinke *et al.*, 1992; Ohmer Schrock *et al.*, 1995)
Pseudomonas aeruginosa	✓	✗		✓ c	✓ c		✓✗	✗		✓✗		(Manz Keinke *et al.*, 1992; Ohmer Schrock *et al.*, 1995; Tino and Wright, 1996; Hickling *et al.*, 1998b)
Klebsiella pneumoniae							✓ c	✓ c	✓ c	✓ c	✓ c	(Kabha *et al.*, 1997)
Gram Positive												
Group B Streptococcus							✓			✓		(LeVine *et al.*, 1997; LeVine *et al.*, 1999b)
Group A Streptococcus							✓			✓		(Ohmer Schrock *et al.*, 1995; Tino and Wright, 1996)
Streptococcus pneumoniae (Untyped)		✓		✓		(Hartshorn *et al.*, 1998)	✓		✓	✓		(van Iwaarden *et al.*, 1990; Ohmer Schrock *et al.*, 1995; Tino and Wright, 1996; McNeely and Coonrod, 1993)
Type 25							✓		✗	✗		(McNeely and Coonrod, 1993)
Staphylococcus aureus		✓		✓		(Hartshorn *et al.*, 1998)	✓	✓	✓	✓✗	✗	(van Iwaarden *et al.*, 1990; Manz Keinke *et al.*, 1992; Ohmer Schrock *et al.*, 1995; Geertsma *et al.*, 1994; Hartshorn *et al.*, 1998; McNeely and Coonrod, 1993)
Fungo												
Pneumocystis carinii	✓		✓	✗		(O'Riordan *et al.*, 1995; Limper *et al.*, 1995)	✓		✓	✓✗		(Zimmerman *et al.*, 1992; Williams *et al.*, 1996; McCormack *et al.*, 1997a)
Aspergillus fumigatus	✓	✓	✓	✓	✓	(Madan *et al.*, 1997a; Madan *et al.*, 1997b; Crouch, 1998)	✓	✓	✓	✓	✓	(Madan *et al.*, 1997a; Madan *et al.*, 1997b)
Cryptococcus neoformans	✓	✓				(Schelenz *et al.*, 1995)	✓	✗		✗		(Schelenz *et al.*, 1995; Walenkamp *et al.*, 1999)
Candida albicans	✓					(Crouch, 1998)	✗			✗		(Rosseau *et al.*, 1997)
Histoplasma capsulatum	✓					(Crouch, 1998)						
Blastomyces dermatitides	✓					(Crouch, 1998)						
Virus												
Hemophilus influenza (not typed)	✓					(Hartshorn *et al.*, 1994)	✓	✓	✓	✓		(Malhotra *et al.*, 1994a; Benne *et al.*, 1995; Tino and Wright, 1996)
type A	✓	✓	✓✗ a	✗ b		(Hartshorn *et al.*, 1994; Hartshorn *et al.*, 1996a; Hartshorn *et al.*, 1996b; Malhotra *et al.*, 1994a)	✓	✓	✓	✓	✓	(McNeely and Coonrod, 1994; Hartshorn *et al.*, 1994; Malhotra *et al.*, 1994a; Benne *et al.*, 1995; Benne *et al.*, 1997)
type B							✓	✗	✗	✗	✗	(McNeely and Coonrod, 1994; Benne *et al.*, 1995)
Respiratory syncytial virus	✓✗					(Ghildyal *et al.*, 1999; Hickling *et al.*, 1999)	✓					(Ghildyal *et al.*, 1999)
Herpes simplex virus							✓			✓		(van Iwaarden *et al.*, 1991; van Iwaarden *et al.*, 1992)
Other bacteria												
Bacillus Galmette-Guérin							✓		✓	✓		(Weikert *et al.*, 1997)
Mycoplasma pulmonis							✓				✓	(Hickman Davis *et al.*, 1997; Hickman Davis *et al.*, 1999)
Mycobacterium tuberculosis	✓	✓	✗			(Ferguson *et al.*, 1999)	✓		✓	✓		(Downing *et al.*, 1995; Gaynor *et al.*, 1995; Ferguson *et al.*, 1999)
Miscellaneous												
Pollen							✓					(Malhotra *et al.*, 1993)
Mite extract	✓					(Wang *et al.*, 1996a)	✓					(Wang *et al.*, 1996a)

Table 1.2 Summary of the pathogens and allergens bound by SP-A and SP-D. The type of binding or response mediated through SP-A or SP-D is indicated. Binds, the surfactant collectin binds; Agglu, the organism is agglutinated; Attach, the surfactant collectin attaches the organism to a phagocytic cell; Uptake, the surfactant collectin enhances uptake of the organism; Killing, the organism is killed. Relevant references are cited in the table. Ticks (✓) indicate a positive finding, crosses (×) indicates that a response is not mediated. Both symbols together (✓×) indicates a discrepancy in the literature. Exceptions; [a] inhibited viral hemagglutinin activity, [b] inhibits infection and [c] strain dependant.

been shown to bind rough LPS derived from only certain organisms; *E. coli*, *Klebsiella pneumoniae* and *Pseudomonas aeruginosa* (Lim *et al.*, 1994). This suggests specificity and that the underlying structural determinant for recognition is a highly conserved and probably an essential element of microbial evolution.

Like SP-D, SP-A appears to bind only rough LPS (Van Iwaarden *et al.*, 1994), through which it probably mediates the phagocytosis and the eventual killing of rough LPS bearing bacteria (Pikaar *et al.*, 1995). SP-A recognises LPS through a different mechanism to SP-D, by binding to the lipid binding region of LPS (lipid A) (Van Iwaarden *et al.*, 1994; Kalina *et al.*, 1995). Despite binding only rough LPS, SP-A appears to have a direct action upon CD14 (the LPS receptor) and consequently influences both rough and smooth LPS-elicited cellular responses via CD14 (Sano *et al.*, 1999). SP-A binds to CD14, increasing rough LPS binding but decreasing smooth LPS binding. This in turn leads to a reduction in tumour necrosis factor alpha (TNF-alpha) release from alveolar macrophages elicited by smooth LPS and enhancement of TNF-alpha release with rough LPS (Figure 1.10) (Sano *et al.*, 1999). In addition, other evidence indicates that the effect of SP-A binding to LPS and alveolar macrophages can proceed through an alternative, CD14 independent, pathway (Stamme and Wright, 1999).

Other native ligands recognised by SP-A are a 95-120 kDa cell surface molecule from *Pneumocystis carinii* (Zimmerman *et al.*, 1992). In addition, both SP-A and SP-D recognise the carbohydrate determinants on the 45 kDa and 55 kDa glycoproteins secreted by *Aspergillus fumigatus* (Madan *et al.*, 1997). Furthermore, the interaction of SP-A with Lombardy polar pollen is mediated via carbohydrate dependent binding to two pollen grain proteins of 57 and 7 kDa (Malhotra *et al.*, 1993).

1.3.2.9. Effects Mediated by SP-A and SP-D Binding

A. Interactions with Lipids

Binding studies indicate that SP-A and SP-D have strong interactions with lipid components of surfactant and that these interactions are independent of the surface tension-lower properties of surfactant. SP-A has been shown to bind to a number of lipids, DPPC, sphingomyelin and to a lesser extent unsaturated PC, while SP-D has been shown to bind to glucosylceramide and PI (Persson *et al.*, 1992; Ogasawara *et al.*, 1992; Kuroki *et al.*, 1992; Ogasawara *et al.*, 1994; Taneva *et al.*, 1997). The physiological significance of these observations is as yet unclear. The interaction of SP-A with lipids is far better understood in relation to the uptake and release of surfactant lipids from type II cells (McCormack *et al.*, 1997). The most conclusive role of SP-A in lipid homeostasis is its requirement in the formation and stability of tubular myelin, demonstrated both *in vitro* (Suzuki *et al.*, 1989; Williams *et al.*, 1991) and *in vivo* with the use of transgenic animals (Clark *et al.*, 1995; Korfhagen *et al.*, 1996).

B. Immune Defence by the Surfactant Collectins

Agglutination

A visually elegant demonstration that the surfactant collectins participate in immune clearance is the cross-linking pathogens to form clusters. SP-D can efficiently and rapidly agglutinate suspensions of *E. coli* (Kuan *et al.*, 1992), *Cryptococcus neoformans* (Schelenz *et al.*, 1995) and *A. fumigatus* (Madan *et al.*, 1997). These clusters of pathogen are presumably dealt with by other arms of the immune system for clearance.

Opsonisation

The surfactant collectins fulfil one of the basic roles of an immune system, the careful specific tagging of foreign structures for disposal by immune cells, or opsonisation (Table 1.2). This is a specific reaction, an example is SP-A specifically opsonising *Hemophilus influenza* type A but not type B (McNeely and Coonrod, 1994) or SP-A acting as an opsonin for only particular encapsulated forms of *K. pneumoniae* (Kabha

et al., 1997). However, it should be noted that opsonisation does not always lead to phagocytosis; SP-D binds to *P. carinii* and enhances their adhesion to alveolar macrophages but it does not promote phagocytosis of this pathogen (O'Riordan *et al.*, 1995).

Chemotactic Responses

The surfactant collectins also mediate immune cell function by recruiting phagocytes at sites of inflammation and immune action. SP-D and SP-A, including a fragment of SP-D composed of the trimeric neck region and CRD (Cai *et al.*, 1999), can act as chemoattractants for neutrophils and monocytes (Pison *et al.*, 1992; Wright and Youmans, 1993; Crouch *et al.*, 1995; Madan *et al.*, 1997). The mechanism, receptors and *in vivo* relevance of this phenomena have yet to be confirmed. However, central to chemotaxis is actin breakdown and repolymerisation and both the surfactant collectins can mediate changes in the cellular architecture of alveolar macrophages, unlike C1q and MBL (Tino and Wright, 1999).

Enhanced Phagocytosis

The surfactant collectins have been shown to act both as opsonins and enhance phagocytosis through other ligand receptor mediated routes. It is well documented that both Fc receptor and CR1-mediated phagocytosis by alveolar macrophages of *S. aureus* can be enhanced in the presence of SP-A (Tenner *et al.*, 1989; van Iwaarden *et al.*, 1990). More recent studies have shown that SP-A can also enhance serum-independent uptake of *E. coli* (Manz Keinke *et al.*, 1992), *P. aeruginosa* (Ohmer Schrock *et al.*, 1995) and *H. influenza* (Tino and Wright, 1996). In addition, SP-A can enhance C1q mediated uptake of pathogens (Geertsma *et al.*, 1994) and mannose receptor mediated uptake of mycobacteria (Gaynor *et al.*, 1995). The mechanism by which the surfactant collectins enhance particle uptake is both specific and possibly mediated by different receptors, as highlighted by the difference in response towards a single pathogen. For example, in the presence of SP-D, *P. carnii* only shows enhanced binding to macrophages (O'Riordan *et al.*, 1995), while SP-A mediates both enhanced phagocyte

binding and phagocytosis of *P. carinii* (Williams *et al.*, 1996). However, SP-A and SP-D both enhance the uptake of *E. coli* by neutrophils (Hartshorn *et al.*, 1998).

There have been many reports of putative receptors described for SP-A and SP-D, but relatively few have been fully characterised and accepted as true receptors, for a detailed review (see (Tenner, 1999)). However, functional evidence for the presence of receptors does exists. Receptors are essential for the surfactant collectins to pursue their roles of enhancing phagocytosis, in chemotaxis, stimulating respiratory burst and mediating surfactant homeostasis through type II cells. SP-D has been described to mediate binding to leukocytes and macrophages by the lectin domain (Kuan *et al.*, 1994; Crouch *et al.*, 1995) and through lectin independent mechanisms (Miyamura *et al.*, 1994). SP-A appears to have three modes of cell binding: via a lectin-mediated event to monocytes and macrophages (Manz Keinke *et al.*, 1991; Wintergerst *et al.*, 1989); by a macrophage cell-surface lectin which binds the N-linked carbohydrate on the CRD (Gaynor *et al.*, 1995); and to alveolar macrophages via the collagenous stalks (Wright and Youmans, 1993). Most of these reports have yet to be substantiated by further studies nor the protein cloned at the cDNA or gene level. There is evidence for a functional receptor for SP-A, however, that is for a common SP-A, MBL and C1q, and enhances Fc-mediated phagocytosis. This common receptor, $C1qR_P$ has been cloned in humans and mice (Nepomuceno *et al.*, 1997; Norsworthy *et al.*, 1999) and is found on monocytes and macrophages. A receptor for SP-D has been identified on macrophages as a new member of the scavenger receptor super family, designated gp-340 (Holmskov *et al.*, 1997) which has also been reported to bind SP-A (Tino and Wright, 1999) and has recently been cloned (Holmskov *et al.*, 1999).

Killing by Reactive Oxygen and Nitrogen Intermediates

An important killing mechanism of phagocytic cells is by producing reactive oxygen and nitrogen intermediates. Both SP-A and SP-D have been shown to enhance the respiratory burst in alveolar macrophages (van Iwaarden *et al.*, 1990; van Iwaarden *et al.*, 1992; Van Iwaarden *et al.*, 1992; Weissbach *et al.*, 1994; van Iwaarden *et al.*,

1995). It appears this response is dependent upon the surfactant collectins forming an effective aggregate. For instance, SP-A will only enhance this response when it is bound to plastic (Weissbach *et al.*, 1994), while SP-D will not cause respiratory burst when the lectin domains are not engaged with ligand (Crouch *et al.*, 1995; Hartshorn *et al.*, 1994).

SP-A enhances killing of *Mycoplasma pneumoniae* through a ·NO dependant mechanism (Hickman Davis *et al.*, 1998). This is especially relevant since reactive oxygen intermediates do not mediate the killing of mycoplasma. SP-A is known to modulate ·NO production (Blau *et al.*, 1997). Mice deficient in SP-A show a striking inability to mount ·NO dependant killing of *Mycoplasma pulmonis.* In fact, the inability to kill *M. pulmonis* by SP-A knockout mice is comparable to that seen with the highly susceptible C3H/He mouse strain which have a defective ·NO pathway (Hickman Davis *et al.*, 1999).

C. Suppression of the Immune Response

Mounting evidence indicates that the functions of SP-A and SP-D extend beyond acting as a molecular 'flypaper' (Acton *et al.*, 1993) or 'Velcro' that clumps pathogens and cross-links pathogens to cells. A clear role in modulation of the immune response is developing for both SP-A and SP-D: quenching the potential immune-mediated damage from occurring to the delicate and labile lung surface. Both these molecules have been shown to inhibit the proliferation of T cells and IL-2 production (Borron *et al.*, 1998; Borron *et al.*, 1998; Wang *et al.*, 1998), perhaps explaining the phenomena of hyporesponsiveness of pulmonary leukocytes. A similar observation has been reported for C1q (Chen *et al.*, 1994). The inhibition of T cell proliferation by SP-D is believed to be a consequence of the suppressed IL-2 production, as the addition of exogenuous IL-2 blocked the inhibitory effect of SP-D. The inhibitory actions of SP-A and SP-D are lectin-independent and analysis of recombinant forms of SP-A indicated that a 36 amino acid region in the collagen region containing an Arg-Gly-Asp (RGD) motif

and interaction through an SP-A receptor were a route for inhibiting T cell proliferation (Borron *et al.*, 1998; Borron *et al.*, 1998). In addition, SP-A suppresses pro-inflammatory cytokine release from pathogen ligand stimulated alveolar macrophages by LPS (McIntosh *et al.*, 1996), and can reduce the release of TNF-alpha from buffy coat cells also challenged with LPS (Hickling *et al.*, 1998).

D. Surfactant Collectins and Allergy

The surfactant collectins may also modulate the inflammatory and allergic response. One of the first observations supporting this was the binding of SP-A to a number of different pollen grains and the identification of a water soluble glycoprotein derived from pollen as the ligand (Malhotra *et al.*, 1993). Later investigations have indicated that surfactant collectins may play an important role in suppressing immune responses to other inhaled allergens of different origins, such as fungal (*A. fumigatus*) and house dust mite (*Dermatophagoides pteronyssinus*) (Madan *et al.*, 1997; Wang *et al.*, 1996; Wang *et al.*, 1998).

The surfactant collectins have also been shown to influence the responses from two important immune cells involved in allergy. Firstly, SP-A has been shown to suppress experimentally induced IL-8 production from eosinophils (Cheng *et al.*, 1998), modulating allergic inflammation. Secondly, research has shown that both SP-A and SP-D can inhibit histamine release from basophils, an early effect of allergen provocation (Wang *et al.*, 1998) and, moreover, inhibit T cell proliferation, an important factor in the late phase of bronchial inflammation (Borron *et al.*, 1996; Borron *et al.*, 1998; Wang *et al.*, 1998). An unexpected finding is that the cross-linking properties of SP-D and possibly SP-A are not required for this anti-proliferative action; a fragment of SP-D composed of the trimeric neck region and CRD has been shown to be as effective as native SP-D (Wang *et al.*, 1998). The chemotactic effects of SP-D can also be mimicked by the trimeric neck-CRD fragment of SP-D (Cai *et al.*, 1999).

1.3.2.10. Functional Sites in SP-A and SP-D

A. Lipid Binding

Mutagenesis studies and epitope mapping using monoclonal antibodies have been used to localise the lipid binding site of SP-A to the CRD (McCormack *et al.*, 1994; McCormack *et al.*, 1997; Kuroki *et al.*, 1994). Lipid binding employs two of the essential residues for carbohydrate selectivity that are also important for Ca^{2+} co-ordination; Glu195 and Arg197 (all numbering is based on the rat SP-A sequence) (McCormack *et al.*, 1994). Alanine replacing mutagenesis in the CRD indicates lipid binding and aggregation require two levels of interaction or affinity. The first is recognition: a low affinity interaction with lipids that is required for adherence to lipid, but which cannot support higher affinity required for lipid aggregation. The two residues, Glu195 and Asp215, appear to be essential for lower affinity lipid binding as shown by complete disruption of all lipid interactions after alanine replacement. Two further amino acids (Arg197 and Glu202) in the CRD are required for increased lipid binding affinity which grants lipid aggregation. Replacement of the acidic Glu202 with alanine, or the basic Arg197 with polar (Asn) or acidic (Asp) residues prevents aggregation, but still permits lipid binding (McCormack *et al.*, 1997; Pattanajitvilai *et al.*, 1998).

B. The Structural Determinants for Carbohydrate Recognition

The X-ray crystallography of the CRD from rat MBL-A complexed with carbohydrate indicates the presence of three calcium binding sites (Weis *et al.*, 1992), as shown in Figure 1.6. In comparison to the crystal structure of rat MBL-C (Weis and Drickamer, 1994) and human MBL (Sheriff *et al.*, 1994), calcium binding site 1 and 2 are conserved. The third lower affinity calcium-binding site is probably crystallisation artefact.

Carbohydrate binding is a combination of hydrogen bonding and van der Waals interactions, in addition to co-ordination bonds with calcium ion 2 (Figure 1.11). Following the crystal structures and site-directed mutagenesis of rat MBL-C and rat MBL-A (Ng *et al.*, 1996; Iobst *et al.*, 1994), six key residues have been identified that are responsible for carbohydrate recognition; Glu185, Asn187, His/Val189, Glu193, Asn205 and Asp206 (numbering based on rat MBL-A). Sequence analysis of the CRD from different C-type lectins reveals Glu193, Asn205 and Asp206 are conserved regardless of carbohydrate specificity. Asp206 only forms co-ordination bonds with the calcium ion. Glu185, Asn187, Glu193, and Asn205 also form co-ordination bonds with the calcium ion and their side chains form hydrogen bonds with carbon 3 or carbon 4 of the monosaacharide and His/Val189 forms van der Waals interactions. In MBL-A, a similar mode of protein-sugar interaction is observed except that the pyranose ring is rotated 180° in MBL-C (Ng *et al.*, 1996).

With regard to carbohydrate specificity, Glu185 and Asn187 are relatively conserved in mannose/glucose binding C-type lectins (Weis *et al.*, 1992), except SP-A where Asn187 is replaced with either Ala or Arg. In galactose binding C-type lectins, Glu185 is replaced with Gln and Asn187 is replaced with Asp. Such mutagenesis of MBL-A, SP-A and SP-D changes the carbohydrate specificity from mannose/glucose to galactose (Drickamer, 1992; McCormack *et al.*, 1994; Ogasawara and Voelker, 1995). By replacing His189 with Trp the affinity for galactose in MBL-A can be increased.

The collectins bind D-hexoses through the equatorial 3- and 4-OH groups of the pyranose ring (Weis *et al.*, 1992). The only exception is L-fucose which mimics the arrangement of the hydroxyl groups with 3- and 2-OH groups. Specificity is governed by these steric constraints as carbohydrates without these requirements have no detectable affinity. Specificity is also governed through higher levels of carbohydrate organisation or structure: the spatial arrangement.

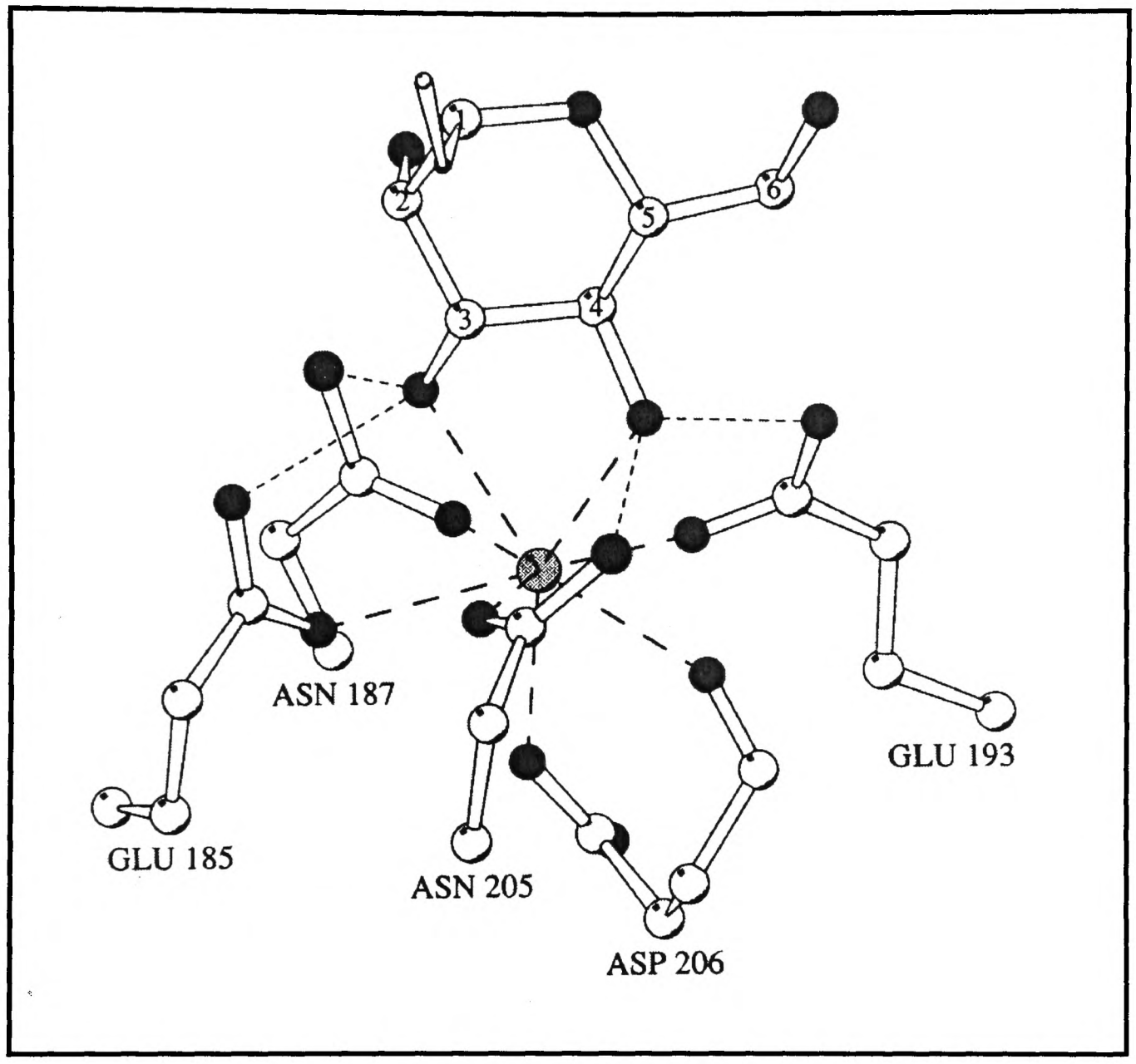


Figure 1.11. The carbohydrate binding site of MBL. X-ray crystallographic analysis of an MBL-A-asparaginal oligosaccharide complex revealed the relative positions of Glu-185, Asn-187, Glu-193, Asn-205 and Asp-206 in the carbohydrate binding site, a Ca^{2+} ion and the terminal mannose residue of the ligand. Dotted lines indicate hydrogen bonds. Dashed lines indicate co-ordination bonds. From Weis *et al.*, 1992.

C. Multimerisation

The amino-terminal 'multimerising' region of SP-A is another essential element of the molecule that is required to maintain liposome aggregation, increasing avidity by increasing the available number of CRDs that can allow cross-linking of liposomes (Figure 1.12). Substitution of Cys6 (which is important for the interchain disulphide bond formation) with serine, leads to a reduction in the oligomeric state of SP-A and consequently disrupts lipid aggregation (McCormack *et al.*, 1997; Zhang *et al.*, 1998). The interaction of SP-A with lipid and glycosylated ligands is relatively weak by individual or trimeric CRD. Greater binding affinity, such as that required for agglutination and phagocytosis/uptake, can be greatly enhanced by simultaneous binding of two or more of trimeric lectin subunits from an individual molecule (Figure 1.12).

D. Increased Avidity

The essential requirement of multimeric trimers for the efficient aggregation of targets is also evident from studies on SP-D. In the case of agglutination of bacteria and viruses, it is becoming clear that mere trimers of SP-D will not work (Eda *et al.*, 1997; Brown Augsburger *et al.*, 1996) and that this requires multimerisation of SP-D through inter-trimeric unit disulphide bridges at the amino terminus (Brown Augsburger *et al.*, 1996). Consequently, even higher order oligomers of SP-D, such as the multimers of the usual tetramers of trimers seen in both preparations of SP-D recombinant protein (Hartshorn *et al.*, 1996; Hartshorn *et al.*, 1996) and SP-D isolated from proteinosis patients (Crouch *et al.*, 1993), have an even higher affinity for glycosylated ligands, such as *H. influenza* type A than the typical cruciform variant of SP-D(Hartshorn *et al.*, 1996).

1.3.2.11. Levels of collectins within the human population

Human MBL is the most characterised collectin as regards to allelic variations and protein levels within the population. Similar studies have yet to be conducted for SP-A and SP-D. The plasma level of MBL in plasma is determined genetically and varies

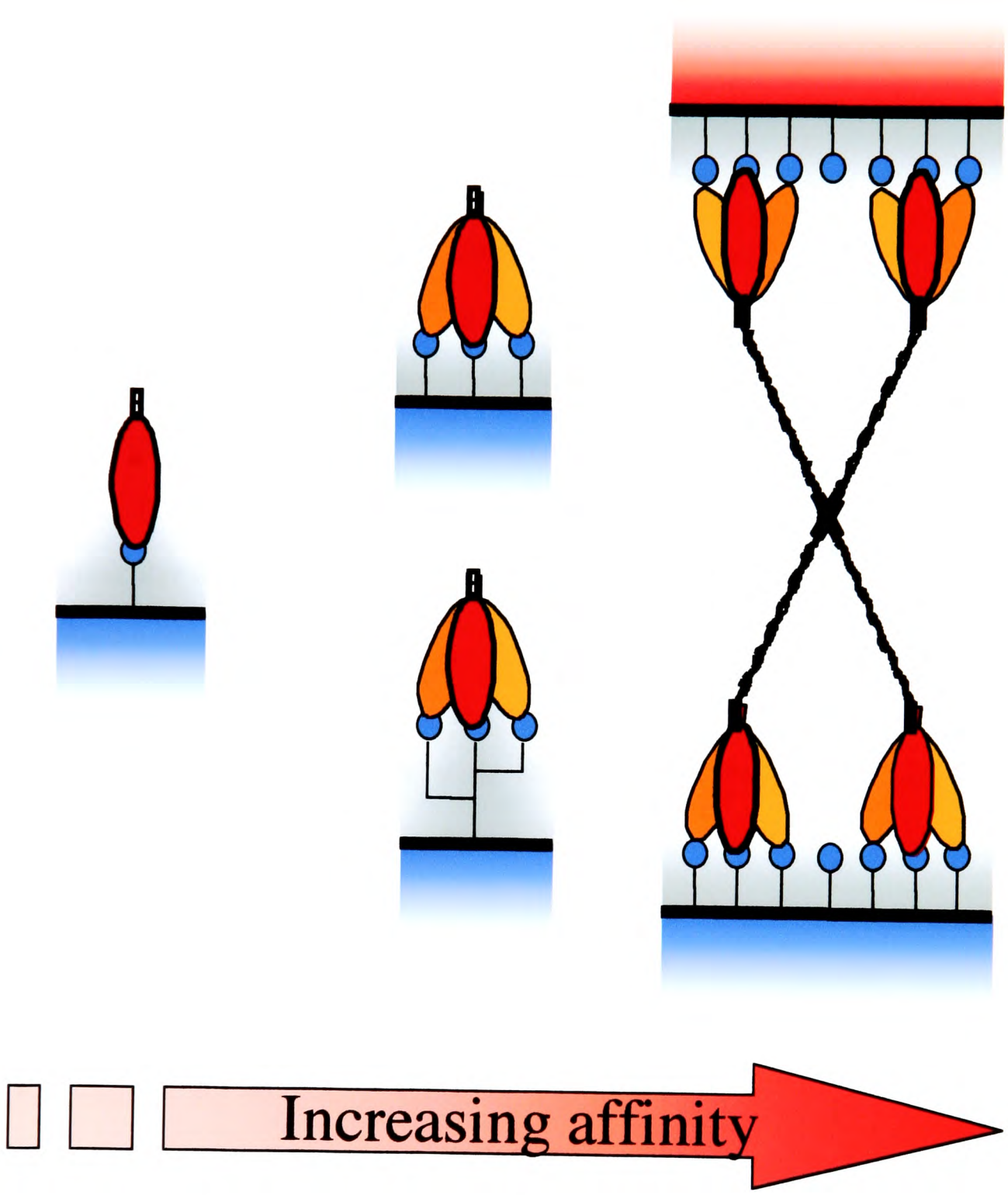


Figure 1.12. Increasing affinity by multimerisation. The increase in affinity for SP-D binding a glycosylated ligand on a surface, for example bacteria. A single CRD is shown interacting with a single glycosylated ligand. Trimerisation through the 'neck' alpha helical coiled-coil region permits three CRDs to interact either with three separate ligands (above) or one branched complex polysaccharide, such as yeast mannan (below). A further increase in affinity is generated by multimerisation of the trimeric polypeptide chain, hereby cross-linking two surfaces. A similar scenario could be described for SP-A. From Lawson and Reid, 2000a.

amongst individuals, from 2 ng/ml to 10 μg/ml. In fact, low levels of the MBL have been associated with a defect in opsonisation and increased risk of infection during childhood(Super *et al.*, 1989; Turner, 1996).

There are three variants in exon 1 of MBL (Madsen *et al.*, 1995; Turner, 1996; Madsen *et al.*, 1994) that have been linked to increased susceptibility to recurrent infections in infants (Sumiya *et al.*, 1991; Summerfield *et al.*, 1997) and adults (Summerfield *et al.*, 1995; Garred *et al.*, 1995). The variant alleles of MBL have also been shown to predispose individuals to enhanced susceptibility to HIV and progression to AIDS (Pastinen *et al.*, 1998; Garred *et al.*, 1997). Individuals with the mutant MBL alleles have low molecular weight forms of MBL, 100-200 kDa, compared with the wild-type MBL of approximately 600 kDa.

1.3.2.12. Association of Alleles with Disease and Protein Levels

In humans there are two genes which code for SP-A; SP-A1 and SP-A2. The levels of SP-A vary widely in the population in comparison to those for another surfactant protein, SP-B. In addition, the hormonal regulation of SP-A levels appears complex. This is accounted for, in part, by different regulation mechanisms of the two SP-A genes. For instance, the response of SP-A1 shows greater inhibition than SP-A2 with dexamethazone treatment (Karinch *et al.*, 1998; Hoover and Floros, 1999). There also appears to be shift in expression ratios during development. SP-A1 is expressed to a larger degree during gestation while in adult life there is a bias towards SP-A2 gene expression (Scavo *et al.*, 1998). Two reports have shown differential tissue expression, SP-A2 is expressed in the human fetal tracheal, bronchial epithelium and in submucosal glands while SP-A1 appears in the lower respiratory tract (Goss *et al.*, 1998; Saitoh *et al.*, 1998). Difficulties in understanding the regulatory differences are compounded by the influence of polymorphisms in the 5' UTR of both human SP-A genes (Floros and Hoover, 1998). There are four alleles for the SP-A1 gene (Karinch and Floros, 1995) and five alleles for the SP-A2 gene (Karinch *et al.*, 1997). One

genotype is associated with low levels of SP-A (Karinch *et al.*, 1997) and another SP-A allele (1A0) together with an SP-B variant seem to associate with RDS in infants (Kala *et al.*, 1998). SP-A may contribute to the either the aetiology of RDS or at least help to define subgroups within this multi-gene disease. Genetic analysis of the surfactant collectins in lung disease and infection are further confounded by the known tight cluster of collectin genes (Hoover and Floros, 1998).

There are no published reports of polymorphisms or alleles for the human SP-D gene; whether promoter region variations, different 5' UTR exons or coding region polymorphisms. This would be an interesting area of study considering the known variation in SP-D levels within the population (Kuroki *et al.*, 1998), the promoter region variations of MBL (Madsen *et al.*, 1995; Madsen *et al.*, 1998) and the SP-A alleles described above.

1.3.2.13. Varying Levels of the Surfactant Collectins in Disease

If the surfactant collectins play a central role in the innate immune defence of the lung, it would be expected that altered levels of these proteins would contribute to lung disease and infection. SP-A and SP-D have already been recognised as potential genes associated with pulmonary disease and respiratory infection (Floros and Hoover, 1998). Certain polymorphisms in SP-A have been found within the coding region, so the function of the protein or structure may be effected. One SP-A2 allele (1A0) is associated with low SP-A mRNA levels and is found in particular subsets of patients with RDS. Disturbances in the levels of the collectins would be expected to have pronounced effects in immuno-compromised individuals. SP-A has been shown to be depleted in certain types of acute lung injury (Jobe and Ikegami, 1998). Low levels of SP-A and SP-D have been found in cystic fibrosis patients who are very susceptible to common lung infections (Postle *et al.*, 1999). Reduced levels of SP-A and SP-D are seen in respiratory syncytial virus (RSV) infection (Kerr and Paton, 1999). Decreases in SP-A concentrations have been reported to occur in lung damage (Gunther *et al.*,

1996). In patients with ARDS, neutrophil-like elastase damage to SP-A has been reported (Baker *et al.*, 1999), thus causing functional depletion of SP-A. Functional depletion of SP-A can also be seen in alveolar proteinosis and pollen allergy patients (Hickling *et al.*, 1998). In these patients there is a shift from the higher to the lower multimeric forms of SP-A. Although this is mediated through an environmental mechanism, this is similar to the allelic variants in exon 1 of MBL which also result in lower molecular forms of MBL in the sera. Since lower order oligomers of MBL are less effective in binding carbohydrate ligands and activating the complement system (Lu *et al.*, 1990; Wallis and Drickamer, 1999) this leads to reduction in the functional concentration of protein. Decreases in the levels of SP-A may be responsible for certain types of lung damage. SP-A has been shown to reduce lung damage induced by lyso-phospholipid by inhibiting its generation by phospholipase A_2 (Arbibe *et al.*, 1998).

In an elegant and intuitive study it was shown that high levels of free glucose in diabetic mice reduced the opsonisation of *E. coli* by SP-D, leading to an increased susceptibility to bacterial infection. The inhibitory effect of glucose could be reversed by administration of insulin (Reading *et al.*, 1998). The free carbohydrate (glucose) simply competes with glycoligands for SP-D binding. This experiment mirrors the ELISA assays performed routinely by investigators of SP-A and SP-D for determining lectin dependant binding to the ligands of pathogens.

There have been no reports that the surfactant collectins act like MBL as acute phase proteins. MBL shows a modest 1.5 to 3 fold increase of serum concentration after malaria infection and trauma (surgery) (Thiel *et al.*, 1992).

1.3.2.14. Transgenic Animals Lacking the Collectins

A. SP-A

The SP-A knockout mouse has unexpectedly marginal defects in surfactant homeostasis and respiratory function; namely the absence of tubular myelin and increased tissue and

alveolar pools of saturated phosphatidylcholine (Korfhagen *et al.*, 1996; Ikegami *et al.*, 1997; Ikegami *et al.*, 1998). Whilst information about the role of SP-A in lipid homeostasis has been limited from these mice, they have helped the understanding of SP-A as an innate immune molecule. The experiments with the SP-A knockout animal have both confirmed and extended the myriad of mechanisms that SP-A utilises against pathogens. To date, SP-A deficient mice have shown enhanced susceptibility to infection against three different pathogenic organisms and a virus (see Table 1.3 and (LeVine *et al.*, 1997; LeVine *et al.*, 1998; LeVine *et al.*, 1999; LeVine *et al.*, 1999; Hickman Davis *et al.*, 1999)). The type of immune response mediated in these knockout animals is dependant upon the type of organism (Table 1.3), which could suggest that SP-A impinges upon many different mechanisms of the humoral and cellular immune systems. This is substantiated by the finding that the free radicals generated during an immune response by these knockouts differ with varying types of infection. For example, superoxide (0_2^-) and hydrogen peroxide (H_2O_2) release is suppressed in both RSV and group B streptococcal infections, but remains unchanged during *P. aeruginosa* infection, where nitrates are increased instead (LeVine *et al.*, 1997; LeVine *et al.*, 1998; LeVine *et al.*, 1999). A major deficiency in the ·NO mediated pathway for killing is seen during *M. pulmonis* infection in SP-A deficient mice comparable to a ·NO defective mouse strain (Hickman Davis *et al.*, 1999). These results are indicative of the multitude of immunological mechanisms that the SP-A collectin contributes towards.

Susceptibility to an infectious agent can be countered in mice lacking SP-A by the administration of exogenous SP-A during immune challenge (LeVine *et al.*, 1997; LeVine *et al.*, 1999). It is of particular interest that SP-D is not upregulated by either the absence of SP-A (Korfhagen *et al.*, 1996) or during infection of SP-A deficient mice (LeVine *et al.*, 1997; Hickman Davis *et al.*, 1999).

Effect	Organisms			
	P. aeruginosa	**RSV**	***M. pulmonis***	**group B streptococcus**
PMNL recruitment	Same number earlier influx	Increased		Same
Proinflammatory cytokines	Increased	Increased		
Reactive toxic intermediates	Same oxygen Increased nitrate	Decreased oxygen	Reduced nitrate	Decreased oxygen
Rescued with exogenous SP-A		Yes		Yes
Phagocyosis	Decreased			Decreased

Table 1.3. Comparison of the immune effects in null SP-A mice after experimental infection with four different organisms. Showing how the polymorphonuclear leukocytes (PMNL) influx is altered, if proinflammatory cytokines (TNFalpha and IL-6) are increased, what type of reactive toxic intermediates are produced or inhibited, if exogenous SP-A, when delivered together with the pathogen, can reduce the infection and finally the effect upon phagocytosis of the organism. Information from (LeVine *et al.*, 1997; LeVine *et al.*, 1998; LeVine *et al.*, 1999; LeVine *et al.*, 1999; Hickman Davis *et al.*, 1999).

B. SP-D

The phenotype of the null SP-D mouse has caused a certain degree of reassessment for the role of SP-D as an innate immune molecule (Korfhagen *et al.*, 1998; Botas *et al.*, 1998). The phenotype of SP-D (-/-) mice shares features with alveolar lipoproteinosis disease, the accumulation of foamy activated macrophages, hypertrophy of alveolar type II cells and an increase in lamaller bodies. This surprising result has been confirmed by two groups (Korfhagen *et al.*, 1998; Botas *et al.*, 1998). However, both groups used similar targeting constructs, so spurious effects from the use of a recombinant SP-D gene or construct cannot, as yet, be ruled out. An experiment that would differentiate between the lack of SP-D or the result of a mutated SP-D gene product would be to cross the SP-D knockout mouse with a SP-D 'knockin' mouse containing the wildtype SP-D allele on a different chromosome.

In the mice lacking SP-D, the study of infection is hampered by the underlying phenotypes of surfactant and macrophage disturbances. It is possibly more informative to examine the innate immune function of SP-D by another route. For example, generating a transgenic mouse with a high level of SP-D mediated expression to study SP-D prevention of infection.

1.4. The Complement System

The complement system acts as an important part of the innate immune system, generating several mechanisms of immune protection (Reid and Law, 1995) as well as complementing and influencing the adaptive clonal immune system (Carroll, 1998; Fearon and Locksley, 1996). Activation of the complement system occurs by three distinct pathways and results in a sequential cascade of specific proteolytic and conformational changes to the complement components that initiate various immune-effector mechanisms. See Figure 1.13 and recent reviews for in-depth discussion on complement (Fearon, 1998; Carroll and Fischer, 1997; Ravetch and Clynes, 1998; Schwaeble and Reid, 1999).

The complement system discriminates between self and non-self, leading to the defence against microorganisms, stimulation of inflammation, clearance of immune complexes and microorganisms and enhancement of the adaptive immune response. Complement is a highly reactive and potentially destructive arm of immunity, but it contains components that maintain tight control over its own activity. Activation proceeds through three different pathways that converge upon the same central component of the complement cascade, C3. The proteolytic cleavage of this component exposes a highly reactive thioester and allows the formation of a self-perpetuating cleavage enzyme complex, or convertase.

This thesis is primarily concerned with two of the three complement activation pathways, the classical and MBL. These converge prior to C3 of the complement and

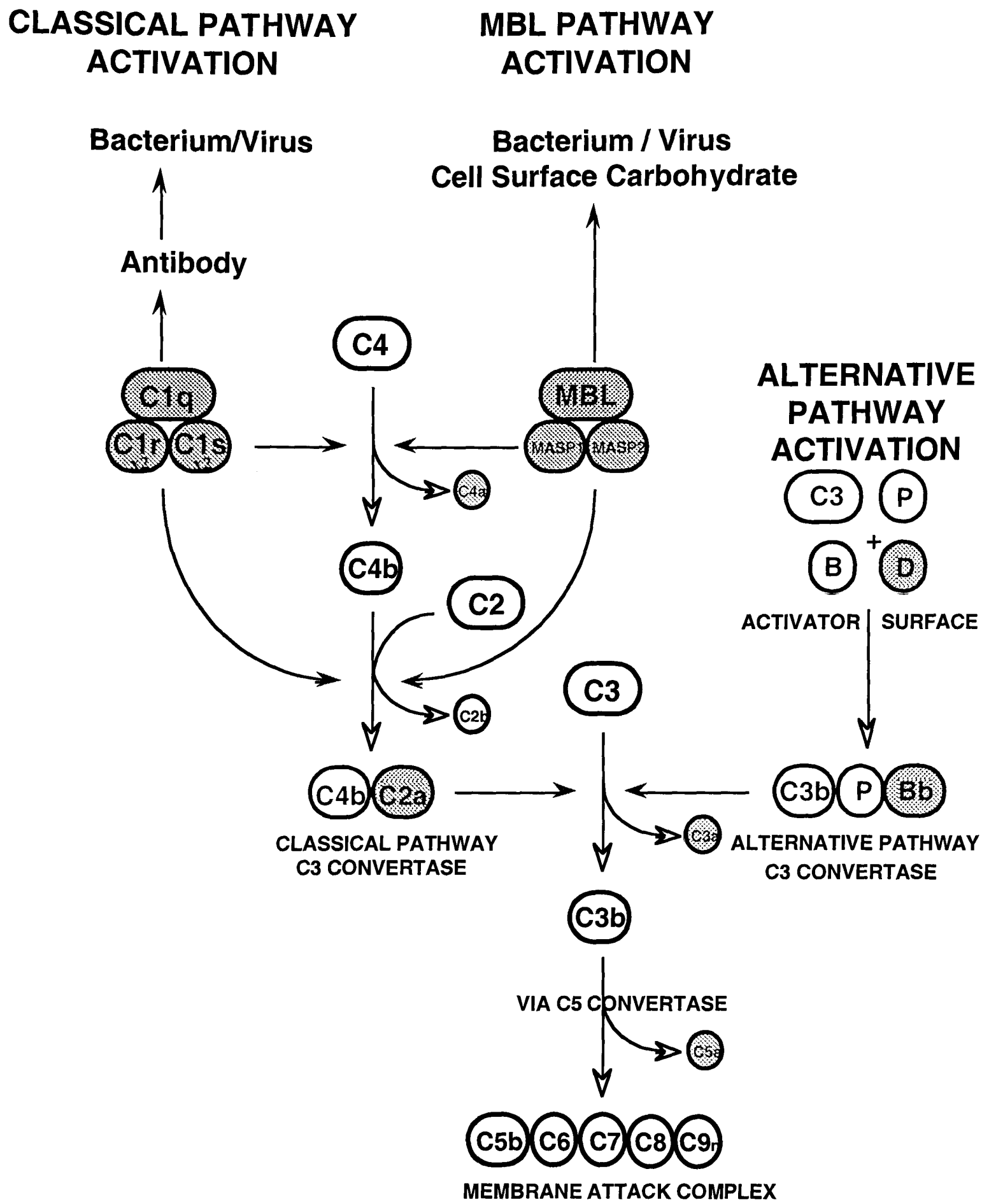


Figure 1.13. The Complement system. The three activation pathways of complement system are shown. The activators are positioned at the top of the figure; immune complexes for the classical pathway and complex carbohydrate for the MBL pathway. Arrows indicate enzymatic cleavage. The anaphylatoxins are shown; C3a, C4a and C5a as shaded small circles. Complement components with serine protease activity are indicated by large shaded circles.

form identical convertases. The pathways differ only in the recognition molecules used, the targets that lead to activation and possibly in the action of their associated serine proteases. For discussion of the alternative pathway see recent review articles (Schwaeble and Reid, 1999; Reid and Law, 1995)

1.4.1. The Classical Pathway

The key recognition protein of the classical pathway is C1q, which primarily recognises antigen-aggregated IgM or subclasses of IgG antibody. The classical pathway can also be activated by a non-antibody dependant mechanism through the interaction of C1 complex ($C1qC1r_2C1s_2$) with C-reactive protein and repeating anionic clusters present on bacterial lipopolysaccharides, a selection of viruses and Gram positive and Gram negative bacteria (Loos and Colomb, 1993).

C1 is a complex of three different components; C1q, C1r and C1s in the stoichiometry of 1:2:2. C1q is the recognition molecule that interacts with classical pathway activators. The C1q molecule is a heterooligomer consisting of 18 polypeptide chains and has a molecular weight of approximately 460 kDa; composed of six A chains, six B chains and six C chains. The three different chains are encoded by three separate genes which probably arose through gene duplication events and are found in a gene cluster of 19 kb (Sellar *et al.*, 1991; Petry *et al.*, 1996). The structure of the collectins resemble that of C1q. Each of the polypeptide chains is composed of an amino-terminal region, a collagenous stretch of sequence and a carboxyl-terminal globular region. The three polypeptide chains trimerise to form collagenous stalks that are kinked due to a skip in the collagen-like sequence (see section 1.3.2.3). A and B chains form interchain disulphides and non-covalent associations with C chains, while the C chains are covalently linked together stabilising the formation of higher order multimers of trimers. Six trimers associate to form the complete molecular assembly of C1q, typically referred to as the arrangement of a 'bunch of tulips'. C1r and Cls form an

alternating complex (C1s-C1r-C1r-C1s) that is Ca^{2+} dependant and associates in a 'figure of eight' structure around C1q to form the C1 complex (Arlaud *et al.*, 1990).

After recognition of a classical pathway activator by C1q, such as clustered IgG, complement activation proceeds by the enzymatic activity of the C1q-associated serine proteases, C1r and C1s (Schumaker *et al.*, 1987). The serine protease activity of C1r is an autocatalytic mechanism and activated C1r proceeds to activate proenzyme C1s by cleavage. C1s then cleaves one of the three chains in C4, liberating the small fragment C4a. Along with two other fragments of complement activation, C3a and C5a, C4a is an anaphylatoxin (see section 1.4.3). The cleavage of C4 leaves the large fragment C4b with an exposed thioester bond that is susceptible to nucleophilic attack from either hydroxyl or amino groups on adjacent surfaces. C4b can either attach to an activating surface or remain in fluid phase and bind C2. C2, like C4 is cleaved by C1s, forming the classical pathway C3 convertase, C4bC2a.

The thioester acts as a mechanism of attaching and concentrating the C3-cleaving complex onto activating surfaces. C3 is the central component of complement activation. Cleavage of C3 not only liberates a small fragment which is the second complement anaphylatoxin, C3a, but generates C3b. C3b can either act as an opsonin or form the alternative pathway C3 convertase or the classical pathway C5 convertase in association with C4bC2a. Cleavage of C5 yields the third complement anaphylatoxin, C5a. Although C5 is structurally related to C4 and C3, it does not contain an thioester. Instead the large fragment of C5 cleavage, C5b remains in liquid phase and is the first component of the terminal pathway then generates the self-assembling the membrane attack complex.

1.4.2. The MBL Activating Pathway

MBL is present in mammals, birds (Oka *et al.*, 1985; Laursen *et al.*, 1995) and distant homologue appear in ascidians (Nonaka and Azumi, 1999; Ji *et al.*, 1997). The MBL

activation pathway is distinct from both the alternative and classical pathways (Matsushita, 1996), although activation through the MBL pathway shares similarities with the classical pathway. The MBL activation pathway proceeds through MBL recognition of carbohydrate structures on targets. MBL circulates in the blood as a complex with the two associated serine proteases MASP1, MASP2 (Thiel *et al.*, 1997) and possibly an alternative splice variant of MASP2 (see section 1.4.2.3). *In vitro*, MBL is also able to activate the classical pathway through C1r and C1s in the absence of C1q (Lu *et al.*, 1990), but these are probably not the physiologically associated proteases. Upon MBL binding carbohydrate, MASP1 and MASP2 are each cleaved to form a heavy chain and a light chain leading to the cleavage of C4, C2 and possibly C3. MBL and the MBL activation pathway of complement are believed to be especially important part of the innate immunity for the immunocompromised individual and throughout childhood (Summerfield *et al.*, 1997).

MBL recognises certain simple carbohydrates such as mannose or *N*-acetylglucosamine in a calcium-dependent fashion (see Table 1.1). Consequently the activators of the MBL activation pathway include a wide spectrum of carbohydrates present on many different microbes (Turner, 1996). MBL recognises these activators through its multiple carbohydrate binding domains. Although, the requirements for simple disaccharide binding are known, the detailed molecular structure of a native activator has not been determined. The binding of MBL to an activating surface is an essential step in activating MASP1 and MASP2. The mechanistic pathway of activation through and by the MASPs is less well understood than the classical pathway through C1q and its associated serine proteases (C1r and C1s). Both pathways lead to a common outcome, but the steps appear different and are as yet not resolved for MBL-MASP. The MBL activation of complement and the classical pathways do converge at the point of C4. Activation of the MASPs can lead to cleavage of C4 and C2 in an analogous fashion to C1, leading to the formation of the same C3 and consequently C5

convertases as formed by the classical pathway (C4bC2a and C3bC4bC2a, respectively).

1.4.2.1. Activators of the MBL Activation Pathway of Complement

The potential list of MBL activators of the complement pathway is large, hence MBL can be considered as a pattern recognition molecule for carbohydrate structures (Medzhitov and Janeway, 1997). The potential activators of the MBL activation pathway include bacteria; *Salmonella montevideo*, *Listeria monocytogenes* and *Neisseria Spp* (Kuhlman *et al.*, 1989; van Emmerik *et al.*, 1994), viruses; influenza A, both HIV-1 and HIV-2 (Hartshorn *et al.*, 1993; Haurum *et al.*, 1993) and yeast; *Candida albicans* (Tabona *et al.*, 1995) and *Cryptococcus neoformans* (Schelenz *et al.*, 1995).

1.4.2.2. The Modular Organisation of the MASPs

Activation of the MBL and C1q associated proteases occurs by the cleavage of an Arg-Ile bond close to the serine protease domain, splitting the protein into two disulphide linked chains; the heavy and light chains (Matsushita and Fujita, 1992; Thiel *et al.*, 1997). The MASPs, C1r and C1s all have the same organisation of six protein modules (Matsushita and Fujita, 1992; Takada *et al.*, 1993; Sato *et al.*, 1994; Thiel *et al.*, 1997). The heavy chain contains two C1r/C1s/sea urchin uefg/bone morphogenic protein (CUB) domains separated by an epidermal growth factor domain, followed by two complement control protein modules. The light chain is the last module; a serine protease domain, which contains the triad of conserved amino acids that form the active site (His, Asp and Ser).

1.4.2.3. The MBL-MASP complex

Human MASP1 and MASP2 are single chain polypeptide pro-enzymes of 94 kDa and 76 kDa, respectively. MASP1 and MASP2 are found in a complex with MBL in serum (Matsushita *et al.*, 1995; Matsushita and Fujita, 1995). The calcium dependent

association of MBL and the MASPs appears controversial (Kawakami *et al.*, 1984) unlike the calcium dependence of the C1 complex. A third component found associated with MBL is MAp19, an alternative splice variant of the human, mouse and rat MASP2 genes (Stover *et al.*, 1999; Takahashi *et al.*, 1999; Stover *et al.*, 1999) consisting of the first two MASP2 domains (CUB and the EGF-like domains). The function is unknown but could be envisaged as a control protein since it forms complexes with MBL yet lacks the serine protease domain.

1.4.2.4. The Chromosomal Localisation of the MASPs, C1r and Cls

The MASP1 gene is located on the long arm of human chromosome 3, 3q27-28, on a different chromosome to the clustered C1s and C1r genes, 12p13 and MASP2 has recently been localised to chromosome 1, 1p36.3-p36.2 (Takada *et al.*, 1995; Nguyen *et al.*, 1988; Stover *et al.*, 1999). The serine protease domain of MASP1 is encoded by six exons while the light chains of C1r, C1s (Tosi *et al.*, 1989) and the recently described MASP2 serine protease domain (Stover *et al.*, 1999) are each encoded by single exons. In Chapter 5 of this thesis, the exon structure of the serine protease domain of murine MASP2 has been determined and the murine genes for the MASPs, C1r and C1s have been genetically mapped. Chapter 5 also includes a discussion on the evolution of this family of molecules within the context of the exon structure and the genetic mapping results.

1.4.2.5. The Activities of the MASPs

The proteolytic cleavages of the complement proteins ascribed to MASP1 in the literature predating the discovery of human MASP2 (Thiel *et al.*, 1997), have to be observed with a degree of caution, because of the possibility of MASP2 contaminating preparations of MASP1. Initial reports describe MASP1 acting in a similar capacity as C1s (see section 1.4.1), cleaving both C4 and C2 (Matsushita and Fujita, 1992), however recent studies suggest C4 is cleaved by MASP2 not MASP1 (Thiel *et al.*, 1997).

In contrast to both C1r and C1s, human MASP (probably MASP1) has been reported to cleave C3, generating the fragments C3a and C3b (Matsushita and Fujita, 1995), thus by-passing the C3 convertase. While the cleavage of C3 by MASP2 (Thiel *et al.*, 1997) required the presence of other serum factors, as pure C3 could not be cleaved directly by MASP2.

The MASP-MBL complex has been previously characterised in mice as Ra-reactive factor (RaRF). It was identified as a serum bactericidal factor that activates C4 and C2 (Ji *et al.*, 1993) by binding core polysaccharides (Ra and R2) that are found on many different strains of enterobacteria (*Salmonella tyhimurium* and *E. coli*) (Ihara *et al.*, 1982; Kawakami *et al.*, 1984). MASP1 was identified in mice as a 100 kDa component of RaRF, composed of two disulphide polypeptide chains of 70 and 29 kDa, also termed P100 (Takada *et al.*, 1993; Takayama *et al.*, 1994; Ji *et al.*, 1993). The mouse MBL-MASP complex (RaRF) has been shown to cleave C4, in fact with approxiamately 100 fold greater activity than mouse C1s (Ogata *et al.*, 1995). Mouse MBL-MASP complex (RaRF), probably containing both MASP1 and MASP2, has been shown to cleave C4 and C2 as well as C3 (Ji *et al.*, 1993; Ogata *et al.*, 1995).

The control of MASP activation is probably regulated in a similar fashion to the C1q assosciated serine proteases. MASP1 can form a covalent complex with C1 inhibitor, like C1s and may also be controlled by interacting with α_2-macroglobulin (Terai *et al.*, 1995; Storgaard *et al.*, 1995).

The components of the C1 complex are present in the serum at the ratio of 1:2:2 (C1q:C1r:C1s), where as MASP1 levels exceed MBL levels and the majority of MASP1 circulates as a free form within the blood (Terai *et al.*, 1997). The catalytic events controlling MASP1 and MASP2 activation do not appear to mirror C1r and C1s, since MASP2 has been reported to have autocatalytic activity. The possibility that

MASP1 and MASP2 act in an analogous fashion to C1r and Cls (Thiel *et al.*, 1997) appears unlikely at present.

1.4.2.6. Evolution and Protein Families

MBL does not rely upon an exquisite repertoire of control proteins, such as for the alternative pathway of complement and use of the adaptive immune molecules for recognition of targets, such as for the classical pathway (section 1.4.1). Therefore in terms of the simpler systems evolving before more complex systems, the MBL pathway may be considered to predate both the classical and alternative pathways in the evolution of the complement system.

The MBL and C1q associated serine proteases (MASP1, MASP2, C1r and C1s) form a distinct family of serine proteases (Thiel *et al.*, 1997; Matsushita *et al.*, 1998). Phylogenetic analysis of the cDNA for the MASPs from vertebrates and ascidians and C1r and C1s from humans, reveals that this family of proteins forms two groups (Ji *et al.*, 1993; Ji *et al.*, 1997). The criteria for the two groups was based upon the functional and structural features of the serine protease domain. The first group of serine proteases have an active site serine that is encoded by a TCN codon, with a loop formed by an extra disulphide bond that contains the active site histidine while the residue three amino acids carboxyl-terminal to the active site serine is proline. The second group has an active site serine encoded by AGY[1] codon, lack a histidine loop and the residue three amino acids carboxyl-terminal to the active site serine is either a valine or an alanine.

Group I contains MASP1 (from humans, mice and frogs) and both ascidian MASPs. Group II contains both C1s and C1r (from humans and mice), MASP2 (from humans, mice, frogs) and the only known MASP from carp and sharks (Endo *et al.*, 1998). It is believed that the group I, TCN-type MASPs arose earlier in evolution based upon the

[1]Where Y stands for either T or C nucleotides.

presence of only MASP1 like proteins in ascidian (Endo *et al.*, 1998). This is discussed further Chapter 5 of this thesis, with experimental support from chromosomal localisation data and the partial murine gene structure of MASP2.

Ascidians have been shown to have homologues of C3, MBL and two MASPs and have the ability to cleave C3 (Ji *et al.*, 1997). These ascidian components could form a simple complement system that could act as a primitive opsonic mechanism (Nonaka and Azumi, 1999). This primitive lectin pathway may act via the MBL-like molecule to recognise carbohydrate targets on pathogens, leading to the cleavage of C3 through the action of the MASPs to form a C3 derived opsonin.

1.4.3. Anaphylatoxin Receptors

Two receptors, the C3a and C5a receptors (C3aR and C5aR) are associated with the complement system and respond to cleavage fragments of the complement cascade generated during activation, named anaphylatoxins. This family of G protein-coupled, seven-transmembrane spanning proteins acts as chemoattractant receptors. The interaction of the complement receptors with C3a and C5a is essential for eliciting various inflammatory reactions mediated by complement, such as the induction of smooth muscle contraction, increasing vascular permeability, directing the migration of phagocytic cells and inducing histamine release. These in turn enhance the local immune defence against pathogens, as emphasised by the findings from the C5aR knockout mice, which is unable to clear *Pseudomonas aeruginosa* from the lungs (Hopken *et al.*, 1996; Hopken *et al.*, 1997).

1.4.4. Gene Clusters

The complement gene localisations carried out to date, in humans and mice, indicate that there are three major gene clusters which account for many of the 40 genes that encode for complement system proteins which share the same protein modules. In addition, there are three clusters that represent simple gene duplications. First, the regulator of

complement activation (RCA) gene cluster forms a linkage group for many of the complement proteins that control the activation and degradation of C3 and C4 (Heine Suner *et al.*, 1997). Second, some of the key early components of complement activation are found in the major histocompatability complex class III region, which includes two of the serine proteases of the classical and alternative pathway, C2 and factor B, together with the two isotypes of C4, C4A and C4B (Carroll *et al.*, 1984). The third cluster constitutes some of the complement components of the membrane attack complex; C6, C7 and C9 (Coto *et al.*, 1991). The three other complement gene arrays represent simpler gene duplications; *C1S* and *C1R* (Kusumoto *et al.*, 1988), two of the genes which encode C8; *C8A* and *C8B* (Rogde *et al.*, 1986) and the three genes which encode for the A, B and C polypeptide chains of C1q (Sellar *et al.*, 1991).

1.5. Gene Mapping

The current approaches for mapping genes in mice can be divided into three groups; chromosomal, physical and genetic mapping. A diagrammatic comparison of these methods can be seen in Figure 1.14.

1.5.1. Chromosomal Mapping

A chromosomal map or cytogenetic map defines chromosomes on the basis of the size and pattern of chromosomal banding. Indirect cytogenetic mapping of genes can be performed with somatic cell hybrid lines which carry previously cytogenetically defined fragments of chromosomes contained within the genome of another species. This method provides chromosome or sub-chromosome assignments for genes and was used to map the porcine SP-A and SP-D genes in this thesis (Chapter 4).
Chromosomal mapping can be conducted in special cases where a chromosomal aberration leads to a mutant phenotype, the mutant loci or gene can then be assigned to the disrupted chromosomal band.

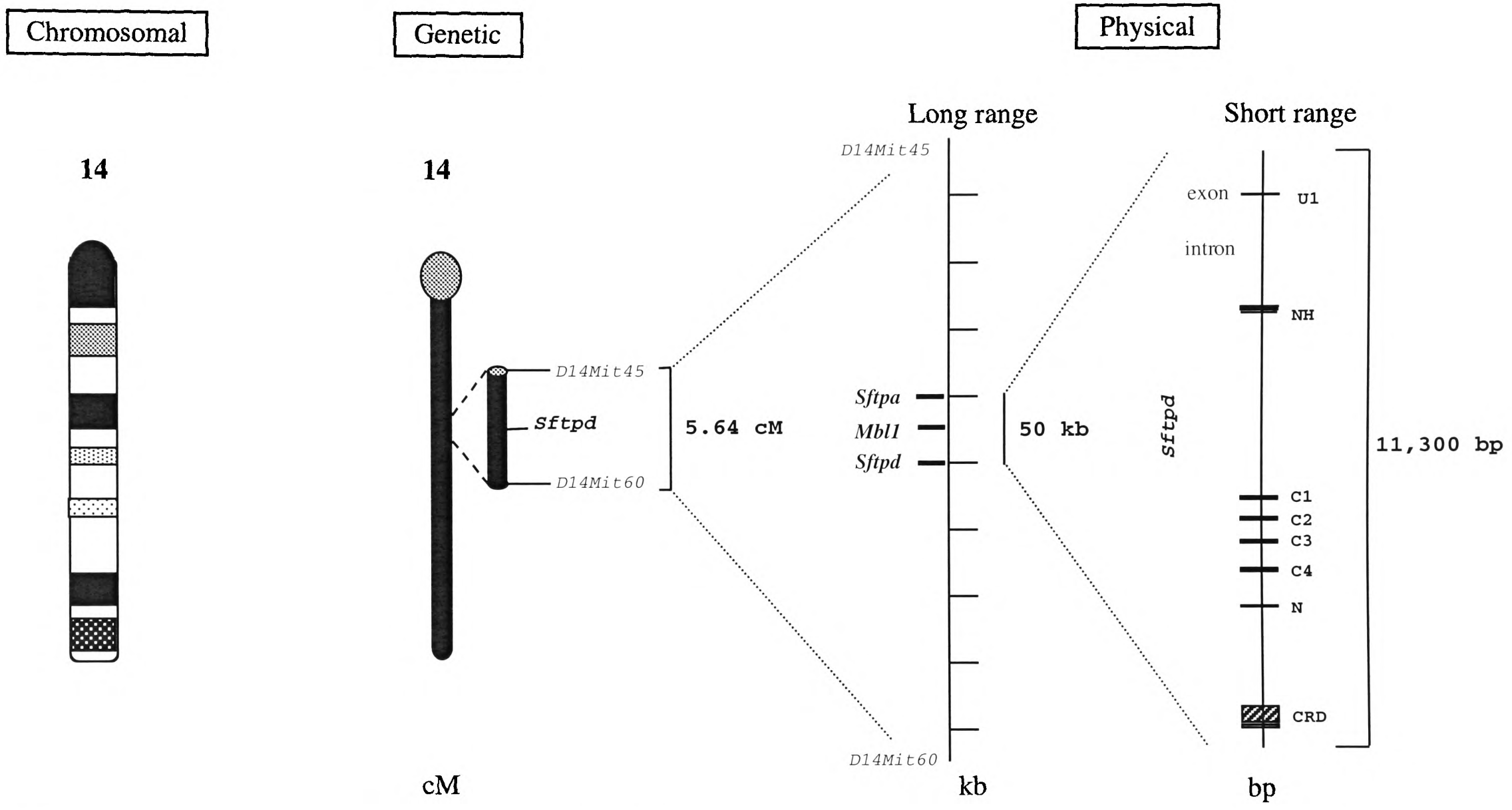


Figure 1.14. Comparison of chromosomal, genetic and physical maps. The three types of maps are shown for loci on murine chromosome 14. The chromosomal and genetic maps extend over the entire chromosome with dashed lines connecting the approximate positions of loci mapped in both. The example of murine SP-D is illustrated here. See Chapter 3 of this thesis and Akiyama *et al.*, 1999 for further details.

Chromosomal mapping of a loci is usually performed by FISH. Gene localisation by FISH provides a cytogenetic indication of the location of the loci relative to a chromosomal band. FISH has the disadvantage of requiring both a high level of specialised training and expertise in addition to specialised equipment and facilities, so it is usually conducted in collaboration with laboratories who carry out the procedure on a routine basis.

All chromosomal assignments made, whether genetic or physical mapping are based upon loci that have been assigned by cytogenetic analysis. However, the resolution achieved by chromosomal mapping is not as high as linkage or physical mapping.

1.5.2. Physical Mapping

Physical mapping involves the direct analysis of DNA. Distances between loci and within genes are measured in base pairs or multiples of. Long range mapping requires large fragments of cloned DNA in Yeast Artificial Chromosomes (YACs) or P1 Artificial Chromosomes (PACs). This type of mapping is reliant upon the use of rare cutting restriction endonucleases and a method of electrophoresis for large fragments of DNA; pulsed field electrophoresis. Short range mapping, typically less than 30 kb, is based upon cloned inserts obtained from cosmids and bacteriophage.

1.5.3. The Genetic Approach for Gene Mapping

The genetic approach for gene mapping relies upon backcross analysis of inbred mice. The inheritance of parental alleles of the gene across the backcross panel is determined to assign the gene a chromosomal location based on the comparison with other previously mapped markers. The distances between markers and the centromere are recombination distances; centimorgans (cM). This requires both finding a restriction length polymorphism near the gene of interest and sufficient backcross genomic DNA for use in southern blotting. Alternatively, backcross analysis can be performed with smaller amounts of genomic DNA by discovering a simple sequence length

polymorphism (SSLP) associated with the gene that can be amplified by PCR, such as those described for murine SP-D in this study (section 3.9). Both these approaches to gene mapping require the isolation of the gene or a fragment of the gene of interest. Moreover, the approach must include either detailed sequence analysis to discover an associated sequence length polymorphisms or require large amounts of backcross DNA for southern blotting once a restriction polymorphism is discovered.

1.5.4. The Use of ESTs in Gene Mapping

Expressed sequence tags (ESTs) have formed a vital part of the human and murine genome mapping project (Marra *et al.*, 1999). EST projects reveal small sequence identifiers for expressed genes within the genome. The very nature of the ESTs libraries means that statistically, the most abundantly transcribed genes are sequenced first. ESTs are a valuable tool in physical mapping, providing an expressed gene content to anonymous or specifically cloned segments of the genome, such as YACs and PACs. They can also be mapped directly onto the genome using FISH. Chapter 5 describes a novel approach that has been developed to map genes rapidly and at high resolution using EST information.

1.6. Aims of the project

The murine gene for lung surfactant protein D was cloned to objectively determine the role of SP-D. There were two hypotheses for the *in vitro* role of SP-D. Firstly, SP-D functioned as an innate immune protein by binding carbohydrate ligands on pathogens and interacting with immune cells. Secondly, SP-D bound lipids of the surfactant, possibly participating in surfactant homeostasis. By cloning the gene for SP-D, transgenic mice could be generated that lacked SP-D and used to objectively understand its role *in vivo*. The SP-D gene was also cloned since the environmental and genetic factors that control its expression were poorly understood. The promoter region could be analysed in comparison to the human SP-D gene and the conglutinin gene (see Chapter 3).

To analyse the expression of SP-A and SP-D in both normal and transgenic mice, specific antibodies would be required. Therefore, fragments of murine SP-A and SP-D were cloned to generate purified recombinant protein in order to raise antibodies against these surfactant components.

During the course of the project, two independent groups generated transgenic mice that lacked SP-D. Consequently, the course of the project changed. The murine collectin loci, consisting of the genes for SP-A, SP-D and MBL-A were mapped on a high resolution genetic map. This was conducted to determine both the size of the collectin loci and to establish if mice lacking individual collectin genes could be crossed to generate double knockouts. This is discussed further in Chapter 3.

The phylogenetic analysis and cloning of a distant species homologues of other innate immune proteins has contributed towards the understanding of these proteins. In light of this, the presence of SP-D in other species was examined, using hybridisation and PCR cloning (see Chapter 4). Leading on from this porcine SP-D was cloned. The porcine homologue for SP-D was also cloned because of the medical relevance of porcine surfactant. Collaborators (Prof. H. Haagsman, Utrecht, The Netherlands) had also identified that porcine SP-D showed pronounced difference to SP-D from other species.

At the beginning of this study a number of murine complement components had not been chromosomally mapped because their genes had not been isolated. However, murine EST information existed for these components. Genetic mapping of the murine collectins and examination of murine EST databases, led to the realisation that murine genes could be mapped with only EST information. Chapter 5 describes the development of a novel approach to murine gene mapping, using EST information, intron prediction and restriction fragment length polymorphisms fragments of genes

amplified by PCR assay. The approach I developed was applied to the murine complement component genes in Chapter 5.

CHAPTER 2: MATERIALS AND METHODS

2.1. Materials

2.1.1. Enzyme and Chemicals Suppliers

Chemicals were obtained from BDH (Merck Ltd, Poole, England), Amersham Pharmacia Biotech (Uppsala, Sweden) or Sigma-Aldrich Company Ltd (Poole, England), unless otherwise stated. Restriction enzymes were obtained from New England Biolabs (Hitchin, Hertfordshire, UK), Stratagene (La Jolla, California, USA), Amersham Pharmacia Biotech and Promega (Southampton, England).

2.1.2. Buffers and Solutions

All buffers and solutions were stored at room temperature, unless otherwise stated. All widely used or complex buffers are listed below, for others the recipe for the buffer is given within the method.

Alkaline lysis solution	0.2 N NaOH, 1% (w/v) SDS. Freshly made.
$CaCl_2$ solution (for transformations)	60 mM $CaCl_2$, 15% (v/v) glycerol, 10 mM PIPES [piperazine-*N*,*N*'-bis(2-hydroxypropanesulfonic acid)], pH 7.0, autoclaved.
Denhardt's solution, 50 ×	1% (w/v) Ficoll 400, 1% (w/v) polyvinylpyrrolidone, 1% (w/v) BSA. Filter sterilised and stored at -20°C.
Glycerol loading dye	30% (v/v) glycerol, 0.1% (w/v) bromophenol blue, 0.1% (w/v) Xylene cyanol FF, 250 mM EDTA.
GTE	50 mM glucose, 25 mM Tris-Cl, pH 8.0, 10 mM EDTA, autoclaved.
Hybridisation buffer	50% (v/v) deionised formamide, 50 mM Tris-Cl, pH 7.4, 1 M NaCl, 0.1% (w/v) sodium pyrophosphate, 5 × Denhardts solution, 10% (w/v) Dextran sulphate, 0.1% (w/v) SDS. Stored at 4°C.
Phosphate buffered saline (PBS)	137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4 \cdot 7H_2O$ and 1.4 mM K_2HPO_4.
Protein sample buffer	200 mM Tris-Cl, 8 M urea, 2% (w/v) SDS, 10% (v/v) glycerol, 0.1% (w/v) bromophenol blue.
Ribonuclease A (RNase A) solution	10 mg/ml RNase A in 10 mM Tris-Cl, pH 7.4, 15 mM NaCl. Heated at 100°C for 15 minutes. Stored at -20°C.

SM buffer	50 mM Tris-Cl, pH 7.4, 0.01% (w/v) gelatine, 100 mM NaCl, 8 mM $MgSO_4 \cdot 7H_20$, autoclaved.
SSC buffer, 20 ×	3 M NaCl, 300 mM tri-sodium citrate, pH 7.0
TAE buffer, 50 ×	2M Tris-acetate, 100 mM EDTA, pH 8.5.
TE buffer	10 mM Tris-Cl, 0.1 mM EDTA, pH 7.4.
Tissue digestion buffer	100 mM NaCl, 10 mM Tris-Cl, pH 9, 25 mM EDTA, pH 8, 0.5% (w/v) SDS, 0.1 mg/ml proteinase K (added on use).
Tris buffered saline (TBS)	100 mM Tris-Cl, pH 7.4, 150 mM NaCl.
Tris-Cl [tris(hydroxymethyl) aminomethane chloride], 1M	1 M Tris base, adjusted to pH 8.0 or pH 7.4 with HCl at working temperature.

2.1.3. Media and Antibiotics

Media were autoclaved for 25 minutes. Antibiotics (10 mg/ml ampicillin, filter sterilised) were added to the media below 50°C. Plates were stored at 4°C, while media were stored at room temperature for convenience. Lambda top agarose was used in preference to lambda top agar when performing plaque lifts and preparing DNA because of the less glutinous nature and fewer impurities in agarose.

LA plates	As Lambda plates with 50 μg/ml ampicillin.
Lambda broth	1% (w/v) Bacto-tryptone, 0.8% (w/v) NaCl, with 0.7% (w/v) agarose.
Lambda plates	1% (w/v) Bacto-tryptone, 0.25% (w/v) NaCl, 1% (w/v) agar, pH 7.4.
Lambda top agar	1% (w/v) Bacto-tryptone, 0.8% (w/v) NaCl, with 0.7% (w/v) agar.
Lambda top agarose	1% (w/v) Bacto-tryptone, 0.8% (w/v) NaCl, with 0.6% (w/v) agarose.
LB medium	1% (w/v) Bacto-tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, pH 7.4.
LB plates	As Lambda broth with 1.5% (w/v) agar.
NZC broth	1% (w/v) NZ amine A, 0.5% (w/v) NaCl, 0.2% (w/v) $MgCl_2 \cdot 6H_2O$. Autoclaved and made to 0.1% (w/v) casamino acids (filter sterilised).

2.1.4. Bacterial Strains

Strain	*Supplier*	*Use*
DH5α	Clontech (Basingstoke, Hampshire, England)	Cloning and storage.
JM110	Stratagene	Cloning. Used to prepare non methylated DNA for *Xba* I and *Cla* I restriction digests.
TG-1	Stratagene	Single-stranded DNA rescue.
TOP10	Invitrogen (Groningen, The Netherlands)	Cloning. Commercially available competent cells.
XL-1 Blue	Stratagene	λFIX-II propagation, single-stranded DNA rescue.
Y1090r⁻	Clontech	λgt11 propagation and screening.

2.1.5. Vectors

Vector	*Supplier*	*Use*
pCRII-TOPO	Invitrogen	Cloning PCR products (section 2.8.4).
PMOS	Amersham Pharmacia Biotech	Cloning PCR products (section 2.8.4).
pBluescript	Stratagene	General purpose cloning.
λgt11	Clontech	Contains porcine cDNA library.
λFIX-II	Stratagene	Contains mouse genomic library.

2.2. Recombinant DNA and Cloning Methods

The majority of molecular biology methods were adapted from two laboratory manuals and these should be referred to for extensive technical support (Sambrook *et al.*, 1989; Ausubel *et al.*, 1998).

2.2.1. Bacterial Manipulation and Storage

2.2.1.1. Growth of Bacteria in Liquid Media

A single bacterial colony from a freshly streaked out plate was used as an inoculant for a 10 ml culture of liquid medium in a 50 ml sterile disposable tube. This was grown overnight or for 6 hours at 37°C, with orbital shaking (250 rpm) and is termed an

overnight culture. For larger cultures 200 ml of medium was inoculated with 2 ml of an overnight culture (1:200 dilution). To ensure adequate aeration, bacteria were grown in a baffled flask with a volume that exceeded the media volume tenfold, that is, 2 L.

2.2.1.2. Growth of Bacteria on Solid Media

For isolating single colonies, bacteria from either frozen stocks, cultures or solid media were streaked out on an agar plate containing the appropriate antibiotic and grown overnight in an inverted position at 37°C.

2.2.1.3. Storage of Bacterial Stocks

Freshly streaked out *Escherishia coli* from frozen stocks were stored on solid media at 4°C for two to four weeks. Long term storage of bacteria was performed by centrifuging 1.5 ml of an overnight culture at 500 × *g* in a 1.5 ml tube and resuspending the pellet in 1 ml of LB medium containing 30% (v/v) glycerol. The stocks were then stored at -20°C or -70°C. To revive the stock splinters of ice containing the bacteria were streaked out onto an agar plate.

2.2.2. Plasmid DNA Preparation

2.2.2.1. Minipreps of Plasmid DNA

Small-scale preparations of DNA were prepared by a modified alkaline lysis procedure (Birnboim and Doly, 1979). A 20 second, 10,000 × *g* bacterial pellet from 1.5 ml of an overnight culture was resuspended in 100 µl of GTE solution. Bacteria were lysed by adding 200 µl of freshly prepared alkaline lysis solution and mixed gently by inverting twice, followed by a 5 minute incubation on ice. The mixture was neutralised with 150 µl of 5 M potassium acetate solution, pH 4.8 and placed on ice for 5 minutes. The mixture was centrifuged for 4 minutes at 10,000 × *g* leaving a pellet of bacterial protein and chromosomal DNA and the plasmid DNA in solution. The supernatant was transferred to a clean tube and ethanol precipitated (section 2.2.5.2). The pellet was

resuspended in 30 µl of TE buffer and stored at -20°C. Contaminating bacterial RNA can be degraded by adding 1 µl of RNase A solution to a restriction digest.

Alternatively, for speed and convenience, plasmid DNA was prepared by a commercially available kit, QIAprep, from Qiagen Ltd (Crawley, West Sussex) based around the alkaline lysis method. This kit was also used to prepare plasmid DNA for sequencing, except the plasmid DNA was eluted from the column with deionised water.

2.2.2.2. Large Scale Preparation of Plasmid DNA

A crude lysate was prepared by alkaline lysis. A saturated bacterial culture ($OD_{600} \approx 4$) of 500 ml was pelleted at 6,000 × *g* for 10 minutes, resuspended in 4 ml of GTE solution and transferred to a 20 ml centrifuge tube. To increase plasmid recovery 1 ml of 25 mg/ml hen egg white lysozyme in GTE solution was mixed with the bacterial suspension and left at room temperature for 10 minutes. Bacterial lysis was performed by adding of 10 ml of alkaline lysis solution for 10 minutes on ice and neutralised for 10 minutes on ice after adding 7.5 ml of 3 M potassium acetate solution, pH 4.8. Chromosomal DNA, SDS-protein complexes and cellular debris were removed by centrifugation for 10 minutes at 20,000 × *g* at 4°C. The plasmid DNA was precipitated from the supernatant with 0.6 volumes of isopropanol (section 2.2.5.2). The DNA pellet could be stored at 4°C before proceeding further.

Caesium chloride/ethidium bromide equilibrium centrifugation was used to purify large plasmids (> 10 kb) to a high quality. The pellet from the crude lysate was dissolved in 4 ml of TE buffer, to which was added, 4.4 g of CsCl and 0.4 ml of 10 mg/ml ethidium bromide solution. This mixture was transferred to a 5 ml ultracentrifuge tube heat-sealed and centrifuged for 16 hours at 350,000 × *g*, 20°C. This procedure results in two discrete bands, a wide plasmid DNA band appears below a narrow chromosomal DNA band approximately halfway down the tube. A third diffuse band below these is RNA. Plasmid DNA can be recovered by inserting a 20G syringe needle at the top of the tube

to provide an inlet for air, while collecting the plasmid DNA with a second needle that is inserted below the plasmid DNA band.

Ethidium bromide was removed by repeated extraction of the plasmid DNA preparation with an equal volume of water saturated butan-1-ol, removing the upper organic phase, until no red colour remains. The plasmid-CsCl preparation was diluted three-fold with TE buffer so that the DNA could be ethanol precipitated (section 2.2.5.2). The pellet was dissolved in 1 ml of TE buffer and stored at -20°C.

Alternatively for preparing smaller plasmids (< 10 kb) a commercially available kit was used (QIAprep® Miniprep, Qiagen). Based upon selective binding of plasmid DNA to silica gel.

2.2.3. Introducing Plasmid DNA into *E. coli*

2.2.3.1. Preparation of Competent Cells for Transformation

Competent cells were prepared by the calcium chloride method (Cohen *et al.*, 1972). *E. coli* were grown to log phase by using 4 ml of an overnight culture to inoculate a large-scale (400 ml) culture. All reagents and tubes were pre-cooled and procedures were carried out at 4°C. This was harvested at an $OD_{600} \approx 0.375$ by centrifuging for 10 minutes at 1,600 × *g* , 4°C (no brake) and then resuspended gently in 10 ml of ice-cold $CaCl_2$ solution. The cells were pelleted and resuspended a second time and incubated for 30 minutes on ice, followed by another centrifugation and resuspended step before being dispensed into 250 µl aliquots and either used immediately or stored by rapidly freezing to -70°C in liquid nitrogen.

Alternatively for convenience, TOP10 cells (Invitrogen), a commercial source of competent cells was used. The manufacturer's instructions were followed for transformation conditions.

2.2.3.2. Transforming Competent *E. coli* with Plasmid DNA

Generally 10 ng of plasmid DNA or less than 1 μg of DNA from a ligation mixture (section 2.3.4) in a volume of 10-25 μl was chilled on ice for 10 minutes. Cold competent cells, 100 μl, were gently mixed with DNA and left on ice for 10 minutes. The cells were then exposed to a heat shock for 2 minutes at 42°C. The cells were allowed to recover by adding 1 ml of preheated LB medium and incubated at 37°C for 40 minutes with shaking (250 rpm). A 100 μl aliquot of the transformation culture was plated onto LA plates. The plates were dried and incubated overnight at 37°C.

2.2.4. Preparation of Lambda Bacteriophage DNA

2.2.4.1. Making Bacteriophage Lysates

Two different techniques were employed to isolate lambda-derived DNA dependent on the amount required.

2.2.4.2 Plate Lysates

Lambda-sensitive strains of *E. coli* (Y1090r⁻ or XL-1 Blue) were grown in LB medium to an $OD_{600} \approx 0.4$. To 1 ml of log phase cells a single fresh bacteriophage plaque was added, using a cut-off pipette tip. To this 7.5 ml of melted (45-50°C) top agarose was added, gently mixed and poured onto a 135 mm diameter lambda plate. The plates were incubated for 5-8 hours at 37°C until 90-95% lysis of the bacterial lawn was observed. SM buffer was added to the plate (3 ml) and the top agarose was scraped off into a 15 ml centrifuge tube. Three drops of chloroform were added to lyse remaining bacteria, vortexed and then left at room temperature for 10 minutes. The tube was centrifuged for 10 minutes at 10,000 × *g*, 4°C and the supernatant of lambda phage was stored (section 2.2.4.4).

2.2.4.3. Liquid Lysates

A single plaque was added to 0.4 ml of lambda dilution buffer (20 mM Tris-Cl, pH 8, 20 mM $MgCl_2$) and the plaque was allowed to elute for 2 hours at 4°C. For bacterial

infection, 100 µl of a saturated overnight culture (Y1090r⁻ or XL-1 Blue) was added to 100 µl of eluted phage as well as 100 µl of 10 mM $MgCl_2$ and 10 mM $CaCl_2$. This mixture was incubated for 15 minutes at 37°C. The magnesium and calcium ions are essential for phage attachment to bacteria. This solution was transferred to 50 ml of NZC medium and incubated at 37°C with shaking until lysis occurred (6 to 8 hours). The remaining cells were lysed by adding two drops of chloroform. Cell debris were removed by centrifugation (10,000 × *g*, 10 minutes, 4°C) and the supernatant of lambda phage was stored (section 2.2.4.4). For optimisation of yield, different dilutions of phages were incubated with the bacteria and the OD_{600} was monitored for 90-95% lysis after 6-8 hours.

2.2.4.4. Storage of Phage Lysates

For imminent use, the phage stocks were stored at 4°C after the addition of two or three drops of chloroform to 100 ml of lysate. For long term storage, the phage stocks were precipitated (section 2.2.4.6) and either an equal volume of SM containing 30% (v/v) glycerol was added to the phage stock before freezing at -70°C.

2.2.4.5. Checking Phage Titre

To check the titre of a phage stock, four serial 100-fold dilutions of the phage stock were made in 100 µl of SM buffer. These were added to 0.3 ml of an overnight culture in lambda broth, supplemented with 0.2% (w/v) maltose and 10 mM $MgSO_4$. The *E. coli*-bacteriophage mixtures were incubated at room temperature for 20 minutes for the phage to attach to the bacteria, followed by a 10 minute incubation at 37°C for the phage to infect the bacteria. The dilutions of phage infected bacteria were mixed with 2.5 ml of top agar, poured onto lambda plates and incubated for 12 hours at 37°C. Using a low-density plaque containing plate, the number of plaques was counted and this number was used to calculate the number of viable phage in the original phage stock.

2.2.4.6. Preparation of Lambda DNA from Phage Lysates

Phage DNA was prepared from phage stock using an ion-exchange resin to selectively bind contaminating RNA, chromosomal DNA and protein. The DNA was then extracted by organic solvents and ethanol precipitation. First, the crude phage lysate was precipitated by addition of 5.8g NaCl and 20g PEG 8000 to 200 ml of a liquid lysate (at ~1×10^{10} pfu/ml). PEG was dissolved with gentle stirring over 1 hour at 4°C. Precipitated phage particles were separated by centrifugation (3,500 × *g*, 10 minutes, 4°C). The resulting pellet was resuspended in 3 ml of TM buffer, transferred to a 15 ml Corex centrifuge tube and after the addition of 3 ml of chloroform, centrifuged (3,000 × *g* for 10 minutes, 4°C). The upper aqueous phase was saved and to reduce sample loss the chloroform was re-extracted with the addition of 3 ml of TM buffer. The aqueous phase was removed as before and pooled with the initial crude lysate extract.

The phage lysate extracted was further purified by anion exchange chromatography. A slurry of DEAE-sepharose (Amersham Pharmacia Biotech) was prepared by adding five volumes of 0.05 N HCl and neutralised with 10 M NaOH. A 9 ml ion exchange column was then poured in a 10 ml disposable syringe, equilibrated with TM buffer. The column was loaded with the crude lysate extract and eluted with an additional 10 ml of TM buffer. The first 3 ml of the eluate were discarded (void volume) and the next 13 ml were collected. The phage particles were precipitated by addition of 2 ml of 5 M NaCl and 10 ml of ice-cold isopropanol and incubated at -20°C for 15 minutes, followed by centrifugation (4,000 × *g*, 10 minutes, 4°C). The pellet was resuspended in 0.8 ml of TE buffer and extracted once with phenol, twice with phenol/chloroform/iso-amyl alcohol (section 2.2.5.1) and finally with chloroform until no visible protein precipitate was visible at the interface. The lambda DNA was then precipitated with ethanol (section 2.2.5.2), dissolved in TE buffer and stored at 4°C for subsequent use.

2.2.5. Extraction and Concentration of DNA

2.2.5.1. Organic Extraction of Protein from Nucleic Acid: Phenol Extraction

Phenol extraction was used to remove protein contaminants from DNA solutions, followed by ethanol precipitation to concentrate the DNA and a 70% (v/v) ethanol wash to remove salts and organic molecules. An equal volume of phenol/chloroform/iso-amyl alcohol (v/v 25:24:1) was added to the DNA solution and vortexed for 10 seconds. The two phases were separated by centrifugation in a microcentrifuge (10,000 × *g*, 10 seconds, RT). The upper aqueous phase was then concentrated by ethanol precipitation (section 2.2.5.2).

2.2.5.2. Ethanol Precipitation

A solution of DNA was precipitated by addition of 1/10 of the volume of 3 M sodium acetate, pH 5.2 and 2 volumes of ice-cold 100% ethanol and incubated at -70°C for at least 15 minutes. After centrifugation (10,000 × *g*, 10 minutes, room temperature) the pellet was washed with 70% (v/v) ethanol, centrifuged briefly, aspirated and dried under vacuum. The DNA pellet was dissolved in either water or TE buffer.

For larger starting volumes, DNA was precipitated with one volume of isopropanol, and incubation at room temperature. However, ethanol precipitation was routinely used as less salt remains in the pellet and ethanol evaporates more readily than isopropanol.

Alternatively, for speed, convenience and small volumes of DNA, a commercially available kit was used (Wizard DNA cleanup, Promega), following the manufacturer's instructions. This kit was based upon the absorption of DNA onto glass beads in the presence of sodium iodide, followed by the removal of contaminants by a high salt buffer and elution of DNA with low salt buffer or water.

2.2.5.3. Preparation of Genomic DNA from Mammalian Tissues

Tissue was rapidly frozen in liquid nitrogen and either used immediately or stored at

-70°C. Between 200 mg and 1 g of tissue was ground to a fine powder with a ice-cold mortar and pestle and then resuspended in 2.4 ml of tissue digestion buffer per 200 mg of tissue. Digestion was carried out overnight at 50°C to ensure the majority of the cellular protein was degraded by the proteinase K in the buffer. The protein was extracted from the sample by repeated phenol extraction (section 2.2.5.1) until the white proteinous interface disappeared and followed by ethanol precipitation (section 2.2.5.2) but using 0.5 volumes of 7.5 M ammonium acetate instead of 3M sodium acetate. Chromosomal DNA was dissolved to 1 mg/ml in TE buffer with gentle shaking over several hours at room temperature and then stored at 4°C.

Alternatively, a commercial kit was used following the manufacturer's instructions, utilising the selective binding of DNA to an anion-exchange resin (Qiagen genomic tips, Qiagen).

2.2.5.4. Isolation and Purification of DNA from Agarose Gels

After the appropriate restriction digest (section 2.2.9) or PCR assay (section 2.8), a DNA fragment from an agarose gel was purified from contaminating bands by a commercially available gel purification kit (QIAquick Gel extraction Kit, Qiagen), following the manufacturer's recommendations. Briefly, the desired band was excised from the gel (section 2.2.6.1) and the agarose was solublised with a guanidine containing buffer. In the presence of high salt concentrations the DNA selectively binds to a silica-gel membrane mounted in a spin-column, this allows contaminants to be washed off the column. The DNA is eluted with water or TE buffer (Heery *et al.*, 1990).

2.2.6. Analysis of DNA

2.2.6.1. Agarose Gel Electrophoresis

Agarose gels were prepared by dissolving agarose in TAE buffer by heating. The percentage of the agarose used in TAE buffer varied from 0.5-1.5% (w/v) for efficient

separation ranges of 1-30 kb to 0.1-3 kb. Ethidium bromide was dissolved in the gel solution (below 50°C) to a final concentration of 0.1 μg/ml. Gels were run submerged in 1 × TAE buffer, using either minigels (103 × 105 × 5 mm) or larger slab gels (200 × 250 × 5 mm) for analysis of larger sample sets or genomic DNA. Samples were loaded in 1/10 volume of 10 × glycerol loading dye. The molecular size markers used were either *Eco*R I/*Hind* III digested lambda DNA ladder (Boehringer Mannheim, Lewes, East Sussex, England), 1 kb ladder (Boehringer Mannheim) or SmartLadder (Eurogentec, Scraing, Belgium). Gels were run at 10 V/cm or for genomic DNA at 1 V/cm overnight. DNA-ethidium bromide complexes were observed under UV light (UVP Inc., Trans-Illuminator TM-20) and were photographed directly using a Cybertech CS1 video system and a Mitsubishi video copy processor.

2.2.6.2. High-Resolution Denaturing Gel Electrophoresis

Sequencing reactions, PCR products or RACE products were resolved on a 32 well, 0.1 mm thick, 25 × 50 cm gel containing 6% (w/v) acrylamide as the resolving matrix and 7 M urea as denaturant. The sequencing gel was prepared and set up according to well established protocols (Sambrook *et al.*, 1989). The gel was run in TBE buffer (89 mM Tris-Cl, 89 mM boric acid, 2.5 mM EDTA), preheated for 30 minutes by running at 1500 V, with a constant power of 50 W.

Gel samples were denatured in 1/5 volume of glycerol loading dye solution by heating at 95°C for 2 minutes. A 2-3 μl aliquot of each sample was loaded per lane and the gel was run at a constant power of 50W until the desired separation had been achieved. The bromophenol blue and the xylene cyanol dyes were used as molecular size indicators during electrophoresis, running at approximately 26 and 106 bases, respectively on a 6% (w/v) polyacrylamide gel.

2.2.7. Southern Blotting (Southern, 1975)

2.2.7.1. Blotting DNA Fragments onto Nylon Supports

After agarose gel electrophoresis of DNA (section 2.2.6.1), the gel was photographed alongside a ruler under UV illumination, for later molecular weight determination. The gel was washed briefly in water. If the bands of interest were greater than 4 kb, the DNA was partially depurinated by acid treatment which in turn leads to DNA cleavage, to enhance the transfer of large molecular weight DNA. Depurination was achieved by washing in 10 gel volumes of 0.25 M HCl for 30 minutes, or until the xylene cyanol and bromophenol blue changed colour to green and yellow, respectively. The gel was rinsed in water and the DNA was denatured in 0.5 N NaOH containing 1.5 M NaCl for 20 minutes. The gel was again rinsed in water and neutralised with 0.5 M Tris-Cl, pH 7.0 containing 1.5 M NaCl.

Next, the gel was placed onto a sponge covered with three pieces of Whatmann 3MM paper, all of which were pre-soaked and placed in a reservoir of 20 × SSC. A piece of uncharged nylon membrane (Hybond-N, Amersham Pharmacia Biotech), pre-soaked in water, was carefully rolled onto the gel to avoid creating air pockets. Three further pieces of Whatmann 3MM paper, pre-soaked in 20 × SSC, were laid on top of the gel followed by a 4 to 5 cm layer of dry paper towels. All paper, sponge and towels were cut to size to avoid any 'short circuits' in solvent flow that could prevent transfer of DNA. A 400 g weight was placed on top of a piece of glass which evenly distributed the weight. Blotting was performed overnight, after which the membrane was washed in 2 × SSC, dried then fixed in a Spectrolinker XL-1500 UV crosslinker, using the automatic optimal setting (Spectronics Corporation, New York, USA).

2.2.7.2. Hybridisation of a DNA Blot with a Radiolabelled Probe

The Southern blot (section 2.2.7.1) was pre-soaked with 6 × SSC then placed in a hybridisation tube (Hybaid, Ashford, Middlesex, UK) containing hybridisation buffer (~1 ml / 10 cm^2 membrane). The membrane was pre-hybridised for 3 hours at 68°C,

rotating in a hybridisation oven (Hybaid). The radiolabelled probe, prepared by random nucleotide priming (section 2.3.1) was denatured by at 95°C for 10 minutes, then rapidly cooled on ice before being adding to fresh hybridisation buffer, preheated to 68°C. The hybridisation of the probe to the membrane was performed overnight at 68°C, with rotation, after which the membrane was washed with an equal volume of 2 × SSC for 10 minutes at room temperature. The membrane was then subjected to a series of buffer washes with increasing stringency monitoring the level of radioactivity before each step. All washing solutions contained 0.1% (w/v) SDS and were performed for 15 minutes with rotation. The low stringency wash consisted of two washes with 2 × SSC at room temperature. Medium stringency was two washes at 42°C, using the same buffer. For high stringency, two washes at 68°C in 0.1 × SSC were performed. After the final wash, the membrane was rinsed in 2 × SSC, covered in plastic wrap and exposed to X-ray film (X-OMAT AR; Kodak, Cambridge, UK) at -70°C.

2.2.7.3. Removal of Probes from the Hybridisation Membrane

Nylon membranes were routinely used for blotting because they could be repeatedly stripped and re-probed. To remove probes, the membrane was also washed in 0.4 M NaOH for 30 minutes at 45°C and rinsed twice in 200 mM Tris-Cl, pH 7.0, 0.1 × SSC, 0.1% (w/v) SDS. If insufficient radioactivity was removed, the membrane was washed in boiling 0.1% (w/v) SDS and allowed to cool to room temperature. Stripping of the membrane was assessed by exposing the blot to X-ray film for 2 days. Membranes were stored in plastic wrap at -20°C.

2.2.8. Measuring DNA concentration

2.2.8.1. Optical Density

The absorbance of an appropriately diluted solution was measured at 260 nm in a CECIL series 2 CE 292 Digital Ultraviolet Spectrophotomomter. The nucleic acid concentration was calculated assuming that an absorbance of 1 corresponded to 50 μg/ml of double stranded DNA or 40 μg/ml of single stranded DNA or RNA.

2.2.8.2. Rapid DNA Concentration Estimation Using Small Volumes

DNA was prepared at known concentrations of 0, 1, 2.5, 5, 7.5 and 10 μg/ml in TE buffer. For convenience DNA molecular size markers (SmartLadder) were used. A volume of 4 μl of the standard DNA solutions and the test DNA samples were mixed with an equal volume of 1 μg/ml ethidium bromide solution on plastic wrap. The concentration of the DNA test solution could be estimated by comparing the fluorescence intensity with the standard solutions of DNA under UV light.

2.2.9. Enzymatic Manipulation of DNA

2.2.9.1. Digestion of DNA with a Single Endonuclease

A restriction digest was usually set up in a 20 μl reaction containing 1 × restriction endonuclease buffer (New England Biolabs), 0.1-4 μg of DNA in TE buffer and less than 2 μl volume of endonuclease (1 to 5U/μg DNA). The endonuclease buffer and reaction temperature were selected according to the recommendations of the suppler of the enzyme. The reaction proceeded for 2 hours to overnight and was stopped either by DNA extraction (section 2.2.5), freezing or by addition of 1/10 volume of glycerol loading dye for electrophoresis samples.

2.2.9.2. Digestion of DNA with Multiple Endonucleases

When using two or more endonucleases, the reaction buffers were selected from a range of commercial buffers to achieve 100% activity of the enzyme according to the manufacturer. When this was not possible, the DNA was extracted after the first restriction digest and the subsequent digestion was performed in the appropriate buffer. Experiments were designed so that multiple endonuclease digestions could be performed using a single buffer.

2.2.9.3. Restriction Mapping with Multiple Endonucleases

For restriction mapping, several separate reactions were set up with the DNA of interest using restriction enzymes that infrequently cleave DNA. Typically restrictions endonucleases with six or five base-pair recognition sequences were used (e.g., *Eco*R I, *Hind* III, *Bam*H I, *Pst* I, *Kpn* I, *Xba* I, *Sal* I, *Xho* I, *Cla* I, *Bst* XI, *Eco*R V, *Apa* I, *Bgl* I, *Sal* I and *Sac* I). Complete cleavage was assessed by electrophoresis of a small aliquot of DNA. The remaining DNA was cleaned (section 2.2.5) and cleaved again with a different enzyme. Single, double and triple digestions were run overnight on a long analytical agarose gel (200 × 250 × 5 mm) and DNA fragments were compared. Gradually, a detailed restriction map was constructed. To add context to the map, the restriction digests were blotted and hybridised with radiolabelled probes (section 2.2.6.3).

2.3. DNA Modification and Labelling

2.3.1. Labelling of DNA by Random Oligonucleotide-Primed Synthesis

DNA was labelled using a commercially available random oligonucleotide primed DNA synthesis kit (Multiprime DNA labelling system; Amersham Pharmacia Biotech) (Feinberg and Vogelstein, 1983). Briefly, a restriction endonuclease fragment or PCR product was gel purified (section 2.2.5.4), resuspended in TE buffer and denatured by boiling and cooling on ice in the presence of 5 μg of random hexanucleotides in a 20 μl reaction volume. The Klenow fragment of *E. coli* DNA polymerase I (2 U) was used to generate uniformly radiolabelled DNA by incorporating [α-^{32}P] dATP (30 μCi) during a one hour synthesis step at room temperature. The radiolabelled product was separated from unincorporated nucleotides by gel filtration using a NICK column (Sephadex G-50, Amersham Pharmacia Biotech), following the manufacturer's instructions.

2.3.2. Terminal Deoxynucleotidyltransferases

Terminal deoxynucleotidyltransferase is a template-independent DNA polymerase, which was used to tag single-stranded DNA by incorporating a string of one type of

nucleotides to the 3'-hydroxyl terminus. The DNA from a 10 µl first strand DNA synthesis reaction (section 2.5.2) was purified using Wizard DNA cleanup (section 2.2.5.3), eluted with 20 µl of water and concentrated under vacuum to 5 µl. This formed part of a 10 µl reaction mix which contained 200 mM sodium cacodylate, 25 mM Tris-HCl, pH 7.4, 1.5 mM $CaCl_2$, 0.1 mM DTT, 100 µM dCTP, 0.5 mg/ml BSA and 25 U terminal transferase (Boehringer Mannheim). After a 30 minute incubation at 37°C, the reaction was stopped by the addition of 0.5 M EDTA or by heat inactivation for 2 minutes at 65°C. The tailed single-stranded DNA could either be purified using Promega Wizard DNA cleanup (section 2.2.5) or an aliquot of this reaction mix could be directly used for PCR (section 2.8).

2.3.3. Alkaline Phosphatase

Calf intestinal phosphatase was used to hydrolyse 5'-phosphate residues from DNA. This was used exclusively to dephosphorylate the vector to prevent self-ligation during ligation. Dephosphorylation of DNA was carried out in a 50 µl reaction containing 20 mM Tris-HCl, pH 8.0, 1 mM $MgCl_2$, 1 mM $ZnCl_2$, 0.1 U calf intestine phosphatase/pmol 5' ends was used, at 37°C for 30 minutes. The DNA was purified by Wizard DNA cleanup (section 2.2.5).

2.3.4. Ligation with T4 DNA Ligases

T4 DNA ligase was used to join duplex DNA fragments. Vector and insert with blunt or compatible cohesive ends were ligated. Typically three different molar ratios of vector to insert were used (1:1, 1:3 and 1:9). For each ligation reaction less than 5 µg of DNA in a 20 µl reaction volume containing 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, 50 µg/ml BSA and 100 U T4 DNA ligase was incubated overnight at 12°C. For transformations 1-10 µl of each ligation reaction was used (section 2.2.3.2).

2.4. Construction of Hybrid DNA Molecules

Both vector and insert were cleaved with the appropriate restriction endonucleases (section 2.2.9). For cohesive end ligation, both vector and insert were either digested with the same restriction endonuclease or those that generated the same cohesive ends. To prevent self-ligation, the vector was dephosphorylated with calf intestine phosphatase (section 2.3.3). If restriction endonuclease digestion resulted in generating more than one fragment, the desired fragment was purified by gel extraction (section 2.2.5.4) following agarose gel electrophoresis. In order to prevent damage to DNA intended for cloning, the DNA was not exposed to UV light. After ligation and transformation, positive transformants were selected by PCR screening or by restriction digest (section 2.8.6) to determine the orientation of the construct.

2.5. Preparation and Analysis of RNA

2.5.1. Purification of Total RNA from Tissue

Tissue was rapidly removed from the animal, frozen in liquid nitrogen and stored at -70°C. To avoid ribonuclease contamination, all glassware had been baked for 4 hours at 300°C and sterile plastic-ware was always used. Water was treated with diethylpyrocarbonate (DEPC), by making it a 0.2% (v/v) solution, shaking vigorously and autoclaving to inactivate the DEPC.

Total RNA was purified from tissue using a commercially available kit (RNeasy Midi Kit, Qiagen), a single step isolation procedure which uses phenol-chloroform phase separation to extract RNA from the tissue (Chirgwin *et al.*, 1979). Following the manufacturer's protocol, tissue was homogenised in a solution of guanidine and extracted with phenol-chloroform. RNA was then precipitated with isopropanol, the pellet was rinsed in 70% (v/v) ethanol and dissolved in DEPC treated water.

2.5.2. Reverse Transcriptase for Single-Stranded DNA Synthesis

Approximately 300 ng of total RNA was heated at 65°C for 2 minutes and cooled rapidly on ice. This removes secondary structure from the RNA, like hairpin loops, that interfere with cDNA synthesis and generate artefactual cDNA copies of the mRNA. Single-stranded DNA was generated in a 10 µl reverse transcription reaction, using the denatured RNA, 5 µg BSA, 1 mM of each dNTP, 500 ng of actinomycin D (Promega), 15 pmol of poly dT primer and 200 U of Moloney murine leukaemia virus reverse transcriptase (Boehringer Mannheim) in the supplied buffer. The single-stranded DNA synthesis was performed for 1 hour at 37°C and the product could be used in anchor PCR (section 2.8.3) or RT-PCR (section 2.8.5)

2.6. Screening of Recombinant DNA Libraries

2.6.1. Plating and Transferring Bacteriophage Libraries to Filters

Before plating, the titre of the library was calculated (section 2.2.4.5). The recombinant phage (~20,000 pfu) was mixed with 5 ml of saturated *E. coli* culture (Y1090r⁻ or XL-1 Blue) and 7 ml of top agarose and poured onto lambda plates. The number of plates used was determined according to the number of phage that give a full representation of the library, complete genome coverage or mRNA expression from the tissue, as detailed for the library by the commercial supplier.

The plates were incubated at 37°C for 6 to 8 hours, until the plaques were 1-2 mm in diameter. The plates were then cooled for 1 hour at 4°C, to prevent the top agarose from peeling when lifting filters. Nitrocellulose filters (Hybond-C, Amersham Pharmacia Biotech) were applied to the plates for 1 minute in the first instance, and two minutes for the duplicate filter to allow transfer of plaques to the membranes, marking the orientation of the filter on the plate with an inked pin. The filters were dried for ten minutes at room temperature and the phages attached to the membranes were lysed, denatured and neutralised. Filters were first placed onto Whatmann 3MM paper saturated with 0.2 M NaOH/1.5 M NaCl, then on paper soaked in 0.4 M Tris-

HCl, pH 7.5 containing 2 × SSC and finally 2 × SSC, each incubation was carried out 2 minutes. The filters were dried and baked at 80°C for 2 hours, then stored at room temperature until further use. The libraries used were porcine lung cDNA λgt11 (Clontech) and a mouse genomic library in λFIX-II (Stratagene).

2.6.2. Purification of Bacteriophage Clones (Helms *et al.*, 1985)

Filters from the bacteriophage primary screening plates were hybridised with radiolabelled DNA probes as described (section 2.2.7.2). The resulting autoradiographs were orientated with the filters, duplicate autoradiographs were aligned to reveal double positive signals disregarding artifactual signals. The autoradiographs were then used to isolate positive plaques from the primary screening plates.

The area containing the positive plaque was picked using a cut-off 1 ml pipette tip (5 mm diameter), the plug was placed into 1 ml of SM buffer for 1 hour. The phage eluate was then plated at a density of less than 500 pfu. This secondary bacteriophage plate was then transferred to nitrocellulose, hybridised with radiolabelled DNA probe and purified as before. This was repeated until the phage stock was pure: all the plaques on the plate were positive for the probe. A final phage stock was made in SM buffer, titred and stored (section 2.2.4).

2.7. DNA Sequencing

2.7.1. Manual Sequencing by Dideoxy Method

DNA was sequenced by the dideoxy method (Sanger *et al.*, 1977; Sanger *et al.*, 1980), using a commercially available T7 DNA polymerase (Amersham Pharmacia Biotech) following the manufacturer's instructions. Briefly, ~1 μg of plasmid DNA was denatured by alkali treatment. An oligonucleotide primer was annealed to the template and extended by T7 DNA polymerase in the presence of four deoxyribonucleoside triphosphates, one of which was ^{35}S-labelled (dATP). Four reactions were set up for each template with each containing one of the four dideoxyribonucleoside triphosphates

(ddNTPs) which when incorporated into the growing radiolabelled DNA, caused chain termination of the elongation reaction. The products of the sequencing reaction were run side-by-side on a high-resolution denaturing polyacrylamide gel (section 2.2.6.2) and visualised by autoradiography.

2.7.2. Automated Sequencing

After the introduction of an automated DNA sequencing service in the Department of Biochemistry, Oxford University, DNA was sequenced using one of various types of commercially available kits (ABI Prism and BigDye; Perkin Elmer Applied Biosystems, Warrington, UK or Thermosequenase; Amersham Pharmacia Biotech) following the manufacturer's recommendations and procedures. These commercial kits were based upon thermal cycle sequencing reaction using either fluorescent labelled primers or fluorescent labelled ddNTPs in conjunction with a heat stable DNA polymerase (Sears *et al.*, 1992). Briefly, this method consists of 25 cycles of denaturation, annealing and elongation steps with a heat stable DNA polymerase, similar to PCR but with only one primer thus performing a linear amplification. The sequencing products were run on a Perkin Elmer ABI 372 automated sequencer, by the Biochemistry Sequencing Service. The resulting fluorescent sequencing traces were analysed using the Macintosh computer program Seqman (DNAstar, London, UK).

2.8. Polymerase Chain Reaction (PCR)

The polymerase chain reaction (Saiki *et al.*, 1988) is a method for enzymatic amplification of a specific fragment of DNA which was used for cloning parts of cDNA and genomic DNA, analysis of allelic sequence variants and direct sequencing.

2.8.1. Enzymatic Amplification of DNA by PCR

The standard PCR assay was a 50 µl reaction mix containing the sense and antisense primers, at a final concentration of 0.5 µM, 0.2 mM of each dNTP and 2.5 U of *Taq* polymerase (Boehringer Mannheim) in a buffer of 50 mM KCl, 10 mM Tris-Cl, pH 9.0,

0.01% (v/v) Triton X-100. The amount of template used was usually 0.1 ng for plasmid DNA or 1 μg for genomic DNA. For each new PCR assay, the magnesium concentration and annealing temperature were optimised. Typically four different annealing temperatures (55°C, 57°C, 60°C and 62°C) and four different final concentrations of magnesium chloride (1.0, 1.25, 1.5 and 1.75 mM) were chosen for optimisation. The *Taq* polymerase enzyme was the last reagent to be added, preventing non-specific extension. The reaction was overlaid with mineral oil to prevent evaporation during temperature cycling. The PCR assays was performed in either a 48 tube DNA thermal cycler (Perkin Elmer Applied Biosystems) with 0.5 ml reaction tubes or for bulk PCR assays in an OmniGene PCR machine (Hybaid) with 96 well plates, each containing a 200 μl reaction volume. The program used for PCR varied for each type of amplification. However, a standard PCR program was established consisting of a denaturing step of 30 seconds at 94°C, a 30 second annealing step at the optimised temperature, followed by an extension step at 72°C for variable time (30 seconds to 3 minutes) depending on the desired size of the product. As an estimate, the extension time was 1 minute for each kb of DNA to be amplified. Between 25 and 35 cycles were used, depending on the efficiency of the reaction and the amount of template DNA. The reaction mix was denatured before cycling in a 1 minute denaturation step at 94°C. After the amplification cycles a final extension step of 7 minutes at 72°C was performed when the product was being directly sequenced (section 2.8.2) or TA cloned (section 2.8.4). This ensured that all the products were fully extended. When optimising the PCR assay, the conditions chosen were those which generated the greatest amount of specific product. Generally 5 to 10 μl of PCR was analysed on an agarose gel.

Usually, *Taq* polymerase was routinely used for amplification. For PCR assays requiring higher fidelity, as in the generation of constructs for use in protein expression, *Pfu* polymerase (Stratagene) was used.

2.8.2. Preparing PCR Products for Direct Sequencing

Only PCR products less than 1500 bp were used in direct sequencing, longer PCR products failed to give sequence information. After performing PCR, the product was analysed on an agarose gel. If more than one product was visible, the desired extension product was gel purified (section 2.2.5.4), but eluted from the column in water, instead of TE buffer, as the salt concentration can interfere with the sequencing reaction. For PCR assays that generate a single product the, DNA was purified from buffers and the short oligonucleotide primers and free nucleotides by DNA Wizard cleanup (section 2.2.5). After eluting the DNA with water, approximately 100-500 ng of PCR product was used in cycle sequencing (section 2.7.2).

2.8.3. Anchor PCR

Anchor PCR is an application utilising the polymerase chain reaction to selectively amplify unknown upstream regions of a cDNA from which only a small stretch of sequencing information is known, otherwise known as 5' rapid amplification of cDNA ends (5' RACE) (Frohman *et al.*, 1988).

Single-stranded cDNA template was synthesised using a reverse oligonucleotide specific for the cDNA of interest (section 2.5.2). The single-stranded cDNA was purified from the buffer salts and the enzyme by Wizard DNA cleanup (section 2.2.5) and tailed with deoxycytidine using terminal deoxynucleotidyltransferase (section 2.3.2). The DNA was purified (section 2.2.5.2) and the tailed single-stranded DNA was used as a template in a 50 µl low stringency PCR. The low stringency PCR was performed on 5µl of the first strand cDNA mixture, using 100 pmol of the 'anchor' primer (GAC TCG AGT CGA CAT CGA $TG_{17}H$) and the sequence specific reverse primer. The anchor primer contained restriction sites for Xho I, Sal I, Cla I. Low stringency PCR was performed to amplify the entire transcripts, using 1.5 mM $MgCl_2$, an annealing temperature of 42°C and 40 cycles of amplification. After DNA purification (Wizard DNA-cleanup system) to remove excess oligonucleotides, a nested

PCR was performed using 1/50 of the first PCR as a template with an upstream sequence specific reverse oligonucleotide and the 'anchor' oligonucleotide (GAC TCG AGT CGA CAT CGA T), 100 pmol of each. Amplification was performed for 35 cycles using a range of annealing temperatures (50, 55, 60 and 65°C) to optimise the specificity and obtain a sufficient yield. The second PCR assay was performed to achieve higher specificity. The 5' RACE products were analysed by agarose gel electrophoresis, gel purified (QIAquick, Qiagen) and then subcloned using cohesive end ligation. A selection of independent clones was then chosen for automated DNA sequencing (section 2.7.2).

2.8.4. Cloning of PCR Products by TA Overhangs

Utilising the ability of *Taq* polymerase to add single nucleotides, preferentially adenosine, to the 3' end of duplex DNA strands, products of *Taq* polymerised PCR assays could be cloned into thymidine overhang vectors (Holton and Graham, 1991). PCR assays were performed with a final 7 minute extension at 72°C, to ensure adenosine tailing of the product. The product was purified then subcloned into a commercially available T-tailed vector (pMOS; Amersham Pharmacia Biotech) using conventional T4 ligase (section 2.3.4) or a T-tailed vector pretreated with DNA topoisomerase I (pCRII-TOPO; Invitrogen) following the manufacturer's instructions.

2.8.5. Reverse Transcriptase PCR

For reverse transcriptase-PCR (Gurr *et al.*, 1991), single-stranded DNA was prepared from total RNA using Moloney murine leukaemia virus reverse transcriptase primed with either oligo dT primer or a sequence specific reverse oligonucleotide (section 2.5.2). An aliquot of the generated single-stranded DNA, typically 1/10 of the reaction volume, was used as the template in PCR. For increased specificity nested primers were generally used (see sections 2.8.4).

2.8.6. PCR Screening for Positive Colonies

E. coli colonies containing plasmids with the desired insert could be identified by a suitable restriction digest (section 2.2.9.3). However, for speed and convenience, the colonies were screened by PCR assay. Briefly, a colony was picked with a pipette tip and used to inoculate a tube containing 20 µl of LA media. 1 µl of the resuspended colony was used as template for PCR. Using the same tip approximately 1 µl of media was transferred to a tube containing 20µl of PCR mix. The primers used were usually a combination of a vector primer and a sequence specific primer to obtain orientation information or vector primers to indicate the presence of an insert. This procedure could easily be performed on 47 or 95 different colonies plus, a reaction containing medium only, as a negative control. The PCR assay was initiated with a 5 minute heating step at 96°C to lyse the bacteria. Approximately 5 µl of the PCR reaction was used to visualise the product on an agarose gel. The presence of the predicted PCR product size was taken as evidence of a positive clone. The inoculated medium could then be used as an inoculant for a seed culture for plasmid DNA preparation (section 2.2.2). The positive clones were usually confirmed in conjunction with another method, such as sequencing.

2.9. Bioinformatics

2.9.1. Submission of Sequences to Databases

Novel sequences were submitted to Genbank (*gb-sub@ncbi.nlm.nih.gov*). For short, single sequences, data was submitted to GenBank (Benson *et al.*, 1997) via the World Wide Web, using the BankIt form on the National Centre for Biotechnology Information (NCBI) home page (*http://www.ncbi.hlm.nih.gov*). For more elaborate sequence information, which required more detailed annotation over a longer period of time, a stand-alone software package Sequin was used. This package was used in the case of several non-overlapping contigs from a gene. Sequin was obtained from the NCBI anonymous FTP server (*http://www.ncbi.nlm.nih.gov/Sequin/index.html*).

Submitted data can be found in Appendices 2, 3, 7 and 8 and as accession numbers AF047741, AF047742, AF132496 and AF133668.

2.9.2. Submitting Sequences for Homology Searches using BLAST

The Basic local alignment search tool (BLAST), was used to search the sequence databases at NCBI using their World Wide Web interface (*http://www.ncbi.nlm.nih.gov/BLAST*). The BLAST suite of programs employ a fast search algorithm for matching the sequence of interest with the daily updated database of sequences and uses statistical analysis to judge the significance of the sequence matches (Altschul *et al.*, 1994). On most occasions BLASTN was used which compares a nucleotide sequence with the nucleotide databases. However, for greater sensitivity for finding similarities, BLASTX to search the protein database with open reading frames from a nucleotide sequence was also performed. To search for murine ESTs, a restricted set of the GenBank database containing only murine ESTs was examined using BLAST.

2.9.3. Sequence Analysis

Sequence analysis was performed using the DNAstar suite of programs for Apple Macintosh. Continuous stretches of sequence data were constructed using SeqMan. Restriction maps were generated using MapDraw. Alignments were performed with Bestfit from the Wisconsin software package. Various other molecular biological tools and Internet sites were obtained from a biomedical research Internet link site (*www.public.iastate.edu/~pedro/reasrch-tools.html*).

2.10. Protein Methods

2.10.1. Electrophoretic Separation of Proteins (Laemmli, 1970)

A Biorad mini-protean II apparatus (Biorad, Hemel Hempstead, Hertfordshire) was used for protein gel electrophoresis, using 72 × 105 × 0.5 mm gels. Prestained high molecular weight markers were purchased from Sigma-Aldrich. A 50 mm long resolving gel was cast with an acrylamide concentration varying from 7.5 to

12.5% (w/v) containing 375 mM Tris-Cl pH 8.8, 0.1% (w/v) SDS, 0.05% (w/v) ammonium persulphate and 0.01% (v/v) TEMED. The gel was levelled with 2 ml of water. After setting, the stacking gel was cast, composed of 4% (w/v) acrylamide containing 125 mM Tris-Cl pH 6.8, 0.1% (w/v) SDS, 0.05% (w/v) ammonium persulphate and 0.01% (v/v) TEMED. A 15 well or 10 well comb was used in each case. Protein samples were mixed with one volume of sample buffer in the presence of reducing agent, 100 mM DTT or 10 mM iodoacetamide to prevent disulphide interchange and heated at 100°C for 5 minutes before loading. Gels were run in 25 mM Tris-HCl, pH 8.3, 190 mM glycine, 0.1% (w/v) SDS at 14 mA per gel for 4 to 5 hours. After separation, the gels were disassembled and the protein detected by Coomassie Blue or Silver Staining (section 2.10.2.1).

2.10.1.1. Detection of Protein

Proteins separated on polyacrylamide gels were visualised using one of two methods. Coomassie blue staining was used to quickly identify proteins with a detection limit of 0.3 to 1 μg/protein band. For lower amounts of protein and for staining DNA gels (section 2.2.6.2), silver staining was used with a detection limit of 2 to 5 ng/protein band.

2.10.1.2 Coomassie Blue Staining

Coomassie blue staining depends upon non-specific binding of the dye to protein. The gel was fixed in 5 gel volumes of 50% (v/v) ethanol, 10% (v/v) acetic acid for 20 minutes. The gel was then stained for 1 to 4 hours in 0.05% (w/v) Coomassie brilliant blue R-250 in fixing solution. The gel was destained in 10% (v/v) ethanol, 10% (v/v) acetic acid. Gels were stored in water or dried onto Whatmann 3MM, but photographed before drying.

2.10.1.3. Silver Staining

Detection of protein bands by silver staining requires binding of silver cations to various chemical groups on proteins (Morrissey, 1981). The gel was first fixed as for Coomassie blue staining (section 2.10.2.2). Appearance of the stained bands was enhanced by fixing for 30 minutes in 5 gel volumes of 10% (v/v) glutaraldehyde. The gel was washed four times in 10 gel volumes of water, for at least 30 minutes each time. To ensure more reproducible staining, the gel was reduced for 30 minutes in 5 μM DTT. The gel was then stained for 15 minutes in 0.1% (w/v) silver nitrate solution. After a brief rinse in water the gel was developed in three changes of 3% (w/v) sodium carbonate supplemented with 0.02% (v/v) formaldehyde. The gel was then washed in water, photographed and dried onto paper.

2.10.2. Protein Blotting and Detection

2.10.2.1. Transfer of Protein to Nitrocellulose Support

The protein samples for blotting (Towbin *et al.*, 1979) were first separated by gel electrophoresis. Prestained markers were used as both molecular weight indicators and as a visual confirmation of protein transfer. After the gel has been equilibrated in transfer buffer (50 mM Tris-base, pH 8.4, 40 mM glycine, 20% (v/v) methanol.) for 5 minutes, the electroblotting cassette was constructed. The gel was laid upon a piece of sponge covered with a piece of Whatmann 3MM for support. The nitrocellulose transfer membrane (Hybond-C, Amersham Pharmacia Biotech), pre-soaked in transfer buffer, was placed upon the gel, followed by a piece of Whatmann 3MM and a sponge. Sponges and Whatmann 3MM were all pre-soaked in transfer buffer to prevent drying of the gel. Finally, the cassette was orientated in the electroblotting apparatus with the membrane facing the anode. Blotting was performed using the mini-transblot apparatus (Biorad). Electrophoretic transfer of proteins to the membrane was carried out at 100 V for 2 hour, with gentle stirring and using an ice pack in the transfer buffer for cooling. The orientation of the gel on the membrane was marked, the membrane was rinsed in

transfer buffer and either stained (section 2.10.3.2) or probed with antibody (section 2.10.4.4).

2.10.2.2 Qualitating Protein with Ponceau S

Ponceau S was used to confirm blotting and location of protein bands on the nitrocellulose membrane, as this method of staining is irreversible, rapid and does not interfere with immunoblotting. Following protein transfer to nitrocellulose, the membrane was stained for 5 minutes in a 0.5% (w/v) solution of Ponceau S in 1% (v/v) glacial acetic acid. The membrane was destained for 2 minutes in water, marked with pencil and photographed. The membrane could be destained completely in water and further used for immunoblotting.

2.10.2.3. Dotblot

To optimise antibody staining, the dilution of primary and secondary antibody was conveniently determined using purified protein or a crude extract, which was directly transferred to nitrocellulose without separation. This method determined the approximate resolution limit of an antibody dilution before proceeding to Western Blotting (section 2.10.3.). The nitrocellulose membrane was first pre-soaked in PBS for 2 minutes, then clamped into a 96 well cassette (Biorad). After rinsing the membrane with PBS, 12 lanes of eight different protein concentrations were added directly to the wells of the cassette, including a buffer control. Typically the amount of protein added to the membrane varied from 1 μg to 1 ng in a 25 μl volume. The protein solution was incubated on the membrane for 30 minutes at room temperature, each well was washed three times with 100 μl of PBS, the cassette was disassembled and the membrane divided into 12 strips. The strips were then blocked with BSA and probed with antibody (section 2.10.4.4).

2.10.3. Antibody Methods

2.10.3.1. Preparation of Polyclonal Antisera

Polyclonal antisera raised against recombinant proteins in rabbits were produced commercially (Regal Group UK, Surrey, UK. Briefly, eight aliquots of 250 μg of antigen were supplied in 1 ml of PBS. The two rabbits to be immunised were first bled to obtain pre-immune sera. The first two aliquots of the antigen were mixed with an equal volume of Freund's adjuvant and injected into the limbs of the two rabbits (New Zealand White Rabbits). The animals were bled two weeks following the priming immunisation (10 ml were obtained each time) and sera was prepared. The first booster immunisation was administered four weeks later and the second bleed was drawn after another two weeks. A further boost immunisation was given three weeks later and test bleeds were collected 10 days after. Anti-sera prepared from the pre-immune sera and the three test bleeds were tested in an indirect ELISA to detect specific antibodies developing against the antigen. The responsive rabbits were then killed by heart puncture and the animal bled.

2.10.3.2. Preparation of Antisera from Blood

Blood was allowed to stand in 50 ml centrifuge tubes for 4 hours at room temperature until a clot developed and then left overnight at 4°C to allow the clot to contract. The clot was gently removed with a wooden applicator and the serum was transferred to a fresh centrifuge tube. Residual cells and debris were removed by centrifugation at 3,000 × *g* for 10 minutes at 4°C. The serum was stored in 1 ml aliquots containing 1 mM sodium azide at -20°C.

2.10.3.3. Indirect ELISA to Detect Specific Antibodies

This assay was used to determine and monitor the titre of antibodies (Schots *et al.*, 1988). Purified antigen (1 μg) was coated onto half of the rows of a microtitre plate (Poly-sorb, Nunc, Denmark) in 100 μl of ELISA coating buffer (15 mM Na_2CO_3, 35 mM $NaHCO_3$, pH 9.6.) per well. The other wells contained no antigen and therefore

acted as a buffer control. After an overnight incubation at 4°C, the plates were washed three times with 150 µl of PBS per well and tapped out onto paper. The residual binding capacity of the plate was blocked with a combination of 0.05% (v/v) Tween-20 and 0.25% (w/v) BSA in PBS (25 µl) for 2 hours at 37°C. The wells were washed as before. The test antibody, typically pre-immune serum, primary, secondary and tertiary bleeds from a rabbit injected with the antigen, was titrated across the ELISA plate containing either a single concentration of antigen or no antigen. The antibody sample (100µl) was added to each well, with a two-fold dilution in each, ranging from a 1 in 10 dilution of antisera to approximately 1 in 10,000 dilution in PBS. After a 1 hour incubation at 37°C, the plate was washed as before and the bound antibody was detected using a commercially available goat antibody against rabbit immunoglobulin conjugated to alkaline phosphatase (Sigma). 100 µl of a 1 in 5,000 dilution of a goat anti-rabbit antibody was incubated in each well for 1 hour at 37°C. The plate was washed as before and the amount of antibody bound was determined by a colourimetric method. The phosphatase substrate, the sodium salt of 4-nitrophenylphosphate (PNPP), was added at a concentration of 1 mg/ml in ELISA developing buffer. The plates were incubated at room temperature and the hydrolysis of PNPP could be visually detected by the appearance of a yellow colour, which was quantitated at 405 nm in a microtiter plate reader (Model MK II, Titertek). Readings were normally complete after 30 minutes of hydrolysis.

2.10.3.4. Probing with Antibody

Nitrocellulose membranes from either a dotblot (section 2.10.3.3) or a protein gel transfer (section 2.10.3.1) were blocked for at least 2 hours in 1% (w/v) BSA in a solution of TBS containing 0.1% (v/v) Tween 20 (TBS-T). All incubations were performed at room temperature with shaking. The membrane was covered in TBS-T buffer containing the desired dilution of antiserum (typically 1 in 5,000) and incubated for 30 minutes. The membrane was washed three times in an excess of TBS-T over 15 minutes and probed for 30 minutes with a 1 in 10,000 dilution of secondary

antibody, goat anti-rabbit conjugated to alkaline phosphatase (Sigma) (Burnette, 1981). The membrane was then washed as before in TBS-T. The signals were developed by incubating the membrane with 20 ml of development buffer (100 mM Tris-HCl, pH 9.5, 5 mM $MgCl_2$, 100 mM NaCl) containing 3.3 mg of BCIP and 2.64 mg of NBT. Colour development was stopped by washing the membrane in water.

CHAPTER 3: THE MOLECULAR CHARACTERISATION OF THE MOUSE LUNG SURFACTANT PROTEIN D GENE

3.1. Introduction and Aims

Upon examining the literature, lung surfactant protein D appears to participate in the two discrete functions within surfactant. Initially, characterisation of SP-D concentrated upon the function of SP-D in surfactant homeostasis and interaction with the lipid components of surfactant (Persson *et al.*, 1990). Later, *in vitro* examination of this protein suggested a role for SP-D in the innate immune system. SP-D has been shown to bind to an ever expanding array of pathogens (Crouch, 1998), in many cases leads to their phagocytosis (Hartshorn *et al.*, 1998). SP-D can also modulate immune effector cells including the release of cytokines (Borron *et al.*, 1998b; Borron *et al.*, 1998a) and has also been shown to provide protective resistance to viral infection in mice (Reading *et al.*, 1998).

It was envisaged that the best strategy to ascertain the function of SP-D was to transgenically generate mice that lacked SP-D. These mice would determine whether or not SP-D was participating in lipid homeostasis and surfactant function or, highlight the role SP-D was playing in innate immunity and the types of pathogens it could provide defence against. The first step in generating knock-out mice was to clone the gene for murine SP-D.

The regulation of expression and the genetic elements responsible for driving expression of SP-D (Rust *et al.*, 1996; Crouch *et al.*, 1991; Ogasawara *et al.*, 1991) had only been studied to a limited degree, in comparison to those of other surfactant proteins, SP-A, SP-B and SP-C (Mendelson *et al.*, 1991; Whitsett and Glasser, 1998; Pryhuber *et al.*, 1998) and those of the human collectin, MBL (Madsen *et al.*, 1995). Consequently, cloning the murine promoter for SP-D would be a valuable tool in

understanding the regulation of this gene, presumably a complex one, since the gene product maybe involved in both innate immunity and lipid homeostasis.

The emphasis of this work was guided by the successful gene knockouts of other lung and collectin related proteins (Korfhagen *et al.*, 1998). The mice null for the SP-B gene die of respiratory failure (Clark *et al.*, 1995), while heterozygotes show decreased lung compliance (Clark *et al.*, 1997). *In vivo* experiments with the null mouse for Clara cell 10 kDa protein (CC10), a protein that in wildtype animals is predominantly expressed in the lung, has revealed that CC10 helps control inflammation within the lung (Harrod *et al.*, 1998). The earliest transgenic mouse for a collectin contained the human MBL gene on a background of the two wildtype murine MBL genes (Tabona *et al.*, 1995). However, the construct used did not contain a large enough region of the human MBL promoter so the human gene lacked the acute phase response of background murine genes. The most informative collectin transgenic was a knockout mouse for the SP-A gene (Korfhagen *et al.*, 1996), which generated an immune compromised animal, susceptible to a number of pathogens (LeVine *et al.*, 1997; LeVine *et al.*, 1998; LeVine *et al.*, 1999b; LeVine *et al.*, 1999a).

During the process of this work, two independent groups successfully generated null mice for the SP-D gene (Korfhagen *et al.*, 1998; Botas *et al.*, 1998) a fact which consequently affected the direction of this project. This chapter describes the characterisation of the murine lung surfactant protein D gene, genetic mapping of the gene and the surrounding collectin genes, phylogenetic analysis of SP-D and the generation of specific antibodies against the murine surfactant collectins. This is followed by a discussion of the work and its implications. Most of this work has been published as a full paper in the American Journal of Respiratory Cell and Molecular Biology (Lawson *et al.*, 1999).

3.2. Generating a Probe: Amplification of a Fragment of Mouse SP-D Using Species Conserved Primers

In order to clone the murine gene for SP-D, a cDNA fragment encoding part of murine SP-D was cloned by PCR and used as the probe for screening a murine genomic library. The three cDNA sequences for SP-D from different species known to date (human, rat and bovine) were aligned using the software program MegAlign (DNAstar) and optimised by eye (section 2.9.3). The two primers designed at regions of high conservation across the three different species to encompass the majority of the coding region for SP-D. As an initial strategy and for convenience, non-degenerate primers were used over a range of PCR assay conditions instead of degenerate primers covering all sequence variants. The annealing positions of the primers are indicated in Figure 3.1. The primers would give an approximate 1100 bp PCR product based on the rat SP-D cDNA sequence.

Total RNA was purified (section 2.5.1) from BALB/c murine lungs. Single stranded DNA was prepared by reverse transcriptase (section 2.5.2) initiated with the complementary species-conserved primer, DCon-R[1]. A standard PCR assay (section 2.8.1) was optimised by trying four different magnesium concentrations and four different annealing temperatures using the primers DCon-F and DCon-R. The PCR assay was monitored for the production of sufficient product after 25, 30 and 35 cycles of amplification. The optimised annealing temperature was 53°C with a magnesium concentration of 1.5 mM. The PCR product was cloned into pBluescript using the *Eco*R I restriction site introduced by the primers. After transformation in *E. coli*, positive colonies containing the insert were identified by PCR screening (section 2.8.6). The cloned PCR product (pMDPCR) was sequenced and identified as the murine homologue of SP-D since it showed greatest identity to SP-D after alignment with the known murine collectins (SP-A, MBL) and bovine collectins (SP-D, conglutinin and CL-43). This PCR derived sequence of murine SP-D showed

[1]The sequence of all the primers referred to in this chapter are shown in Appendix 1

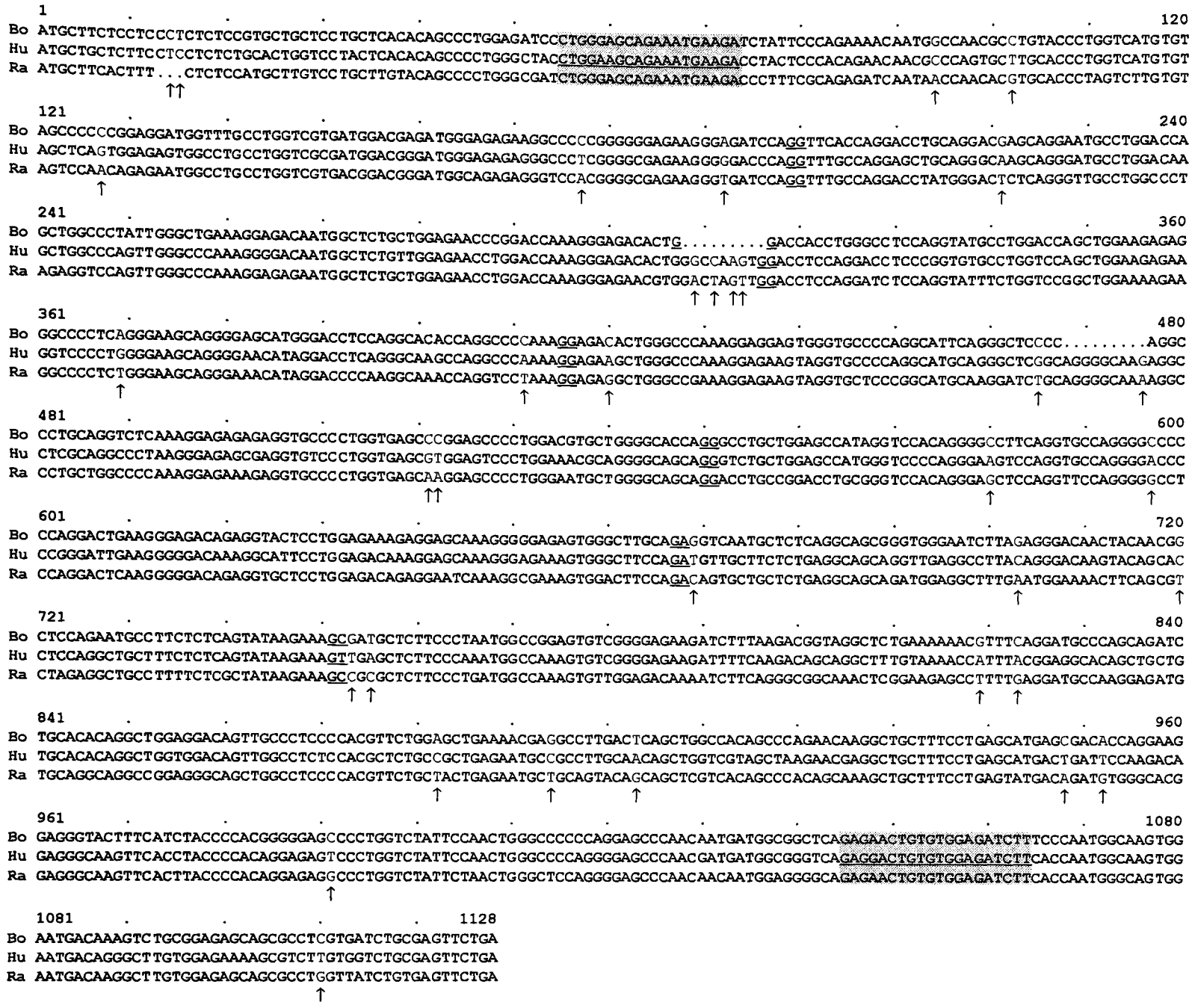


Figure 3.1. Alignment constructed for the design of conserved oligonucleotides. The conserved oligonucleotides, Dcon-f and Dcon-r, (see Appendix 4) were designed for amplifying mouse SP-D based upon the human (hu), rat (ra) and bovine (bo) sequences known at the time of the study. Nucleotide positions conserved between 2 or all species are shown in bold. Unconserved nucleotide positions are indicated by an arrow. The annealing points of the primers are shown (underlined and shadowed). The positions of introns, based upon the human gene structure (Crouch *et al.*, 1993), are indicated with two underlined nucleotides marking the 3' and 5' nucleotides of each pair of exons. The numbering is based upon the human SP-D sequence, running from the first ATG codon. The positions of the introns are consistent with those observed for murine SP-D (section 3.3). The alignment was performed with Bestfit from the Wisconsin software package (section 2.9).

complete identity with the later published murine SP-D cDNA sequence (Motwani *et al.*, 1995).

3.3. Cloning and Characterisation of the Murine Gene for SP-D

The insert from the cloned murine SP-D PCR product (pMDPCR) was excised by *Eco*R I restriction digestion and purified by gel extraction (section 2.2.5.4). This cDNA fragment of murine SP-D cDNA was radiolabelled by incorporation of [α-^{32}P] dATP with the Klenow fragment of *E. coli* DNA polymerase I (section 2.3.1) and subsequently used as the probe for identifying the murine gene for SP-D. A commercially prepared 129/SV murine embryonic D3 stem cell genomic library constructed in lambda FIX-II (Stratagene, obtained from Prof. J. Heath), was screened with the radiolabelled probe using Y1090r$^-$ as a bacterial host. From approximately 0.5×10^6 plaques, seven positives clones were identified which hybridised to the probe under high stringency conditions (0.2 × SSC, 0.1% (w/v) SDS at 68°C) with a 3 day exposure to X-ray film. These were purified to homogeneity after quaternary screening (section 2.6). To determine which positive clones contained the full length gene for mouse SP-D, sense and antisense primers were designed at the 5' and 3' ends of each exon based on the human intron-exon structure and the mouse SP-D cDNA sequence previously obtained (see Figure 3.1 for exon-intron arrangement). PCR assay were then performed on each positive phage clone with various primer pairs to ascertain which exons were represented in each clone. Four out of the seven positive clones contained all the translated exons (Figure 3.2). The four clones were then restriction mapped (section 2.2.9.3) with several restriction endonucleases to ascertain which clone had the largest coverage of the promoter region. After analysis of the fragments by 0.8% (w/v) agarose gel electrophoresis, the fragments were transferred to nylon and probed with [α-^{32}P] radiolabelled PCR products generated from different regions of the murine SP-D gene (Table 3.1), to add both context and orientation to the restriction map (see Figure 3.2).

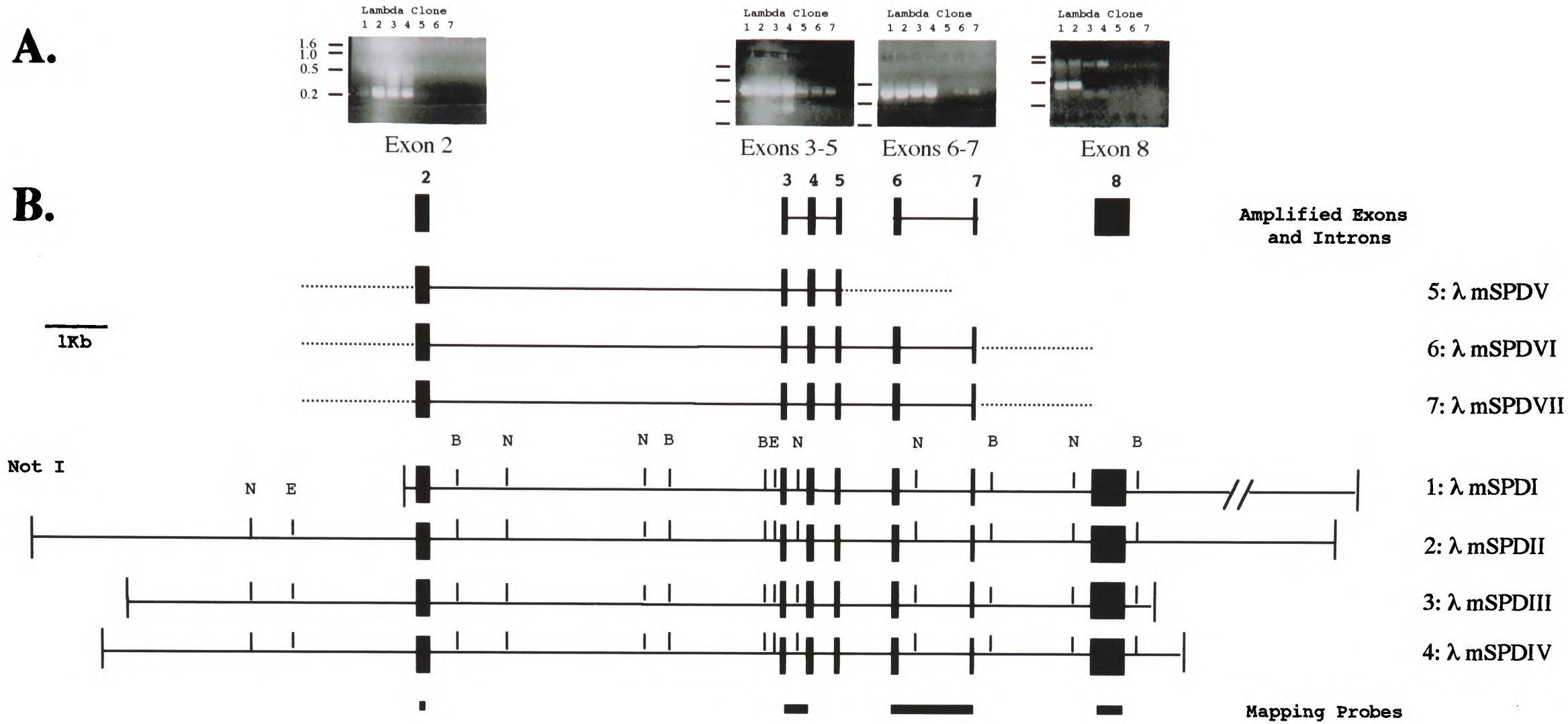


Figure 3.2. Characterisation of the bacteriophage clones for the murine SP-D gene before subcloning. **(A)** PCR analysis of clones 1 -7 using the primers listed in Table 3.1 to determine which exons were present in each clone. The molecular weight of four DNA markers are indcicated. **(B)** A diagrammatic representation of the PCR analysis. The amplified regions are indicated as parts of the final gene structure. The exon content of each clone is shown. Only clones 1 - 4 contained all protein encoding exons, these were briefly restriction mapped to determine the extent of their 5' and 3' regions. The restriction enzymes used were *Bam*H I, *Eco*R I and *Nco* I. The broken line in λmSPDI represents 4.5 Kb of DNA. The probes used in restriction mapping are indicated below.

Fragment	Fragment Size	Amplification Primers
Exon 2	188	mD1f, mD1r
Exons 3-5	890	mD2f, mD4r
Exons 6-7	1241	mD5f, mD7r
Exon 8	362	mD7f, mD7r

Table 3.1. PCR generated probes of the mouse gene for SP-D used in restriction mapping of the gene. The fragment name refers to the region of the gene covered for example Exon 3-4 covers exons 3 to exon 4 including the intervening intron. The size of the fragment generated is listed. The primers used in a standard PCR amplification are listed. An optimised annealing temperature of 53°C was used for each primer pair.

The insert from the bacteriophage clone, λmSPD-II, was excised by *Not* I restriction digestion and judged to be over 24 kb by 0.5% (w/v) agarose gel electrophoresis. To avoid problems with gel purifying large DNA fragments using conventional methods, the *Not* I digested bacteriophage clone was ligated without further purification into pBluescript, giving pBmSPD-II. A more detailed restriction map was generated for this clone of the mouse gene for SP-D for the envisaged future use of generating a null transgene construct and for use in analysis of the promoter. A higher detailed restriction map of the 3' region of the gene was established using a 10 kb *Not* I-*Eco*R I fragment from pBmSPD-II subcloned into pBluescript.

3.4. Sequencing the Murine Gene for SP-D

The approximately 24 kb genomic plasmid clone was partially sequenced using a combination of primer walking and shotgun cloning of the promoter region, followed by sequencing of these clones. Primer walking was used to sequence the entire coding region exons and the majority of the intervening introns. Primer walking was initiated with primers designed on the PCR cloned cDNA sequence, using fluorescent-dye terminator cycle sequencing.

Sequencing of the 5'-region of the gene was deemed important since it would reveal information on the promoter region of the murine SP-D gene and confirm the sequence of the first exon required for the complete mouse SP-D cDNA sequence. A 4.5 kb *Hind* III fragment containing the entire 5' region of the gene, including a few base pairs of vector sequence, was randomly sheared by sonication and 0.6-1.5 kb fragments were selected by gel purification and cloned into M13mp18 (section 2.7). Of approximately 60 clones, 48 of these clones were successfully sequenced using the M13 reverse primer, using fluorescent dye primer cycle sequencing. Remaining gaps and inconsistencies were filled by primer walking on pBmSPD-II.

Two contiguous stretches of sequence were constructed by sequencing. The first contig (accession number AF047741) is 6090 bp long and contains 3671 bp of the promoter region plus the first and second exons of mouse SP-D (shown in Figure 3.3). The first exon encodes 42 bp of the 5' untranslated region (see section 3.5) while the second exon encodes the remainder of the 5' untranslated region (3 bp), the signal peptide, N-terminal region and the first seven triplets of collagen-like sequence. The second contig (accession number AF047742) is 5593 bp long and contains the remaining 6 exons that encode SP-D. Four of the exons (3-6) are 117 bp in length and encode the remainder of the collagen region. Exon 7 encodes the coiled-coil neck region and, finally a single exon (8) encodes the C-type lectin domain and the 3' UTR. To ensure reliability of the sequence each nucleotide was sequenced at least three times on both the sense and anti-sense strands of the gene with independently primed sequences. The sequences of the two contigs, including annotations, are shown in Appendices 2 and 3.

3.5. Determining the 5' Extremity of the Mouse SP-D Gene

The 5' extremity of the mouse SP-D cDNA was absent from the published sequence obtained from conventional screening of a cDNA library (Motwani *et al.*, 1995). In order to define the 3' end of the promoter region of mouse SP-D and to generate a

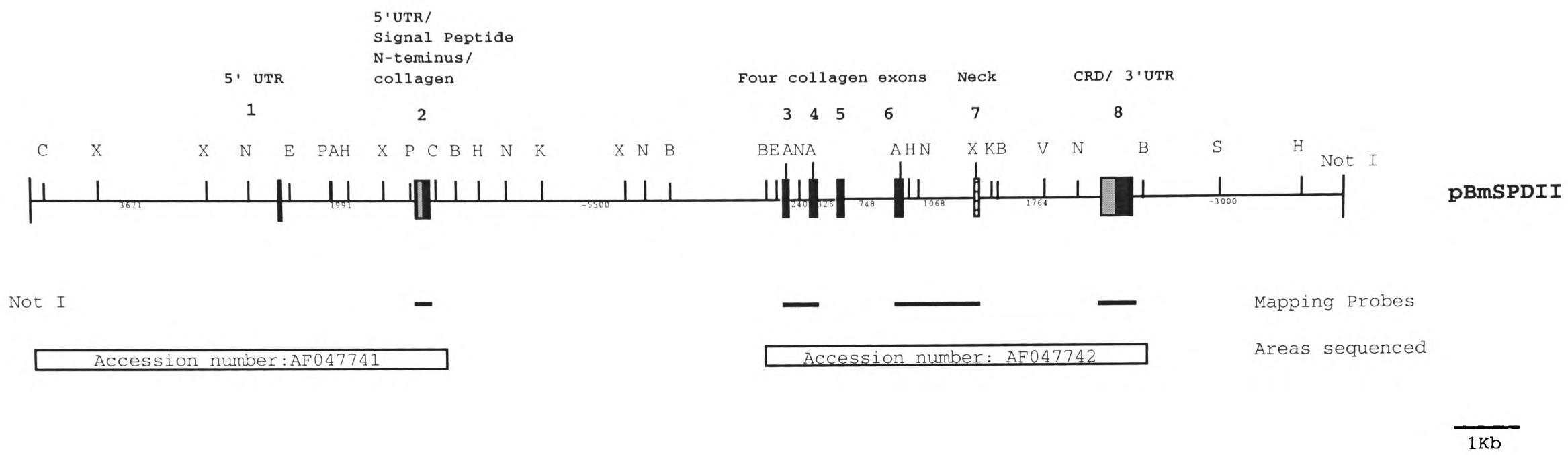


Figure 3.3. A partial restriction map of the mouse SP-D gene. The gene organisation of the mouse SP-D bacteriophage clone (λmSPDII) that was subcloned in pBluescript (pBmSPDII). The boxes represent exons, with the name and number of the exon listed above. Lines represent introns with the sizes in base pairs shown below. Bold horizontal bars indicate the probes used in Southern blotting. Open horizontal bars indicate the areas sequenced together with the accession number. Vertical bars represent the positions of restriction sites, abbreviated to a single letter (underlined): Apa I, Bam HI, Cla I, Eco RI, Hind III, Kpn I, Nco I, Pst I, Sma I, Eco RV, Xba I. The Not I restriction sites from the lambda FIX II polylinker are shown. A pBluescript subclone of the 10 Kb Eco RI-Not I fragment was used in defining the restriction map in the 3' end of the gene.

complete gene sequence, the 5' sequence of the murine SP-D transcript was obtained through a PCR cloning approach: 5' rapid amplification of cDNA ends (RACE), termed 5' RACE (section 2.8.3).

A single stranded DNA template was generated from mouse (BALB/c) lung total RNA using Moloney leukeamia virus reverse transcriptase (section 2.5.2) primed with an anti-sense oligonucleotide specific for exon 7 of SP-D (mD6r). The SP-D enriched single stranded DNA was tagged with a tail of deoxycytidine using terminal deoxynucleotidyltransferase (section 2.3.2). PCR amplifications were performed for 30 cycles on the tailed single stranded DNA template using an 'anchor' primer[2] ($XSCdG_{17}$) and an anti-sense primer specific to exon 3 (mD2r). A second, nested[3] PCR was performed to introduce a higher level of stringency to the amplification by using an anti-sense oligonucleotide specific for the 3' end of exon 2 (mD1r) and a sense oligonucleotide (XSC) that anneals to the anchor sequence. The 5' RACE yielded the expected product of approximately 250 bp. The RACE products were subcloned into pBluescript and a total of ten independent clones were sequenced. Seven independent subclones of the longest sequences of 242 bp, which were identical to the second exon of SP-D, were identified. This was followed by an extra 5' 42 bp sequence, which was identical to a region of sequence located in the 5' region of the SP-D gene, now termed exon 1. It also showed striking homology to the published sequences for the human and rat SP-D 5' UTRs (Lu *et al.*, 1992; Rust *et al.*, 1996). This was consistent with this extra sequence being the 5' extent of the mouse SP-D gene transcribed in the lung. Three other subcloned 5' RACE products were identified, two of which initiated at position 24 and a second at position 31 of the published cDNA. These are probably the result of RNA degradation although

---

[2]An anchor primer is a primer that will introduce a novel piece of known sequence at the end of a cDNA to allow subsequent amplification. In this case GACTCGAGTCGACATCGAT was added: a complementary primer termed XSC anneals to the anchor sequence during amplification.
[3]A nested PCR, is a second PCR performed upon a PCR product. One or both of the second set of PCR primers attach 3' of the first amplification primers; that is attach to amplified sequence rather than primer derived sequence.

initiation of transcription at another point can not be completely ruled out. The complete sequence of mouse SP-D is shown in Appendix 4.

3.6. The Presence of the SP-D Gene in Other Species

Innate immune defence is a primitive immune system present both in higher animals and simpler organisms, unlike the adaptive immune system which is only present in vertebrates. Consequently, an important area of research in innate immunity is investigating the immune system of lower organisms to understand the more elaborate immune system of higher organisms. To detect the presence of the SP-D gene in other species and to examine the emergence of SP-D, both low stringency hybridisation using part of the murine cDNA and low stringency PCR were performed across the carbohydrate recognition domain (CRD) on genomic DNA from a range of animals.

3.6.1. Zooblot

Genomic DNA was extracted from either soft tissue or whole blood from several different species of animals (section 2.2.5.3). Approximately 10 μg of genomic DNA was subjected to endonuclease digestion using *Bam*H I followed by 1% (w/v) agarose gel electrophoresis and transfer to a nylon membrane. The membrane was probed with an [α-^{32}P] labelled mouse SP-D probe. The template used to generate the probe was a PCR product derived from pBmSPD-II, spanning part of exon 8 using the primers mD7f and mD7r in a standard PCR assay (section 2.8.1) with optimised conditions (see Table 3.1). The membrane was washed under medium stringency conditions (0.1 × SSC, 0.1% (w/v) SDS at 50°C) and exposed to X-ray film for 32 hours.

Under medium stringency conditions, cross hybridisation of the murine probe was seen with a single DNA fragment within all the mammals tested (human, cattle, sheep and rabbit), birds (goose) and amphibians (frog) which indicated that single copy

SP-D-like gene existed in all these mammals (Figure 3.4). The genomic probe was considered specific, since it hybridised to a single band in each species: this indicated no cross reactivity against the other collectin genes was present. This was especially evident in the case of the cattle, which have two additional collectins (CL-43 and conglutinin) that bear the highest identity to SP-D. Conglutinin and CL-43 shares about 71% and 72% sequence identity to mouse SP-D over the CRD at the cDNA level.

3.6.2. Low stringency PCR on different species and direct sequencing

Low stringency PCR was performed on genomic DNA from several different species to clone SP-D. A pair of primers (mD7f and mD7r) were used that amplified the single exon encoding the CRD of murine SP-D. Approximately 1 μg of genomic DNA isolated from each species was used as a template in a standard PCR assay (section 2.8.1). The PCR assays were optimised to generate a discrete band at approximately 370 bp using four different temperatures in the range of 42°C to 55°C and four different magnesium concentrations ranging from 1.0 mM to 1.75 mM. Discrete products were obtained from the positive control (mouse), sheep (ovine), goose, trout and horse genomic DNA. These fragments were gel purified and sequenced directly using fluorescent DNA sequencing with the amplification primers. The sequence of the positive control PCR from murine genomic DNA was identical to murine SP-D. The sequences of PCR products from goose and horse showed no similarity to sequences in GenBank and the trout sequence showed identity to the rainbow trout mRNA for SoxP1, unrelated to the collectins (accession number D83256). Of more interest was the sheep PCR derived genomic sequence that showed 91% identity over 332 bps to the bovine collectin, CL-43. Protein, cDNA or genes for CL-43 and conglutinin have only been reported for cattle. The role of CL-43 is unknown, it is the only collectin that does not exist as higher order oligomers of trimeric subunits. This is the first indication that one of the bovine collectins is present in other members of the bovidae: sheep. The partial gene

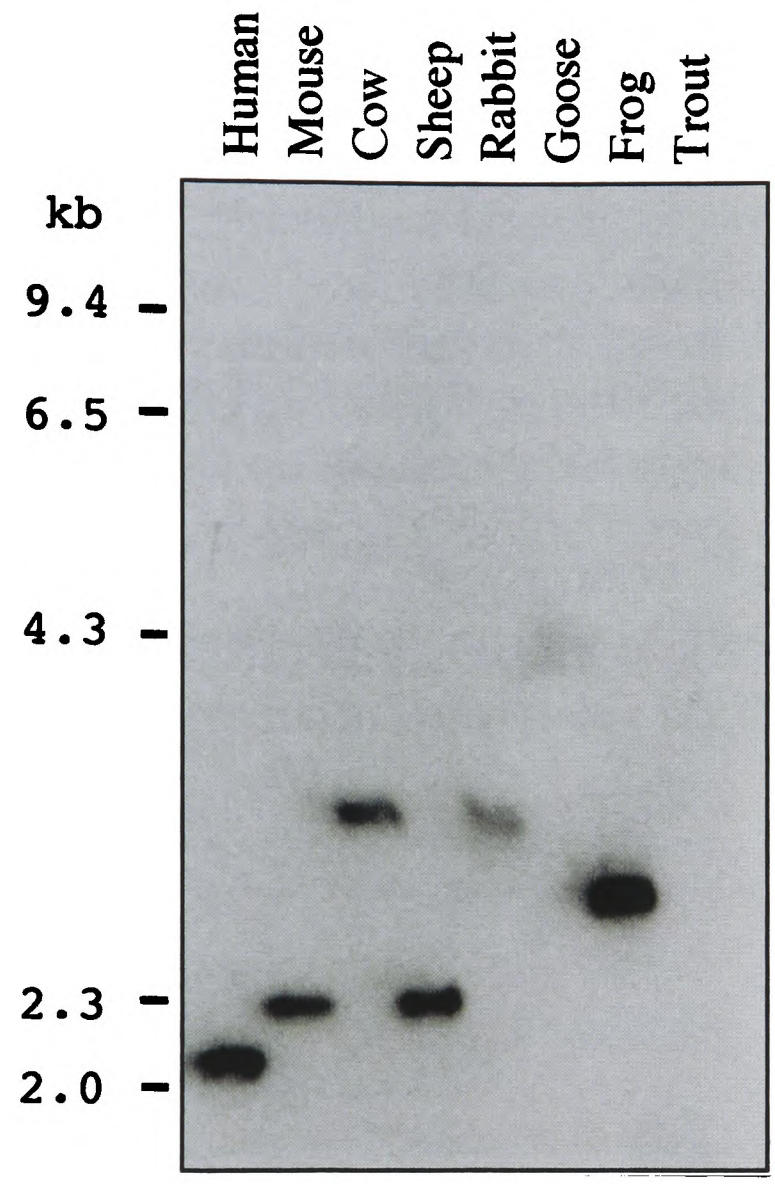


Figure 3.4. Zooblot: Southern blot analysis of genomic DNA from different species. Ten μg of genomic DNA from eight different animals was digested with *Bam*H I, separated on a 1 % (w/v) agarose gel then transferred to nylon membrane and probed with radiolabelled DNA encoding the CRD exon of mouse SP-D. A single band of varying molecular size is seen in all the species of animals examined, except for fish. The molecular sizes are indicated on the left.

sequence of ovine CL-43 is shown aligned against the other known bovine collectins (MBL, SP-D, Conglutinin and CL-43) as well as to human SP-A (Figure 3.5B and C). The phylogenetic relationship to the other bovine collectins shows the similarity to bovine CL-43 (see Figure 3.5D). RT-PCR assays, northern blotting or tissue staining would be required to determine if this gene is transcribed or the mRNA is translated in sheep.

The use of low stringency PCR can be quite limited when using a single set of primers. For example, bovine SP-D was not amplified, notably because of 3' mismatch within the second and third nucleotides of both primers. However, the technique could be made more amenable for cloning species homologues of the collectins. Improvements could be made by using more than one set of primers, designing primers at regions of higher conservation or using codon degeneracy at the 3' ends of primers. However, this technique has discovered the first species homologue for CL-43.

3.7. Dispersed Repetitive Elements in the Murine SP-D Gene

The larger size of the murine SP-D intron 5 compared to that of the human gene (see Figure 3.6) could be attributed to an insertion of a 360 bp dispersed repetitive element. The presence of this repetitive element was identified by comparing the murine SP-D gene to the GenBank database. The repetitive element in the murine SP-D gene (*Sftpd*) was responsible for a number of surprising homologies to other rodent genes, including those for manganese superoxide dismutase (accession number L35525) and murine IgE binding lectin (accession number M97896). The identity of the repetitive element was determined by screening the gene against the internet database of repetitive elements (*ftp.genome.washington.edu/cgi-bin/RepeatMasker*). The element was identified as an ORR-1 (Caddle *et al.*, 1990) which is a subgroup of rodent specific mammalian apparent long terminal repeat (LTR) retrotransposons (MaLRs) (Smit, 1993). The murine SP-D MaLR was shown to have the greatest

A.

```
 E  K  I  F  K  T  S  G  V  V  K  S  Y  S  D  A  X  Q  V  C  R  E  X  T  G  Q  L  P
GAGAAGATCTTCAAGACGTCAGGTGTTGTAAAATCATATTCGGATGCCSXGCAGGTCTGCAGGGAGGSTACGGGACAGCTGCCC

 S  P  R  S  A  A  E  N  E  A  V  T  Q  L  V  R  A  K  N  K  H  A  Y  L  S  M  N  D
TCCCCACGCTCTGCAGCCGAGAATGAGGCCGTGACACAGCTGGTCAGAGCCAAGAACAAGCATGCTTACCTGAGCATGAATGAY

 I  S  R  E  G  R  F  T  Y  P  T  G  X  S  L  V  Y  S  N  W  A  P  G  E  P  N  N  R
ATCTCCAGAGAGGGCAGGTTCACCTACCCTACGGGGCMGTCACTGGTCTATTCCAACTGGGCCCCTGGGGAGCCCAACAATAGG

 A  K  D  K  G  P  E  N  C  L  E  I  Y  S  N  G  N  W  N  D  I  E  C  R  E  E  R
GCCAAAGACAAAGGCCCAGAGAACTGTCTTGAGATCTATTCTAATGGCAACTGGAATGACATAGAATGCCGGGAGGAGCGC
```

B.

```
   175  .         .         .         .         .         .         .         .
sh          ...EKIFKTSGVVKSYSDAXQVCREXTGQLPSPRSAAENEAVTQLVRAKNKHAYLSMNDISREGRFTYPTGXS
43 VLFPDGQAVGEKIFKTAGAVKSYSDAEQLCREAKGQLASPRSSAENEAVTQLVRAKNKHAYLSMNDISKEGKFTYPTGGS

   255  .         .         .         .         .
sh LVYSNWAPGEPNNRAKDKGPENCLEIYSNGNWNDIECREER...
43 LDYSNWAPGEPGNRAKDEGPENCLEIYSDGNWNDIECREERLVICEF
```

C.

```
   175  .         .         .         .         .         .         .         .
sh          ...EKIFKTSGVVKSYSDAXQVCREXTGQLPSPRSAAENEAVTQLVRAKNKHAYLSMNDISREGRFTYPTGXS
43 VLFPDGQAVGEKIFKTAGAVKSYSDAEQLCREAKGQLASPRSSAENEAVTQLVRAKNKHAYLSMNDISKEGKFTYPTGGS
BK VLFPDGQAVGEKIFKTAGAVKSYSDAEQLCREAKGQLASPRSSAENEAVTQMVRAQEKNAYLSMNDISTEGRFTYPTGEI
bD MLFPNGRSVGEKIFKTVGSEKTFQDAQQICTQAGGQLPSPRSGAENEALTQLATAQNKAAFLSMSDTRKEGTFIYPTGEP
bM LIFSLGKRVGKKAFFTNGKKMPFNEVKTLCAQFQGRVATPMNAEENRALKDLVTEE---AFLGITDQETEGKFVDLTGKG
hA SLQGSIMTVGEKVFSSNGQSITFDAIQEACARAGGRIAVPRNPEENEAIASFVKKYNTYAYVGLTEGPSPGDFRYSDGTP

   255  .         .         .         .         .
sh LVYSNWAPGEPNNRAKDKGPENCLEIYSNGNWNDIECREER...
43 LDYSNWAPGEPGNRAKDEGPENCLEIYSDGNWNDIECREERLVICEF
BK LVYSNWADGEPNNSD-EGQPENCVEIFPDGKWNDVPCSKQLLVICEF
bD LVYSNWAPQEPNN---DGGSENCVEIFPNGKWNDKVCGEQRLVICEF
bM VTYQNWNDGEPNN---ASPGEHCVTLLSDGTWNDIACSASFLTVCEFSL
hA VNYTNWYRGEPA----GRGKEQCVEMYTDGQWNDRNCLYSRLTICEF
```

D.

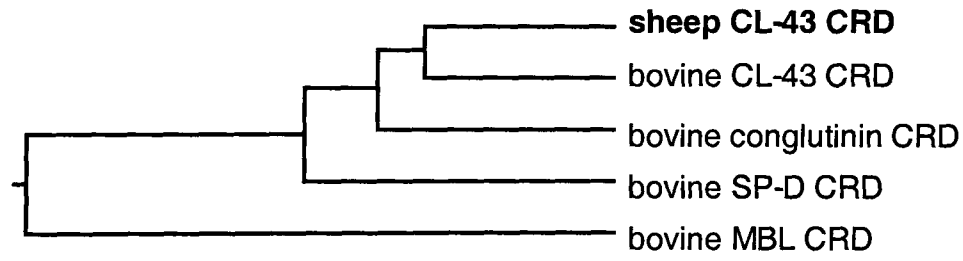
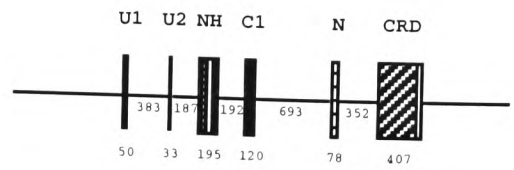


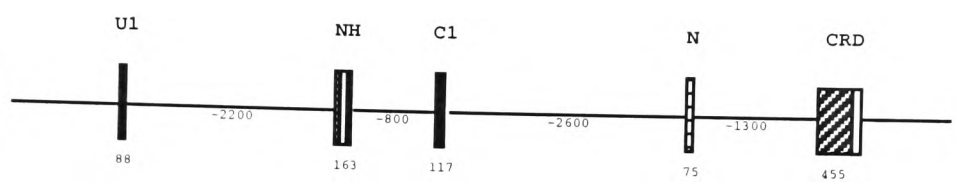
Figure 3.5. Sequence analysis of sheep CL-43. (A) The region of the partial, putative sheep CL-43 gene encoding part of the CRD. The nucleotide (lower) and derived amino acid (upper) sequences of the only full-length open reading frame (ORF) are shown. Alignments of **(B)** the putative sheep CL-43 (sh) ORF and bovine collectin CL-43 (43) CRD region, as well as **(C)** bovine SP-D (bD), MBL (bM) and conglutinin (BK) and human SP-A (hA) CRD peptides. Sequences are numbered according to bovine CL-43. Residues conserved between the putative sheep CL-43 and other species are shown in bold. Dashes have been introduced to maximise the alignments, which were constructed using Bestfit from the Wisconsin package (section 2.9). Dots indicate the beginning and end of the sheep CL-43 sequence obtained. **(D)** Phylogenetic relationship of the fragment of sheep CL-43 and the CRDs from the bovine collectins. Human SP-A was excluded as bovine MBL formed a suitable out-group.

Mouse SP-A

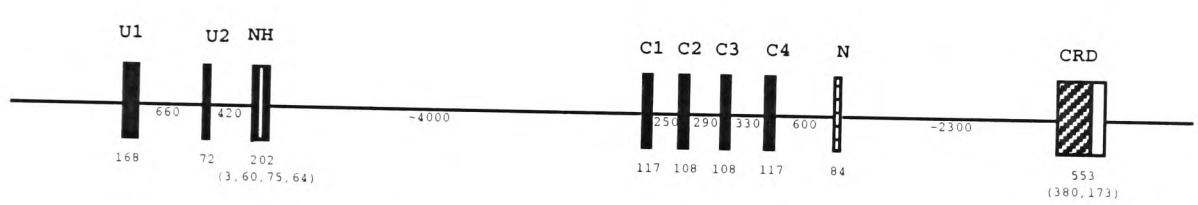
1 kb



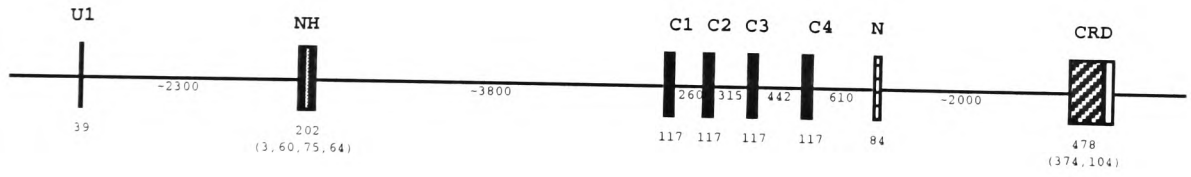
Mouse MBL-A



Bovine Conglutinin



Human SP-D



Mouse SP-D

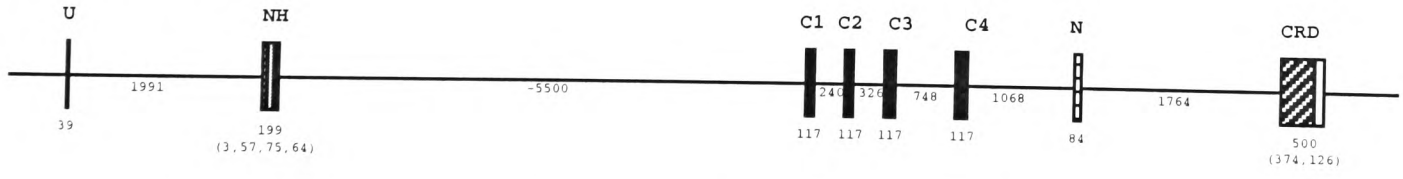


Figure 3.6. The organisation of the mouse SP-D gene in comparison to other collectins. The boxes represent exons, with the name and number of the exon listed above. Lines represent introns with the sizes in base pairs shown below. The exons are indicated; **U**, untranslated; **NH**, amino terminal, cysteine rich and beginning of the collagen region; **C1-C4**, collagen encoding exons; **N**, neck exon; **CRD**, CRD exon.

similarity to the consensus sequence of ORR-1A (accession number U17093), displaying typical hallmarks of a MaLR (Smit, 1993). This is shown in Figure 3.7. Like most MaLRs, the *Sftpd* ORR-1 is a solitary LTR, with the flanking sequences showing no similarity to those around other repeats (Kelly, 1994).

The MaLR element found in murine SP-D is of particular interest because it contains five minisatellites (CAAAA) (Figure 3.7). This type of organisation resembles the two hypervariable minisatellite loci present within MaLR elements in mice (Kelly, 1994) and in two, human MaLR type elements, *MstII* elements (Mermer *et al.*, 1987). The microsatellites present in these elements have been shown to be highly polymorphic, theoretically because of their localisation within MaLR elements (Kelly, 1994). The two rodent MaLRs show a high degree of polymorphism with between 500 to 2000 minisatellite arrays and are believed to be the site of germline and somatic recombination events (Kelly *et al.*, 1991; Gibbs *et al.*, 1993).

3.8. Polymorphisms in the Murine SP-D Gene

From the sequence of the gene, three repeat regions were discovered. Two dinucleotide repeats (CA) and a polyadenosine stretch in the 3' UTR (shown in Figure 3.8A). The repeats were investigated for simple length polymorphisms. Primers were designed to generate PCR products in the order of 150 bp including the repeat region (see Table 3.2).

3.8.1. Detection of polymorphisms by PCR

To determine if the repeats were polymorphic, the size of the repeat was measured by PCR assay across the repeat from a number of different inbred strains of laboratory mice (provided by Yvonne Boyd, MRC Mammalian Genetics Unit, Harwell, UK). A standard PCR was performed for 30 cycles, using an annealing temperature of 55°C, 58°C or 60°C (see Table 3.2) with 25 ng of murine genomic DNA in a 25 µl reaction. The size of the amplified repeat was then measured by running approximately 5 µl of

```
ORR1A                                                          TGTG 4
                                                               ||||
sftp4   CTCCACAGCACAGGCTTTACTTAAAGGGTAAATTTTGGAAGGGATCTGTG 1666
                .         .         .         .         .
ORR1A   ATGGTTTGTATATGCTCGGACCA.GGGAGTGGCACTATCTGRAGGTGTGG 53
        |||||||||||||| || |||||| |||||||||| || ||:||||||||
sftp4   ATGGTTTGTATATTCTTGGACCACGGGAGTGGCACCATTTGGAGGTGTGG 1616
                .         .         .         .         .
ORR1A   CCTT..................GTTGGAGTAGGTGTGTCACTGTGGGCG 84
        ||||                  ||||||||||||||||||||||||  |
sftp4   CCTTGTTGGAATTGGGTGACCTGGTTGGAGTAGGTGTGTCACTGTGGATG 1566
                .         .         .         .         .
ORR1A   TGGGCTTTAAGACCCTCACCCTAGCTGCCTGGAAGTCAGTCTTCTSCTAG 134
        ||||| |||||| |||||||||| | |||||||||||| ||   ||||
sftp4   TGGGC.TTAAGATCCTCACCCTAGTTTCCTGGAAGTCAGTTTTTCACTAG 1517
                .         .         .         .         .
ORR1A   CAGCCTTCGGATGAAGATGTAGAACTCTCAGCTCCTCCTGCACCATGCCT 184
        ||||||| |||||||||  |||||||||||||||||||||||||||| |||
sftp4   CAGCCTTTGGATGAAGACATAGAACTCTCAGCTCCTCCTGCACCATTCCT 1467
                .         .         .         .         .
ORR1A   GCCTGGATRCTGCCATGCTCCCGCCTTGATGATAATGGACTGAACCTCTG 234
        ||||||||:|| || ||||||| |||
sftp4   GCCTGGATGCTACCCTGCTCCCACCT.....................A 1440
                .      Poly A signal.         .         .
ORR1A   AACCTGTAAGCCAGCCCCAATTAAATG..................... 261
        ||||||||||||| ||||||||||
sftp4   AACCTGTAAGCCAGGCCCAATTAAATGTGTGGTTTTGTTTTGTTTTGTTT 1390
                .         .         .         .         .
ORR1A   .....................TTGTCCTTTATAAGACTTGCCTTGGTC 288
                             |     |  ||||||||||||||| ||
sftp4   TGTTTTGTTTTGTTTTGTTTTTATACATATAAATAAGACTTGCCTTGATC 1340
                .         .         .         .         .
ORR1A   ATGGTGTCTGTTCACAGCAGTAAAACCCTAACTAAGACA            327
        |||||||||||||||||||||||||||||||||||||
sftp4   ATGGTGTCTGTTCACAGCAGTAAAACCCTAACTAAGACAGGATCTAGGCC 1290
```

Figure 3.7. The inserted mammalian apparent LTR. The retrotransposon from intron 5 of the mouse SP-D gene (bottom) is aligned with the consensus sequence for ORR1-1A (top). The alignment was performed using GAP of the GCG package. Gaps (.) were inserted to maximise similarity. Nucleotide numbering refers to the mouse SP-D gene (AF047742). The sequence of the mouse SP-D gene has been reversed and complemented. The sequence of the minisatellite repeat is outlined. Characteristic features of the MaLR LTR include the terminal inverted repeats TGT and ACA (underlined), a 5 bp direct repeat surrounding the MaLR (bold) and polyadenylation signal (underlined AATTAAA).

the reaction product on a 6% (w/v) non-denaturing polyacrylamide sequencing gel (section 2.2.6.2) and visualised by silver staining (section 2.10.1.3).

Repeat	Primers	Sequence of the Amplified Repeat	Tm (°C)
CA1	CA1f	CAGAGTGTAGGTGGAGTAAGAGTAGGTGAGTCAACAGGAAAAAATTGTGT	55
	CA1r	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGACTTTTCTCAGAA	
		GACCAATGCTGGGTAGCTCCACTTACAGGGCAGGCATAAATAGCC	
CA2	CA2f	TACACAGTAACTAATATTATCATGAAAAGTCAGAGTTTTGTTTTAGGGGT	60
	CA2r	AGCTCCGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	
		GTGTGTTCAAGCTAGAGGTTGATATCAGATGTT	
PolyA	pAf	AATGATAAGGCTTGTGGAGAGCAGCGCCTTGTTATCTGTGAGTTCTGAGC	58
	pAr	CACTGGAGGTAAAAAAAAAAAAAAAAAAAAAACAGGACAATATTTGGCCA	
		GGAGTTTAACCAGTGTACTATGGCTCAGACCTGTATGTTGCCAACATCC	

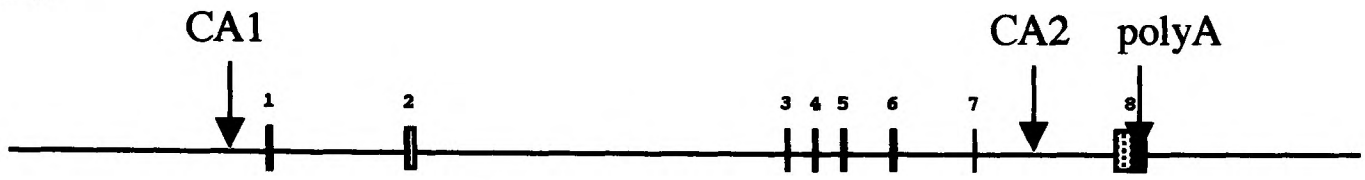
Table 3.2. Information on the repeat and PCR conditions used to amplify the polymorphic repeats in the murine gene for SP-D. The underlined sequence indicates the location of the primers.

3.8.2. The CA Repeats

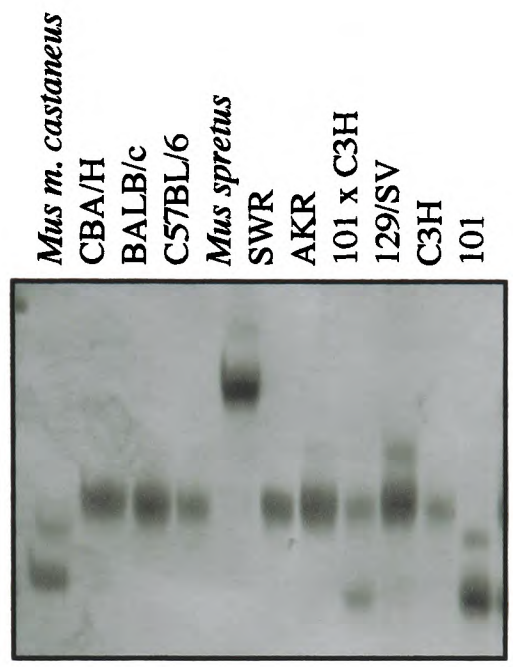
The first CA repeat found in the promoter region contained 23 CA units in the 129/SV SP-D gene. This repeat is also present in the rat but absent from both the human SP-D gene and the bovine conglutinin gene (Kawasaki *et al.*, 1994; Rust *et al.*, 1996). The second CA repeat found in 129/SV*Sftpd* contained 22 CA units and was present in intron 7. It is not known if a similar repeat is present in the rat and human SP-D genes, however, a CA repeat is present in the analogous intron of rat MBL-A (Drickamer and McCreary, 1987).

Both of the dinucleotide repeats in *Sftpd* were shown to be polymorphic (Figure 3.8C). The polymorphic CA repeat in the promoter region had three different alleles across seven different mouse strains. The CA repeat in intron 7 displayed a higher degree of polymorphism and variation (see Figure 3.8C) with five different alleles observed across the seven mouse strains studied.

A.



B.



C.

	CA1	CA2	Polyadenosine
Mus m. castaneus	0	18	0
CBA/H	5	6	3
BALB/c	1	6	3
C57BL/6	1	7	3
Mus spretus	5	0	8
SWR	-	-	3
AKR	-	-	3
129/SV	1	7	3
C3H	-	-	3
101	-	-	0
DBA	5	5	-

Figure 3.8. The nucleotide and dinucleotide repeats. (A) The location of the repeats, the CA repeats (CA) and the polyadenosine repeat (polyA). (B) Analysis of the polyadenosine repeat across different mouse strains ran upon a 6 % (w/v) acrylamide gel, silver stained. Note, the heterozygote (101 x C3H) with an allele from each parent. (C) The size variation of the three different nucleotide repeats in murine SP-D across different inbred mouse strains. Dash indicates not determined.

3.8.3. The Polyadenosine Repeat

The polyadenosine tract within the 3' UTR of mouse SP-D was found to be polymorphic, showing three different alleles across ten strains of mice (Figure 3.8B). This repeat is the least polymorphic of those found in *Sftpd*, but shows the greatest variation, being between 19 and 27 bp in length. The significance of a second polyadenosine tract close to the polyadenylated tail of a mRNA upon RNA stability or translation is not known.

3.8.4. The Implications of Repeats

The CA repeat in the promoter region may also prove to be interesting at the protein level. Dinucleotide repeats within promoter regions may be responsible for intraspecies phenotypic variation (Kashi *et al.*, 1997), as repeat array variations have been shown to modulate *in vitro* expression of protein (Hamada *et al.*, 1984). It would be interesting to determine if variation in repeat numbers with the promoter region causes quantitative variations of SP-D within an outbred population of mice. Repeats regions in the 3' UTR are believed to act a potential source of quantitative variation in protein levels that occur at the post-transcriptional level (Kashi *et al.*, 1997).

3.9. Chromosomal Localisation Of Murine Gene for SP-D

To localise the gene for murine SP-D to a chromosome, the simple sequence length polymorphism (SSLP) in the promoter region was used in genetic mapping. A random selection of genomic DNA samples from 50 mice from the European collaborative interspecific backcross (EUCIB) panel (Breen *et al.*, 1994) was obtained from the human genome mapping project resource centre (Cambridge, UK). The animals used were from a C57BL/6 × *Mus spretus* F1 generation backcrossed with either of the parental strains. PCR was carried out as before (section 3.8.1). The inheritance of the SSLP flanking *Sftpd* CA1, was used to determine the chromosomal

location and fine scale mapping position of the gene with respect to known EUCIB markers on the mouse genome. Gene order, linkage to the EUCIB anchor loci and statistical analysis was carried out on the MBx database (*www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html*) and by visual inspection of the data. Initial genotyping confirmed an earlier report that *Sftpd* lies in the central region of mouse chromosome 14 (Motwani *et al.*, 1995). The EUCIB data indicates that the murine gene for SP-D is linked to *D14Mit45* (χ^2 value of 71.14) and *D14Mit60* (χ^2 value of 27.14) (Figure 3.9).

For fine scale mapping, a panel of informative recombinant mice for *D14Mit45-D14Mit5* was analysed. Analysis of the critical recombinants, while minimising the number of double recombinants, established the locus order *D14Mit45-D14Mit212-Sftpd-D14Mit141-D14Mit60* (Figure 3.9). Of the 27 scored recombinant mice between *D14Mitt45* and *D14Mit60*, 16 were recombinant between *D14Mit45* and *Sftpd*, while 11 were recombinant between *Sftpd* and *D14Mit60,* showing tighter linkage of *Sftpd* to *D14Mit60* on the genetic map.

3.10. Genetic Mapping of The Collectin Locus

The principal reason for mapping the murine collectin locus (consisting of MBL-A, SP-D and SP-A, see section 1.3.2.6) was to determine if double and triple null animals for the murine collectins could be made by crossbreeding different single transgenic mice of each collectin. The EUCIB panel of mice was utilised to investigate if recombination occurred between collectin loci present in this panel of 1,000 backcross mice. If the three genes co-segregated across the 1,000 mice then it could be assumed that it was technically impractical that a SP-A (-/-) × SP-D (-/-) cross would lead to double knockouts. This assumption excludes possible variation in crossover events between different inbred mice strains.

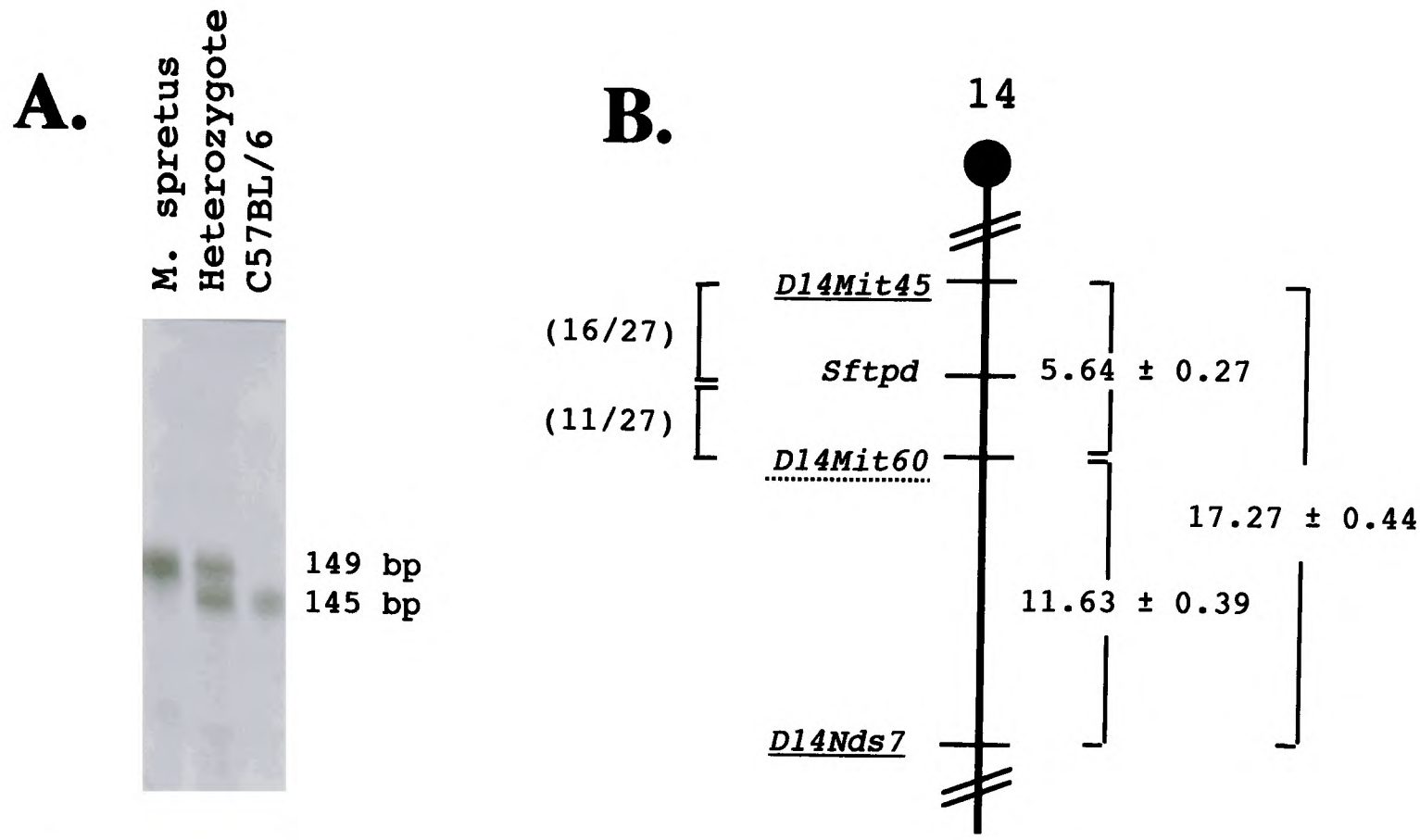


Figure 3.9. Chromosomal mapping of the mouse SP-D gene. (**A**) 6 % (w/v) denaturing polyacrylamide gel showing the PCR products generated for haplotyping; alleles for *M. spretus*, heterozygote and C57BL/6 are shown and their sizes are indicated. (**B**) Mouse chromosome 14. Recombination distance and standard error are shown in centimorgans (cM) between EUCIB reference (underlined) and secondary reference markers (dotted underline). The fraction of recombinant mice in the fine mapping panel (*D14Mit187* - *D14Mit45*) that have been scored for *Sftpd* is shown to the left. (**C**) Haplotype analysis of critical recombinants. Animals used were (C57BL/6 x *M. spretus*) backcrossed with either parental strain. Black boxes represent heterozygotes; white boxes represent homozygotes for either *M. spretus* or C57BL/6 alleles.

To genetically map the murine SP-A and MBL-A genes onto the EUCIB panel of mice, polymorphic markers had to be found. Examination the sequences of these genes showed that potential SSLPs, such as polynucleotide tracts, di, tri or tetramer repeats were absent. Since SSLP could not be used to map these genes, a different approach was employed. SP-A and MBL-A were mapped by first identifying restriction length polymorphisms between the two strains. SP-D was also mapped by this method for confirmation of authenticity of the procedure.

To accomplish this, polymorphic restriction sites within amplified gene fragments were used to define the two parental mouse strain alleles. The restriction pattern of the backcrossed mice could then be determined by monitoring the inheritance of the different alleles. Briefly, PCR primers were designed at the 3' and 5' ends of two adjacent exons to amplify the two exons and the intervening intron. The regions of amplification were chosen so that they had introns of approximately 1 kb, as judged by the sequence of each gene or restriction maps generated by others (Korfhagen *et al.*, 1992; Sastry *et al.*, 1995). The intron size of 1 kb was selected because it was technically difficult to obtain sequence by direct fluorescent-dye sequencing of larger PCR products, which were required for confirmation of amplified products. Secondly, based upon empirical evidence, 1 kb of intron DNA was sufficient to discover a restriction length polymorphism between two different mice strains, using approximately 41 restriction enzymes.

The standard PCR assay for each gene was optimised for magnesium concentration (1.0-2.0 mM) and the annealing temperature (55-67°C) to generate a single discrete band of the expected size (see Table 3.3). The PCR assay was performed for 30 cycles on 1 μg of genomic DNA from the parental mouse strains (C57BL/6 and *Mus spretus*). The PCR products from each gene were partially sequenced (section 2.7.2) and shown to be identical to the published gene sequences.

Polymorphic restriction sites were found by screening the amplified fragment from both parental mice with 41 different restriction enzymes (see Table 5.3) and comparing the restriction pattern on a 1% (w/v) agarose gel (see Figure 3.10.A). For convenience and speed, 10 µl of the PCR assay was cleaved directly, without purification, in a 20 µl reaction containing 10 U of restriction enzyme. A *Bst*E II polymorphism was found in the SP-D gene and SP-A contained a *Pvu* II polymorphism (see Figure 3.10A and Table 3.3). The same restriction pattern was observed for both the purified and unpurified PCR products. The fragment amplified from MBL-A gene from *Mus spretus* and C57BL/6 differed in length, this length polymorphism could be used to map *Mbl1* (see Figure 3.10A).

Gene	Primers	Tm	Time	[Mg^{2+}]	Enzyme	C57BL/6	MS
Sftpa	mAnif	66°C	60	1.25	*Pvu* II	1600	1700
	mAnlr					+100	
Sftpd	mD2f	65°C	120	1.5	*Bst*E II	1500	1800
	mD5r					+ 300	
Mbl1	mMBL1f	65°C	90	1.5	none	1700	1600
	mMBLr						

Table 3.3. The PCR-RFLP used to localise the three collectin genes; SP-D, SP-A and MBL-A. Conditions used to amplify the genes (primers, annealing temperature (Tm), elongation time in seconds (time) and magnesium concentration ([Mg^{2+}]). The restriction enzyme used and the approximate size of the allele from each inbred mouse strain after endonuclease digestion is listed (MS, *Mus spretus*)

Using genomic DNA from 50 random mice of the EUCIB backcross, the presence of each allele for the three different murine collectin genes was assessed. *Sftpa*, *Mbl1* and *Sftpd* were shown to map to the middle of chromosome 14, thus confirming the chromosomal mapping result obtained for *Sftpd* using the SSLP (section 3.9). Analysis of fine scale mapping for each gene revealed that *Sftpa*, *Mbl1* and *Sftpd* all co-segregated (see Figure 3.10).

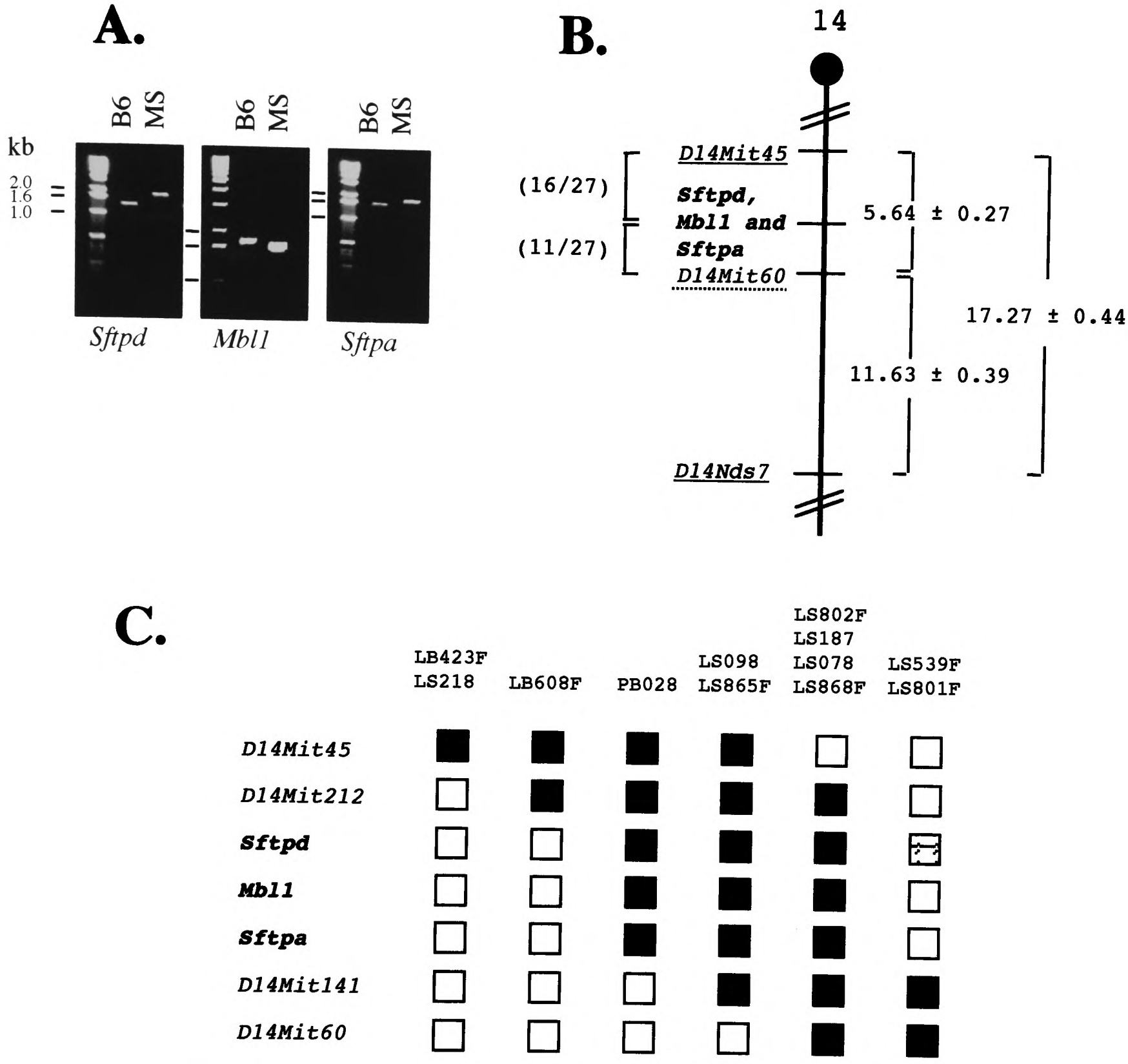


Figure 3.10. Chromosomal mapping of the murine collectin cluster. (**A**) 1-1.25 % (w/v) agarose gels showing the PCR or PCR-RFLP products of *Sftpd*, *Mbl1* and *Sftpa* generated for haplotyping (see Table 3.3 for details). *M. spretus* (MS) and C57BL/6 (B6) alleles are shown. Three molecular weight markers are indicated **(B)** Mouse chromosome 14. Recombination distance in centimorgans (cM) between EUCIB markers and the collectin cluster. For details of the figure see figure 3.9B. (**C**) Haplotype analysis of critical recombinants. The three collectin genes; *Sftpd*, *Mbl1* and *Sftpa* all co-segregate on the EUCIB backcross. For details of the figure see figure 3.9C.

Chromosomal mapping of the three collectin genes, which constitute the collectin locus in mice, indicates that the locus lies on the middle of chromosome 14, on a section of DNA that is smaller than the resolution of the EUCIB genetic map, less than 0.3 cM (Breen *et al.*, 1994). Based on the use of the EUCIB panel of mice, which represents 1000 backcrossed mice, no recombination breakpoints were observed within the collectin locus. This result indicates that crossbreeding mice containing null alleles of the different murine collectins genes to generate double or subsequently triple null collectin locus transgenic animals is impractical.

3.11. Analysis of the Promoter Sequence of Murine SP-D

The sequence around the first nucleotide of the transcribed cDNA for *Sftpd* fits the consensus sequence for the initiation of transcription (YNNNYAYYYYY) (Breathnach and Chambon, 1981). This is consistent with the transcription start site identified in this work by 5' RACE and that identified in humans and rats (Rust *et al.*, 1996).

3.11.1. The Promoter Region Contains LINEs and SINEs

Using the internet database of repetitive elements (section 3.7) in the initial analysis of the promoter region, the 5' sequence flanking the transcription start was found to be clearly divided into two as regards the location of repetitive sequences. The more distal, upstream half, -3671 to -1809 was found to contain several different dispersed repetitive elements, occupying 55% of the sequence, including three short interspersed repetitive elements (SINEs) located upstream of three long interspersed repetitive elements (LINEs). The sequence immediately upstream of the transcription start site was devoid of any known dispersed repetitive elements (see Figure 3.11).

3.11.2. Conservation in the Promoter Region

The size of the functional promoter region for murine SP-D could be ascertained by using dotplot (Wisconsin package, Genetics Computer Group, Wisconsin, USA) to

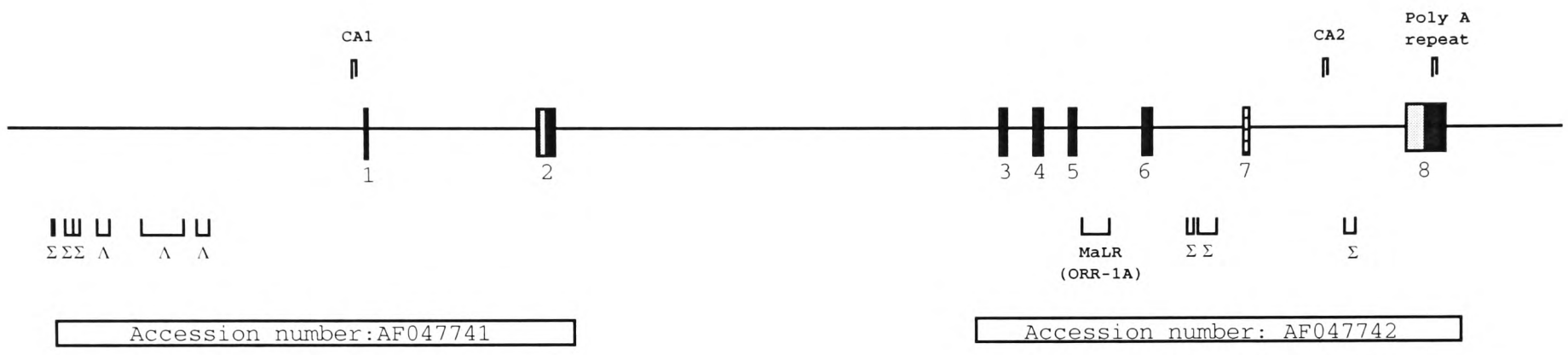


Figure 3.11. The location of repeat sequences within the murine gene for SP-D. The boxes represent exons, with the number of the exon listed below. Lines represent introns. The location of repetitive sequences are indicated by the square brackets, Dinucleotide repeats are shown above the gene structure; CA and polyA repeats. Repetitive elements are shown below the line; transposon (MALR) , LINE sequences (Λ) and SINE sequences (Σ). Open horizontal bars indicate the areas sequenced together with the accession number

compare the promoter regions between members of the collectins. Dotplot comparison between the human and murine SP-D promoters revealed a region of conservation stretching 700 bp upstream from exon 1 (Figure 3.12A). A similar area of similarity was seen when comparing murine SP-D with the promoter of bovine conglutinin, although fewer clusters of identity were found immediately adjacent to the transcription start site (Figure 3.12B). This suggests that the minimal promoter region for murine SP-D is approximately 700 bp long. It may also suggest that the more proximal region of the promoter is responsible for specific functions regarding SP-D gene function, such as cell-type specificity, whereas the more distal regions may confer functions that pertain to general functions of the collectins, such as acute phase reactants.

3.11.3. Transcription Factor Binding Sites

To establish an idea of the transcription control mechanisms influencing the expression of SP-D, the promoter region of murine SP-D was scanned for transcription factor binding sites using the transcription database (*http://pdap1.trc.rwcp.or.jp/research/db/TFSEARCH.html*) and a list of well characterised transcription sites (Faisst and Meyer, 1992). An array of transcription factor binding sites were identified within the 5' region flanking the mouse SP-D gene (Figure 3.13). Those which were positionally conserved between the murine and human SP-D genes or conglutinin were judged as having a higher probability of being functionally significant (listed in Table 3.4).

Two blocks of high sequence identity in the alignment between the mouse and human SP-D genes (Figure 3.13) are of particular interest, since they both contain conserved cell-specific transcription factors. Within one of these conserved blocks lies a degenerate TATA box (CATAAAT) at -30. The sequence of this degenerate TATA box is conserved in humans and mice, differing by one nucleotide from the consensus TATA box (TATAWAW). Also found within this conserved block of sequence are

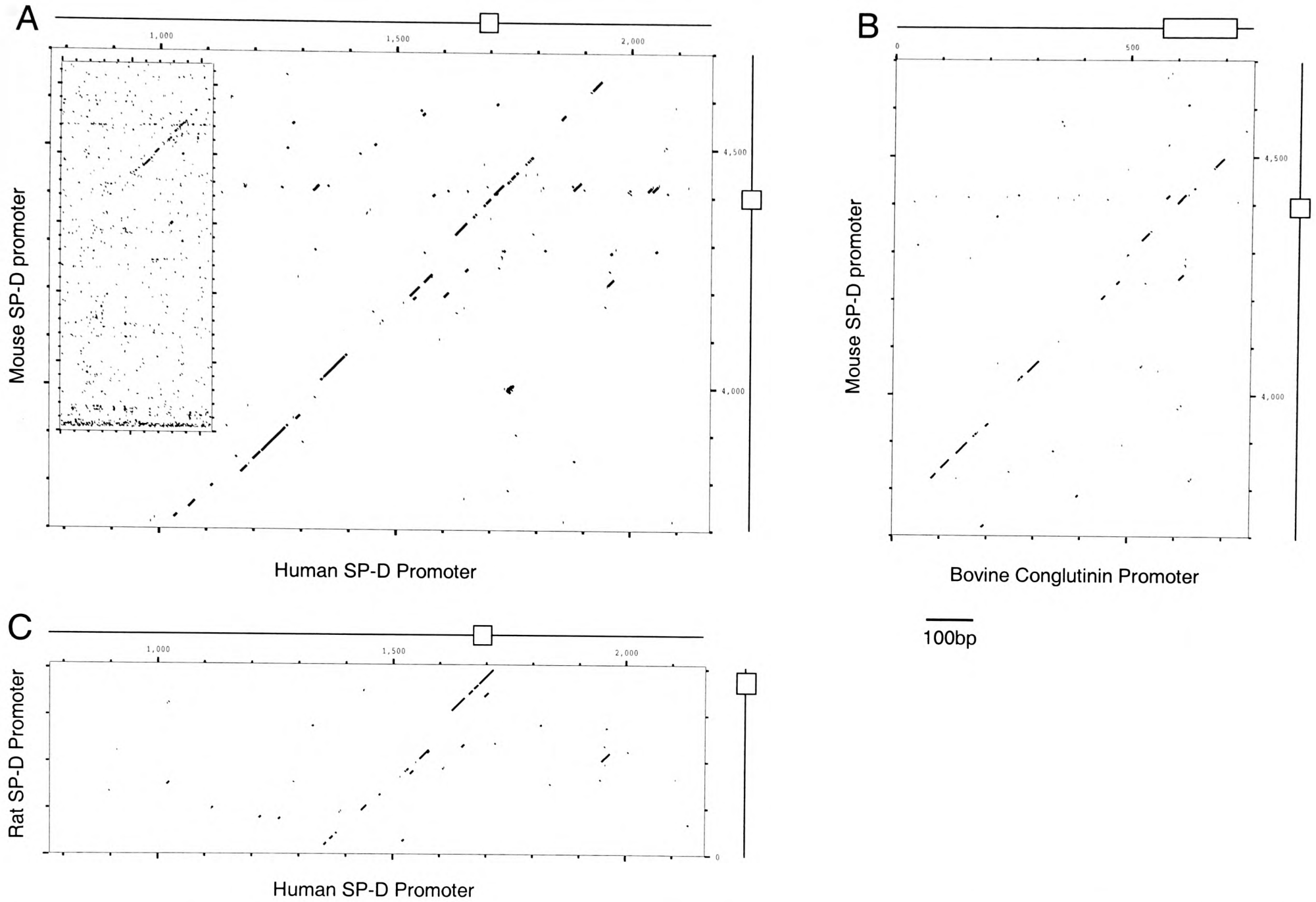


Figure 3.12. Dotplot sequence analysis of the promoter regions of SP-D and bovine conglutinin. **(A)** Mouse SP-D with Human SP-D (Rust *et al*, 1996). **(B)** Mouse SP-D with bovine conglutinin (Kawasaki *et al*, 1993).**(C)** Rat SP-D (Rust *et al*, 1996) and Human SP-D. A pictorial representation of a section of the genes with the promoter region (line), first exon 1 (box), and part of intron 1 (line) is shown surrounding the dotplots. The scale is shown at the bottom right. The insert in A shows the extent of similarity between the murine and human SP-D promoters, each tick represents 200 bp. Numbering starts at the 5' end of the sequence.



Figure 3.14. Alignment of the mouse and human SP-D promoter regions. The human SP-D (top, AF034862) and mouse (bottom, AF04774) promoter regions, including the first exon and part of the first intron, are shown in alignment. The transcription initiation point is indicated (+1). The alignment was performed using Bestfit from the Wisconsin package (see section 2.9); gaps have been introduced to maximise the similarity. Transcription factor sites are shown above and below the alignments of human and mouse, respectively; their positions are indicated by a single bar. The sequence of the first exon is indicated in green characters. Bold characters represent sequence identity between species.

two cell-specific transcription factor binding sites. These are the transcription factor binding site, Cdx (MTTTATR), which is responsible for intestinal-specific gene transcription (Suh *et al.*, 1994) plus the core recognition sequence (RTAAAYA) for the forkhead family (fkh) of transcription factors (Overdier *et al.*, 1994) are involved in organ specific gene expression along the foregut axis (Bohinski *et al.*, 1994).

Transcription factor site	Consensus	Location	Conserved	Reference
Fkh	RTAAAYA	-27	Hu, CNG	(Overdier *et al.*, 1994)
TATA box	TATAWAW	-30	Hu, CNG	(Breathnach and Chambon, 1981)
Cdx	MTTTATR	(-30)	Hu, CNG	(Suh *et al.*, 1994)
H-APF-1	CTGGYAA	-66	Hu	(Faisst and Meyer, 1992)
AP-1	TGASTMA	-96	Hu, CNG	(Faisst and Meyer, 1992)
LF-A1	TGRMCC	-170	CNG	(Hardon *et al.*, 1988)
E-Box	CANNTG	-221	Hu	(Faisst and Meyer, 1992)
		-284	CNG	
		-822	Hu	
TTF-1	CTAAG	-308	Hu	(Yan *et al.*, 1995)
		-530	Hu	
NF1-like	TGGCA	(-350)	CNG	(Faisst and Meyer, 1992)
PEA3	AGGAAR	(-362)	Hu, CNG	(Faisst and Meyer, 1992)
AP-2	CCCMNSSS	-484	Hu	(Faisst and Meyer, 1992)
demi-GRE	AGAACA	(-505)	Hu	(Faisst and Meyer, 1992)

Table 3.4. Transcription factor binding sites present in *Sftpd*. The position of each site is indicated relative to the transcription start site in the mouse gene. Parenthesis indicate inverted elements. Conservation in the human gene (Hu) and bovine conglutinin gene (CNG) is indicated.

Upstream within a second larger block of conserved sequence (-512 to -459) a conserved binding site for the transcription factor AP-2 (CCCMNSSS) at -509 and a conserved central core motif for TTF-1 (Yan *et al.*, 1995) at –489, were identified. Positional conservation of the SP-1, H-AP-1 and NF-IL6 regulatory elements could

not be seen in the *Sftpd* as described for SFTPD and the rat SP-D gene (Rust *et al.*, 1996).

It is interesting to note that the rodent and human genes for SP-D do not contain CAAT boxes within the vicinity of the TATA box, unlike the conglutinin gene. Indeed alignment between *Sftpd* and CGN1, showed that the analogous region of *Sftpd* has been disrupted by the polymorphic microsatellite (CA1).

3.12. Generation of Specific Antibodies for mouse SP-A and SP-D

3.12.1. Cloning of the cDNA fragment for the neck-CRD of murine SP-A

A fragment of the murine cDNA for SP-A was cloned by PCR. Briefly, a single strand cDNA was generated using moloney murine leukaemia virus reverse transcriptase (section 2.8.5) on total BALB/c lung RNA primed with the anti-sense oligonucleotide specific for the mouse SP-A cDNA, mAr. The single stranded DNA enriched for the cDNA for SP-A, was used as a template for amplification using primers mAr and mAf. A standard PCR assay (2.8.1) was optimised to 1.5 mM magnesium chloride and a 60°C annealing temperature to obtain a single discrete PCR product of approximately 480 bp. The primers were designed to introduce an *Eco*R I restriction sites at the 5' and 3' ends for cloning into pGEX-2T (Pharmacia). This construct was sequenced using vector and internal SP-A specific primers for confirmation.

3.12.2. Cloning of the cDNA fragment for the neck-CRD of murine SP-D

The neck CRD fragment of mouse SP-D was cloned by PCR, as described for murine SP-A (3.12.1). The antisense primer, mDn1-r, was used it initiate single-stranded cDNA synthesis. This product was used as a template in a standard PCR assay, using primers, mDn1-f and mDn1-r, with an annealing temperature of 58°C and 1.5 mM $MgCl_2$. The discrete product of 501 bp was then cloned into pGEX-2T and sequenced for confirmation.

3.12.3. Confirming the Sequence of Murine SP-A

It was observed that two bases within the cDNA coding region for the neck region of SP-A in the sequence obtained by PCR, did not correspond to the cDNA derived from the published gene sequence (accession number S48768). Alignment of the cDNAs covering the neck region of SP-A in other species plus the sequences of four different expressed sequence tags (EST) from the NCBI EST database (see Figure 3.14), suggested that the published gene sequence contained either a sequencing error or a polymorphism. The corrected sequence has leucine (CTC codon) at position 111, which replaces a serine (TCC codon). The CTC codon is present in all four murine ESTs, and is conserved across all species examined at the cDNA level to date, except in the guinea pig SP-A that contains a TTC codon for phenylalanine.

3.12.4. Murine SP-A and SP-D Used to Generate Antibodies

In order to raise antibodies, soluble protein[4] was obtained using the *Pichia pastoris* yeast expression system (Invitrogen, Groningen, The Netherlands), performed by N. K. Lim (MRC Immunochemistry Unit, Oxford). The cloned cDNAs encoding the neck-CRD fragments of murine SP-A and SP-D were used as a template for PCR and suitable restriction were introduced for cloning into the *Pichia pastoris* expression vector pPIC9K (Invitrogen). The constructs were designed to express the neck-CRD of mouse SP-A from Gly95 to the carboxyl terminal phenylalanine (F248) and the neck-CRD of mouse SP-D from Gly215 to the carboxyl terminal phenylalanine (F374). The construct was sequenced for confirmation and used to transform yeast. Once the DNA was integrated into the yeast genome, individual colonies were screened for high level protein expression. The protein was purified by N. K. Lim, using maltose affinity and gel filtration chromatography. This material was used as an antigen for generating polyclonal antibodies.

---

[4]The bacterial expression of the neck-CRD fragments of murine SP-A and SP-D were found to be more than 95% insoluble, despite varying the growth conditions, concentration of inducer and *E. coli* strain (results not shown).

A.

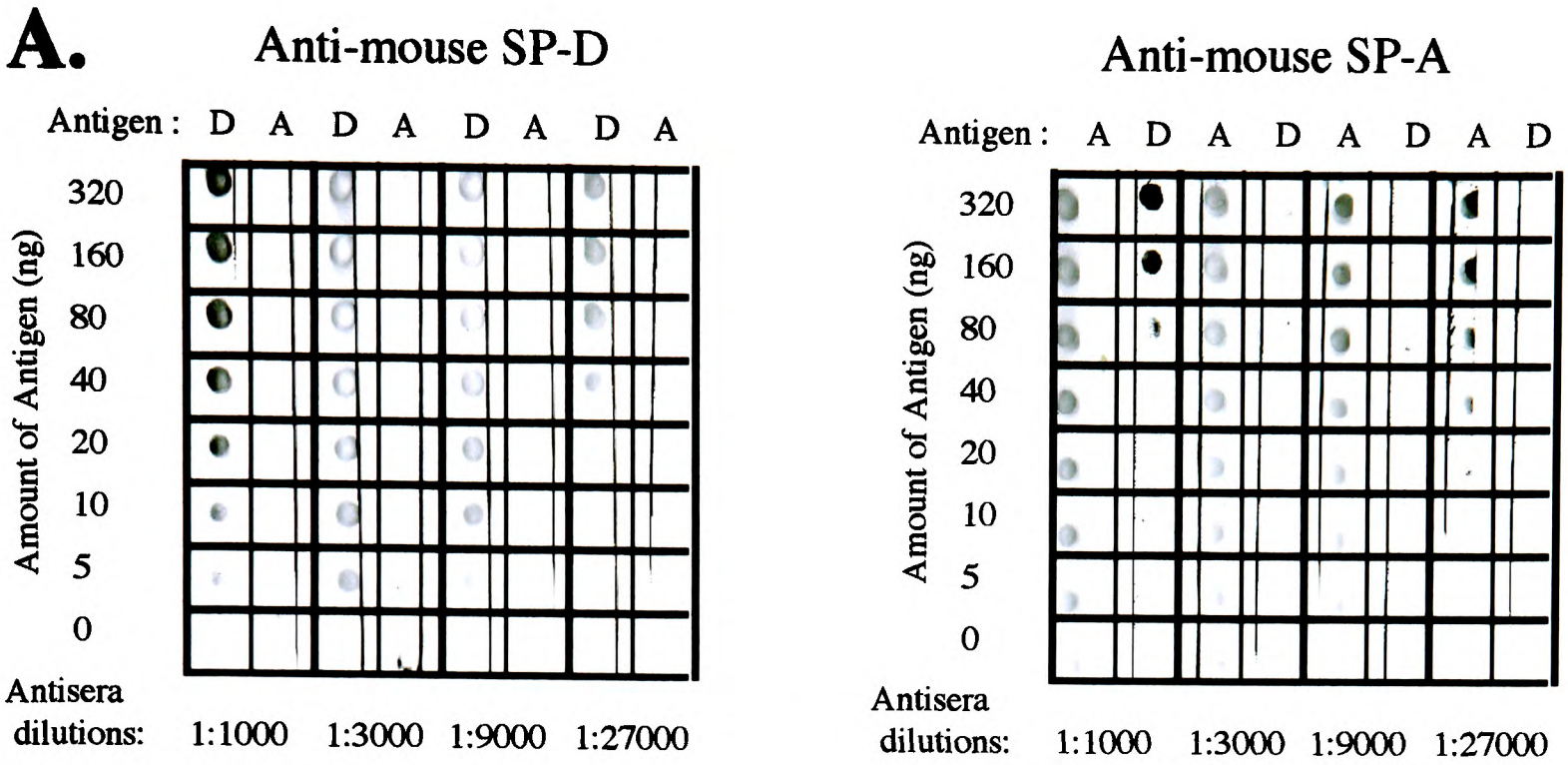
```
   293                                                
```

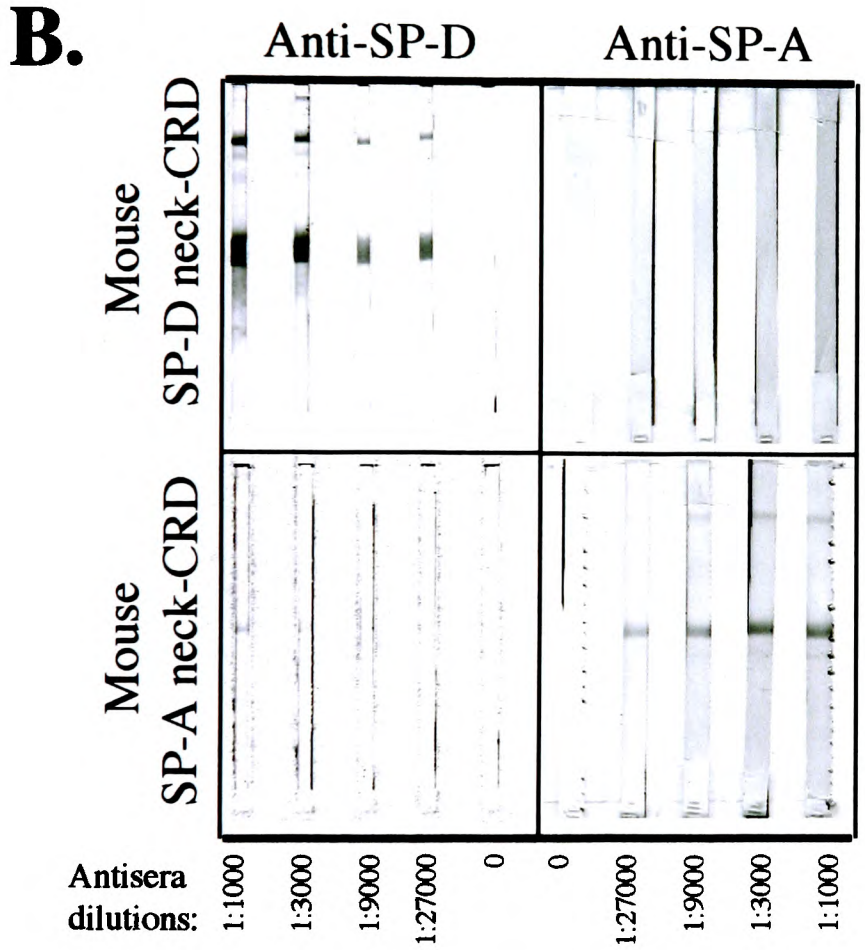
3.12.5. Raising Polyclonal Antibodies to Murine SP-A and SP-D

Polyclonal antibodies against the recombinantly expressed fragments of murine surfactant protein A and D were raised in rabbits (section 2.10.3.1). Indirect ELISA was used to monitor the increasing titre of the polyclonal antibodies during the immunisation protocol (section 2.10.3.3).

Both antibodies were considered specific for either SP-A or SP-D as judged by the lack of cross-reactivity of the antibodies with the respective non-cognate antigens (see Figure 3.15A). This removed any concern when using antibodies raised against native SP-A or SP-D, as low-level contamination with the unwanted surfactant collectin in the preparations to be used as immunogens, is difficult to avoid.

The optimal dilution of a primary antibody required to visualise approximately 5-50 ng of surfactant protein A or D was determined by dotblot (section 2.10.2.3) before performing Western blotting. This amount of protein was estimated as the quantity of SP-D or SP-A that was present in 10 µl (a typical volume of sample for SDS-PAGE) from a 1 ml total mouse lung lavage. The result of the dotblot using different dilutions of antibody versus protein amount (section 2.10.2.3) can be seen in figure 3.15A. The working concentrations was a compromise between discernible signal and the background from an antibody specific an irrelevant antigen. The antisera to mouse SP-A and mouse SP-D was deemed usable at a 1 in 9,000 dilution. The specificity of the antisera to the recombinant proteins of SP-D and SP-A at different dilutions, was also judged by Western blotting (Figure 3.15B). Approximately 5 µl and 10 µl of a 1 ml total lung lavage from normal mice was subjected to SDS-PAGE, followed by protein blotting. The membrane was incubated with the optimised dilutions of antisera and the protein blot was developed for analysis with a colourimetric method (section 2.10.3.4). From the protein blot, both antibodies could be judged as specific, detecting only in each case, either a 30-34 kDa





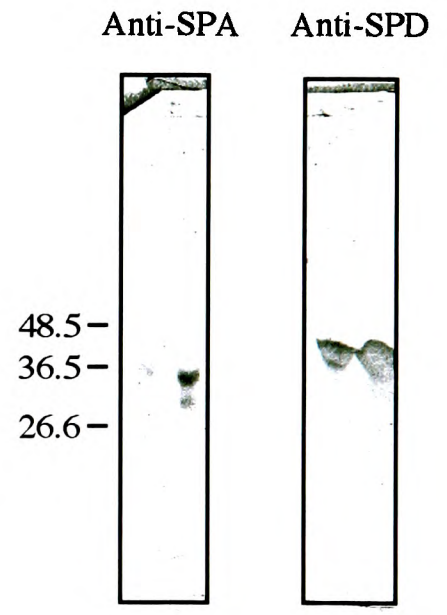


Figure 3.15. Characterisation of the antibodies directed against the neck-CRD fragments of murine SP-A and SP-D. (**A**) Dotblots: different antisera dilutions are titrated against different concentrations of ligand which is either the neck-CRD freagment of murine SP-A (A) or murine SP-D (D). (**B**) Dilutions of antisera against a fixed concentration of murine neck-CRD, SP-A or SP-D (20ng). (**C**) Western blot of 5 and 10 μl of mouse surfactant from a 1 ml total lung lavage, probed with a 1:9000 dilution of the antiseras. Molecular weight markers are shown to the left (kDa).

or a 42 kDa band, which corresponds to the predicted molecular weights for SP-A and SP-D, respectively (Figure 3.15C).

3.13. Discussion

This chapter describes the molecular cloning and the characterisation of the mouse gene for SP-D. The 5' extremity of the mouse SP-D transcript was cloned, demarking the transcription start site for the murine SP-D gene. The mouse SP-D gene was partially sequenced and this information has been submitted to GenBank as two individual contiguous stretches of sequence (Figure 3.3; Appendices 2, 3). Three new polymorphic simple sequence length polymorphic markers and three restriction length polymorphisms in the murine genes for SP-A, SP-D and MBL-A were described and used to genetically map the three murine collectins; SP-A, SP-D, MBL-A to a single recombination breakpoint.

3.13.1. The Mouse SP-D Gene Sequence

The murine gene was found to extend over 14 kb and consist of 8 exons (Figure 3.3). All intron-exon boundaries conformed to the GT-AG rule, as introns begin with a 5' GT and terminate with a 3' AG (Table 3.5). The positions of introns were conserved between the human (Crouch *et al.*, 1993b) and mouse SP-D genes (Figure 3.6), with all the protein coding exons being in phase I.

The mouse SP-D gene was sequenced by a combination of primer walking and shotgun cloning. The sequenced regions covered two areas of the eight exon gene. The first contig includes 3.5 kb of the 5' flanking region of the gene, exon 1 and exon 2 (accession number AF047741). The second contig encompasses exon 3 to exon 8 (accession number AF047742).

Intron	Intron Size	Intron Phase	3' Junction	Exon	Exon Size	5' Junction
				1	39	..AASCAATCT**gt**gag
						- - -
1	1991	0	gc**ag**GACATGCTG..	2	199	..GGTGATCCAG**gt**aag
			- M L			G D P
2	~5500	1	tat**ag**GTTTGCCA..	3	117	..GGACTAAGTG**gt**gag
			G L P			G L S
3	240	1	gcc**ag**GACCTCCA..	4	117	..GGGCCCAAAG**gt**ggg
			G P P			G P K
4	326	1	ctc**ag**GAGAAGTA..	5	117	..GGAGCAGCAG**gt**gag
			G E V			G A A
5	748	1	ttc**ag**GACCTGCC..	6	117	..GGGCTTCCAG**gt**aag
			G P A			G L P
6	1070	1	tat**ag**ACAGTGCT..	7	84	..TATCAGAAAG**gt**gag
			D S A			Y Q K
7	1764	1	ttc**ag**CTGCATTG..	8	500	
			A A L			

Table 3.5. The intron-exon boundaries of the mouse SP-D gene. Lower case and upper case characters represents intronic and exonic sequences, respectively. Part of the protein sequence is translated underneath the exons, where the first methionine is underlined. Bold characters indicate the acceptor and donor sites, which conform to the GT-AG rule. Sizes of the introns and exons where determined by sequencing, except intron 2 which was estimated from restriction digestion fragmentation and by PCR.

3.13.1.1. The Exons of the Mouse SP-D Gene

Since the 5' untranslated region of the mouse SP-D transcript was not present in the published cDNA sequence (Motwani *et al.*, 1995), PCR cloning (5′ RACE) was employed to clone the remaining 5' untranslated region. This 5' UTR was identified as being 926 bp in length and 39 bp corresponded to a sequence in the promoter region, termed exon 1. An additional level of confidence that the 5' UTR was correctly identified was that exon 1 showed 78% identity to the first exon of human SP-D and 80% identity to the first exon of rat SP-D (Rust *et al.*, 1996). The remaining exons (exons 2 to 8) of *Sftpd* showed complete identity to the published

mouse cDNA sequence (Motwani *et al.*, 1995), with the exception of the polyadenosine tract present within the 3' UTR. This was subsequently shown to be polymorphic (see below). Exon 2 codes for the last 3 bp of 5' UTR, the signal peptide, the N-terminal region, important for the quaternary protein structure (Brown Augsburger *et al.*, 1996a) and 40 amino acids of the collagen region. The first translated exon (exon 2) is followed by four 117 bp exons encoding an uninterrupted collagen-like sequence. The neck region, believed to initiate trimerisation of the polypeptide chain (Hoppe *et al.*, 1994) and the CRD domain are encoded by separate exons (exons 7 and 8). Exon 8 also codes for a 126 bp 3' untranslated region.

3.13.1.2. The Collagen Encoding Exons

At the boundaries of collagen encoding exons, glycine codons were found to be split by phase I by introns (Table 3.5), consistent with other members of collectin gene family (Figure 3.6) and resembling the organisation of the non-fibrous vertebrate and invertebrate collagens (Exposito and Garrone, 1990). Split glycine codons have also been identified in the type IV collagens (Oohashi *et al.*, 1995), the complement C1q genes (Sellar *et al.*, 1991), the macrophage scavenger receptor gene (Emi *et al.*, 1993) and the human ficolin genes (Endo *et al.*, 1996b). Phase I interrupted glycine codons are believed to be the most ancestral collagen exons (Exposito *et al.*, 1991). Phase I introns are the most prevalent type of introns hence they are compatible with exon shuffling and the formation of mosaic proteins, composed of different modules (Patthy, 1987).

The gene structure of SP-D is clearly suggestive of collagen exon duplication, as the collagen region is encoded by four identically sized 117 bp exons. Despite the presence of these four genomic interruptions, the Gly-Xxx-Yyy collagen-like motif is not perturbed in contrast to that seen for other collagen-containing proteins, such as C1q and MBL (Sellar *et al.*, 1991; Drickamer and McCreary, 1987).

3.13.2. Repeats

3.13.2.1. Microsatellites

Three polymorphic simple sequence length polymorphisms were described from the study of murine SP-D gene. Two dinucleotide repeats were shown to be present in the promoter region and in intron 7; in addition a polyadenosine tract was found within the 3' UTR. These represent the only simple sequence length polymorphism for the collectin gene locus. The polymorphic nature of these markers permits their use with many different backcross and recombinant inbred panels (Figure 3.8), making them useful for positional cloning and disease linkage studies.

3.13.2.1.1. Significance of the Repeats

The dinucleotide repeat within the promoter region of SP-D is present in rodents but not the human promoter. This repeat may prove to be interesting at the protein level, as dinucleotide repeats within the promoter region may be responsible for intraspecies phenotypic variation (Kashi *et al.*, 1997). It would be interesting to determine if variations in repeat length within the promoter or the 3' UTR give quantitative variations in the expression levels of SP-D. The significance to transcription of having a polyadenosine tract within the 3' UTR, close to the polyadenylated tail of a mRNA, resulting in two nearby polyadenylated stretches is not known.

3.13.2.2. Minisatellites and Repetitive Elements

The suggestion by Motwani and colleagues that the sequence of the 3' UTR of murine SP-D shared sequence similarity with human ALU repetitive elements and a human hypervariable minisatellite (D1S8) (Motwani *et al.*, 1995) could not be confirmed by screening repeat databases. It is possible that the polyadenosine stretch in the 3' UTR has been confused with the polyadensosine tracts that ALU repetitive elements possess. Two interesting features regarding dispersed repetitive elements were revealed.

First, the promoter region of *Sftpd* can essentially be defined by the absence of repetitive elements proximal to the start of transcription. This is in contrast to the distal region flanking the gene which contains three LINE-like and three SINE-like sequences (see Figure 3.11), confirming the promoter region defined by homology studies between the human and murine SP-D genes (section 3.11.2). Second, a dispersed repetitive element, a solitary LTR (Figure 3.7), is present in intron 5 of the mouse SP-D gene and belongs to the rodent-specific ORR-1A subfamily of mammalian apparent LTR retrotransposons (MaLRs) (Smit, 1993). This element is atypical as it contains a minisatellite insertion (Figure 3.7). Minisatellites contained within dispersed repetitive elements have been identified as hypervariable and are regarded as 'hotspots' of recombination (Kelly *et al.*, 1991). However, as suggested from the co-segregation of the murine collectin genes, this region of DNA does not appear as a recombination hotspot.

3.13.3. Genetic Mapping of the Collectins

The first CA polymorphic repeats (CA1) were used to genetically map the murine SP-D gene to chromosome 14 (Figure 3.9), in agreement with an earlier report (Motwani *et al.*, 1995). Further fine-scale mapping of *Sftpd*, shows that it lies between *D14Mit212* and *D14Mit141* (Figure 3.9). The three collectin genes were also genetically mapped using restriction length polymorphisms. All three genes co-segregated across the panel of approximately 1000 backcross mice. This suggests that cross breeding of single collectin null mice is not a feasible method for generating double or triple collectin gene null mice.

3.13.4. The Promoter Region

A valuable tool for the identification of genes and regulatory regions with gene promoters is the comparison of syntenic regions from different genomes (Koop and Hood, 1994). Dotplot analysis (Figure 3.12) and sequence alignment of the human and murine 5' flanking regions of SP-D genes (Figure 3.13) suggested a conservation

in regulatory sequence extending 700 bp upstream of the transcription start site. This analysis fits with the transcription analysis performed on the human SP-D promoter (Rust *et al.*, 1996).

3.13.4.1. Conservation Within the Promoter Region

Numerous stretches of sequence identity were revealed by alignment of the mouse and human SP-D promoter (Figure 3.13). Some clearly show importance as they coincide with consensus transcription factor binding sites (Table 3.4) (Faisst and Meyer, 1992). Other regions of identity warrant further investigation as they may include unknown transcription binding sites important for the regulation of SP-D expression. Areas with no identity may represent intraspecies differences in transcriptional control or simply random genetic drift

3.13.4.2. Conservation of Transcription Factor Binding Sites

Rust and colleagues (Rust *et al.*, 1996) have defined a minimal promoter for human SP-D that stretches 161 bp upstream from the transcription start site and confers cell-type specificity. Within this region, several transcription factor binding sites can be seen that are positionally conserved between the murine and human SP-D promoters (Figure 3.13). These include AP-1, H-APF-1, Cdx and the core motif for the forkhead family of transcription factors (Table 3.4). Within the human SP-D promoter, the AP-1 site has been shown to be functional (Rust *et al.*, 1996); this transcription factor binding site appears to be positionally conserved in the bovine conglutinin promoter, but it is not known if it is functional.

3.13.4.3. Lung-Specific Transcription Factor Binding Sites

The only conserved lung-specific transcription factor binding site within the minimal promoter is for the forkhead family of transcription factors (Overdier *et al.*, 1994). This is intriguing since members of this family also contribute to lung-specific expression of CC-10 and SP-B (Hellqvist *et al.*, 1996).

The other well characterised transcription factor which plays important roles in lung-specific expression of SP-A, SP-B, SP-C and CC-10 (Bruno *et al.*, 1995; Bohinski *et al.*, 1994; Kelly *et al.*, 1996; Toonen *et al.*, 1996) is thyroid transcription factor (TTF-1). The transcription factor binding site core motifs of TTF-1 (Yan *et al.*, 1995) were used to find two positionally conserved sites in the human and murine SP-D promoters, at -497 and -620 bp upstream from the transcription start site. Both of these sites were situated within larger blocks of sequence identity and warrant further investigation.

E-box motifs are known to be present and may be responsible for the action of the functional enhancer regions of the SP-A promoter (Gao *et al.*, 1993); two E-boxes are positionally conserved in the human and murine SP-D promoters.

A transcription factor binding site commonly found in acute phase response genes and those induced by lipopolysaccharides is H-APF-1. This site is conserved within the minimal promoter region for murine SP-D. This is interesting in light of the fact that lipopolysaccharides have been shown to up regulate rat SP-D *in vivo* (McIntosh *et al.*, 1996b).

3.13.5. Collectins in Other Species

The phylogeny of collectins has concentrated on the serum collectin MBL. MBL is the 'first component' of complement activation in early animals, as it is considered that C1 arose later to establish link with an antibody-mediated immune system (Dodds and Petry, 1993). Both an MBL-like molecule (Raftos *et al.*, 1997) and MBL associated serine protease (MASP)-like molecule have been identified in Urochordata (Ji *et al.*, 1997). Of other collectins, the cDNA and protein sequences of SP-D and SP-A have only been described for mammals (humans, rodents, cows and guinea pig). A SP-A-like RNA and protein has been described in fish (Sullivan *et al.*, 1998).

3.13.5.1. SP-D-like Genes in Other Species

Here, cross hybridisation of a murine SP-D CRD probe has been used to identify SP-D-like genes in other species. All the animals with lungs tested here; mammals, avians and amphibians were shown to have a single SP-D-like gene (Figure 3.4). Even though SP-D expression is not restricted to the lungs and has been detected at various intestinal sites in other species, a gene for SP-D could not be detected in fish. This raises the possibility that SP-D may not be present in animals without lungs, having arisen for lung related functions. It can be further hypothesised that in animals in which SP-D is present at non-pulmonary sites, the transcriptional control of the initially lung-specific gene must have been later adapted for expression at these secondary sites. This hypothesis could help explain the low expression levels of surfactant proteins A and D at their secondary sites of expression.

SP-D was found to be a single copy gene across all mammals, birds and amphibians tested. This stands in contrast to the cases of MBL and SP-A, where the copy number of identified genes varies between different species. For example, mice contain two functional MBL genes (Sastry *et al.*, 1995) and one SP-A gene (Korfhagen *et al.*, 1992) while humans have only one MBL gene and two functional SP-A genes. The diversity in the number and array of collectin genes suggests different selective pressures acting upon this group of innate immune pattern recognition receptors, within different animals, during the course of evolutionary divergence.

3.13.5.2. Sheep CL-43

Low stringency PCR was used to identify SP-D genes in other species. Instead, the sheep homologue for bovine CL-43 was identified. Bovine Conglutinin and CL-43 are believed to have arisen in bovidea from gene duplication of the SP-D gene. The putative sheep CL-43 is the first species homologue of CL-43 and displays 91% identity over 332 nucleotides. Indicating that if CL-43 duplicated from SP-D, this

event occurred in the common ancestor of cows and sheep. The amplified product is only a fragment of CL-43, the CRD. It remains to be determined the extent of the sheep CL-43 gene and whether it is transcribed or constitutes a pseudo-gene.

CHAPTER 4: CHARACTERISATION OF THE cDNA FOR PORCINE LUNG SURFACTANT PROTEIN D[1]

4.1. Introduction and Aims

The primary interest in investigating the protein components of porcine surfactant lay in its medical uses. Surfactant from both porcine and bovine origins is used as a successful replacement therapy for many different pulmonary and respiratory disorders, most notably respiratory distress syndrome (Spragg, 1998). At present the extracted surfactant used in therapy is principally composed of lipid and two hydrophobic proteins of the surfactant; surfactant proteins B (SP-B) and surfactant protein C (SP-C) (section 1.3). However, there is mounting evidence that the other protein components of surfactant may play a beneficial role. First, there is an improved disease outcome with the use of natural, non-extracted, surfactant compared to extracted surfactant (Halliday, 1997). Second, surfactant protein A (SP-A) protects the surfactant functions of extracted porcine surfactant from serum-induced inactivation when co-administered (Sun *et al.*, 1997).

Of the porcine surfactant proteins, SP-A has been characterised to a limited degree at the protein level (Ruano *et al.*, 1996; Ruano *et al.*, 1998). Porcine SP-B and SP-C have been isolated and their roles examined in lung surfactant function (Nag *et al.*, 1997; Tashiro *et al.*, 1999; Dico *et al.*, 1997; Perez Gil *et al.*, 1993). The cDNA sequence of porcine SP-A has not been published but has been submitted to GenBank (accession number L41350).

---

[1]Cloning of lung surfactant protein D was carried out in collaboration with the visiting scientist Martin van Eijk (Laboratory of Veterinary Biochemistry and Graduate School of Animal Health, Utrecht University, Utrecht, The Netherlands) and Professor, Henk P Haagsman (Department of the Science of Food of Animal Origin, Faculty of Veterinary Medicine, Utrecht University, The Netherlands). Porcine chromosomal localisation was carried out in collaboration with Dr Thomas Skinner of the laboratory of Professor Alan Archibald (The Roslin Institute, Roslin, Midlothian, Scotland).

As a starting point for investigating the porcine lung surfactant proteins in innate immunity, the surfactant collectins have been characterised at the cDNA and gene level. Lung surfactant protein (SP-D) was cloned following earlier work by collaborators on this project who demonstrated the presence of a porcine homologue for SP-D with a spurious molecular weight when compared to SP-D from other species (Henk P Haagsman, personal communication). The porcine SP-D-like molecule was shown to have a molecular weight of approximately 52 kDa on reduced SDS-PAGE, approximately 9 kDa larger than expected. From comparisons with other collectins there are two known potential sources for this variation. Firstly, the size of the collagen region varies between bovine and other mammalian SP-D sequences (Lim *et al.*, 1993). Secondly, the number of potential N-linked glycosylation sites on both SP-A and MBL varies between different species.

The second reason for investigating the porcine surfactant collectins was the hope that the characterisation of homologous proteins in different species could lead to a better insight into the roles and functions of these proteins. To date, SP-D has been cloned from four different species (mouse, rat, bovine and human) with each species containing a single SP-D gene, while SP-A has been cloned from eight different mammals (mice, rats, dogs, rabbits, baboons, humans, guinea pigs and pigs). Both humans and baboons have two functional genes for SP-A (Katyal *et al.*, 1992; Gao *et al.*, 1996), while other animals appear to have only one functional gene.

While the strategy of cloning the collectins follows standard cDNA library approaches, the design of the probe was important to obtain the full sequence of porcine SP-D. Empirical observations from the laboratory demonstrated that the mRNA encoding the repetitive collagenous stretch often leads to premature chain termination by reverse transcriptase during cDNA library construction. To avoid isolating prematurely terminated cDNA clones, a SP-D cDNA probe was used that lacked the collagen region but contained the amino and carboxyl terminal modules. In addition, the designed

probe was also believed to reduce cross hybridisation with other collagen encoding cDNA clones.

The characterisation of porcine SP-D described here has been published in the Journal of Immunology (Van Eijk *et al.*, 2000). Just prior to publication, a short communication by others described a partial PCR-derived cDNA sequence for porcine SP-D and its localisation in the Eustachian tube (Paananen *et al.*, 1999).

4.2. Construction of the Probe for Screening the Porcine Lung cDNA Library

A recombinant probe was constructed from the full-length human SP-D cDNA (Lu *et al.*, 1992) to lack the majority of the collagen encoding region. The probe was generated by cleaving the full-length human SP-D clone with two blunt end restriction endonucleases; *Nru* I and *Sma* I, thereby releasing the collagen encoding region. Following re-ligation, the generated construct encoded the signal sequence and the amino terminal region joined directly to the neck region and carbohydrate recognition domain (CRD) of human SP-D. The 0.6 kb recombinant insert was then excised from the plasmid with *Eco*R I and *Xho* I restriction digestion, gel purified (section 2.2.5.4) and radiolabelled (section 2.3.1).

4.3. Identifying the Lambda Bacteriophage Clones

A commercially available porcine lung cDNA library, constructed in λgt11 (Clontech) was screened with the recombinant human SP-D cDNA probe. A representative sample of lung cDNA clones were screened (0.6 x 10^6 plaques). After three rounds of screening, the probe hybridised with a total of 25 clones under high stringency hybridisation conditions (section 2.2.7.2).

4.4. PCR Sequencing of the Positive Clones

Positive clones were characterised by directly sequencing PCR products derived from the bacteriophage clones (section 2.8.2) to eliminate cloning false positive cDNAs. Inserts from the positive clones were amplified by PCR using commercially available primers[2] (Clontech) to the left and right lambda arms of λgt11. A standard PCR assay (section 2.8.1) was performed with an annealing temperature of 55°C, 1.5 mM magnesium chloride and an elongation step of 1 minute using a single plaque from a stock plate as the template. The PCR products derived from the positive plaques were sequenced directly with the amplification primers. Of the 25 positive plaques identified by hybridisation, only two were identified by sequencing as containing sequences homologous to human SP-D. The two clones covering the entire coding region of porcine SP-D were subcloned into pBluescript, generating the plasmid clones pPIG17d and pPIG20j.

4.5. Characterisation of the Porcine SP-D cDNA Clones

The two plasmid clones isolated were sequenced by primer walking, using pBluescript vector primers and internal SP-D sequence specific primers[2] designed from the sequence that was derived from the PCR products of the lambda clones.

The two plasmid clones were each sequenced on both strands with at least three independent sequence runs confirming each nucleotide. Sequencing was performed with fluorescent-dye terminator cycle sequencing chemistry (section 2.7.2). Continuous stretches of sequence were assembled for each clone using the DNAstar suite of Macintosh computer programs (section 2.9.3). Clone pPIG17d was 1123 bp long, covering nucleotides 41-900 (see section 4.5.1) and clone pPIG20j was 606 bp long and covered nucleotides 799-1385 (numbering based upon the porcine SP-D cDNA sequence shown in Figure 4.1). Alignment of the two plasmid sequences

[2]The sequence of all the primers referred to in this chapter are shown in Appendix 5.

```
                                                  M  L  L  L  P  L  S  V  L  I  L  L
CGAGTTTGCCTGGAGATTCTGAGCTCTAGAGGACGCAACTGACATGCTTCTCCTCCCTCTCTCCGTGCTGATCCTGCTC 79

 T  Q  P  P  R  S  L  G  A  E  M  K  T  Y  S  Q  R  A  V  A  N  A  C  A  L  V
ACACAGCCCCCGAGGTCACTGGGAGCAGAAATGAAGACCTATTCCCAGAGAGCAGTGGCCAACGCCTGCGCCCTGGTC 157

 M  C  S  P  M  E  N  G  L  P  G  R  D  G  R  D  G  R  E  G  P  R  G  E  K  G
ATGTGTAGCCCCATGGAGAATGGCCTGCCTGGTCGTGATGGTCGGGATGGGAGAGAGGGCCCTCGGGGCGAGAAGGGG 235

 D  P  G  L  P  G  A  V  G  R  A  G  M  P  G  L  A  G  P  V  G  P  K  G  D  N
GATCCAGGTTTGCCAGGAGCTGTAGGGCGAGCGGGGATGCCTGGACTGGCTGGCCCAGTTGGGCCCAAAGGGGACAAC 313

 G  S  T  G  E  P  G  A  K  G  D  I  G  P  C  G  P  P  G  P  P  G  I  P  G  P
GGCTCTACTGGAGAACCCGGAGCAAAGGGAGACATTGGACCATGCGGGCCTCCAGGACCTCCAGGTATACCTGGTCCA 391
                                                  ↑
 A  G  K  E  G  P  S  G  Q  Q  G  N  I  G  P  P  G  T  P  G  P  K  G  E  T  G
GCCGGAAAAGAAGGTCCCTCAGGGCAGCAGGGGAACATAGGACCTCCAGGCACACCAGGCCCCAAAGGAGAGACTGGG 469

 P  K  G  E  V  G  A  L  G  M  Q  G  S  T  G  A  R  G  P  A  G  L  K  G  E  R
CCCAAAGGAGAAGTGGGTGCCCTGGGCATGCAGGGCTCTACAGGGGCAAGAGGCCCTGCAGGTCTTAAAGGAGAGAGA 547

 G  A  P  G  E  R  G  A  P  G  S  A  G  A  A  G  P  A  G  A  T  G  P  Q  G  P
GGTGCCCCCGGTGAGCGTGGAGCCCCTGGAAGTGCTGGGGCAGCAGGGCCTGCTGGAGCCACGGGCCCTCAGGGCCCT 625

 S  G  A  R  G  P  P  G  L  K  G  D  R  G  P  P  G  E  R  G  A  K  G  E  S  G
TCAGGTGCCAGGGGCCCCCCAGGACTGAAAGGGGACAGAGGTCCTCCTGGAGAAAGAGGAGCCAAGGGAGAGAGTGGA 703

 L  P  G  I  T  A  L  R  Q  Q  V  E  T  L  Q  G  Q  V  Q  R  L  Q  K  A  F  S
CTCCCAGGCATCACTGCTCTGAGGCAACAGGTGGAGACCTTACAGGGGCAGGTACAACGCCTCCAGAAGGCCTTCTCT 781

 Q  Y  K  K  V  E  L  F  P  N  G  R  G  V  G  E  K  I  F  K  T  G  G  F  E  K
CAGTATAAGAAAGTGGAGCTCTTCCCCAATGGCCGAGGTGTCGGGGAGAAGATCTTCAAGACGGGAGGCTTTGAAAAG 859

 T  F  Q  D  A  Q  Q  V  C  T  Q  A  G  G  Q  M  A  S  P  R  S  E  T  E  N  E
ACTTTTCAGGATGCTCAGCAGGTATGCACACAGGCCGGGGGACAGATGGCCTCCCCACGCTCTGAGACTGAGAACGAG 937

 A  L  S  Q  L  V  T  A  Q  N  K  A  A  F  L  S  M  T  D  I  K  T  E  G  N  F
GCCTTGAGCCAGCTGGTCACAGCTCAGAATAAGGCTGCTTTCCTGAGCATGACTGACATCAAGACGGAGGGCAATTTC 1015

 T  Y  P  T  G  E  P  L  V  Y  A  N  W  A  P  G  E  P  N  N  N  G  G  S  S  G
ACCTACCCCACGGGGGAGCCCCTGGTCTATGCCAACTGGGCCCCTGGGGAGCCCAACAACAATGGTGGCAGCAGCGGA 1093

 A  E  N  C  V  E  I  F  P  N  G  K  W  N  D  K  A  C  G  E  L  R  L  V  I  C
GCAGAGAACTGTGTGGAGATCTTTCCCAATGGCAAGTGGAATGACAAGGCCTGCGGAGAACTGCGCCTCGTGATCTGC 1171

 E  F  .
GAGTTCTGAGCCCCTGGGGAGGGAGGGGCGGTGTCCAGAGCTGTGTGCTACCAACGTCCCAATAAATAGGTGACCTTC 1249
TGCTGGCCAGGGCTTCTCCACAGAGCCGTGGGACGAGGCCAGAAGGTAGGGAGCCTATGGAACGCCTCCCTCAGAATA 1327
AAGTACGAAACTGGCCTCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA                      1385
```

Figure 4.1. The cDNA sequence of porcine SP-D. The numbering refers to the nucleotide sequence beginning at the start of transcription. The predicted protein sequence is shown above the nucleotide sequence, starting at the first methionine codon, ATG (bold). The protein sequence in italics indicates the predicted leader peptide. The underlined DNA sequence indicates that identified by 5' RACE. The remaining sequence was obtained from the two lambda clones isolated from a porcine lung cDNA library. The two potential polyadenylation signals are shaded. The arrow indicates an exon-intron-exon junction identified in the gene by PCR sequencing.

revealed an overlap of 102 bp (see Figure 4.2) that showed complete sequence identity between the two clones.

4.5.1. Analysis of the 3' Clone for Porcine SP-D

Sequence analysis of the plasmid clone, pPIG20j, showed 85% identity across its entire length with the 3' end of bovine SP-D. The clone encoded for part of neck region, the entire CRD and the 3' UTR. The clone appears to contain the complete 3' UTR due to the presence of a polyadenosine tract.

4.5.2. Analysis of the 5' Clone for Porcine SP-D

The plasmid clone pPIG17d, also showed just 85% identity to bovine SP-D, spanning 855 bp, although the clone itself was 1123 bp in length. This clone contained the cDNA sequence which encoded the cysteine rich amino terminal region, the collagenous region, the neck region and part of the CRD. However, the 5' region of pPIG17d was homologous to the cDNA sequence of human protein kinase C (PKC), showing 86% identity (see Figure 4.2B; speckled region of the solid bar). Appendix 6 shows the complete sequence of clone pPIG17d. This clone was considered to be a chimera of PKC and SP-D, probably a result of a cloning artefact generated from the construction of the cDNA library. Curiously, the point of transition appears to occur at a predicted intron-exon boundary in porcine SP-D. To identify the entire 5' sequence of porcine SP-D and to confirm that the clone was chimeric, 5' rapid amplification of cDNA ends (RACE) (section 2.8.3) was employed to clone the 5'-region of SP-D from porcine lung RNA.

4.6. 5' RACE to Clone the 5' Region of Porcine SP-D cDNA

Single-stranded lung cDNA transcripts enriched in SP-D-like cDNA sequences were generated from total mouse lung RNA by reverse transcriptase (section 2.5.2) with an anti-sense oligonucleotide that was specific for nucleotides 225-243 of porcine SP-D (PD-1R). After tagging the cDNAs with deoxycytidine (section 2.3.2), an anchor PCR

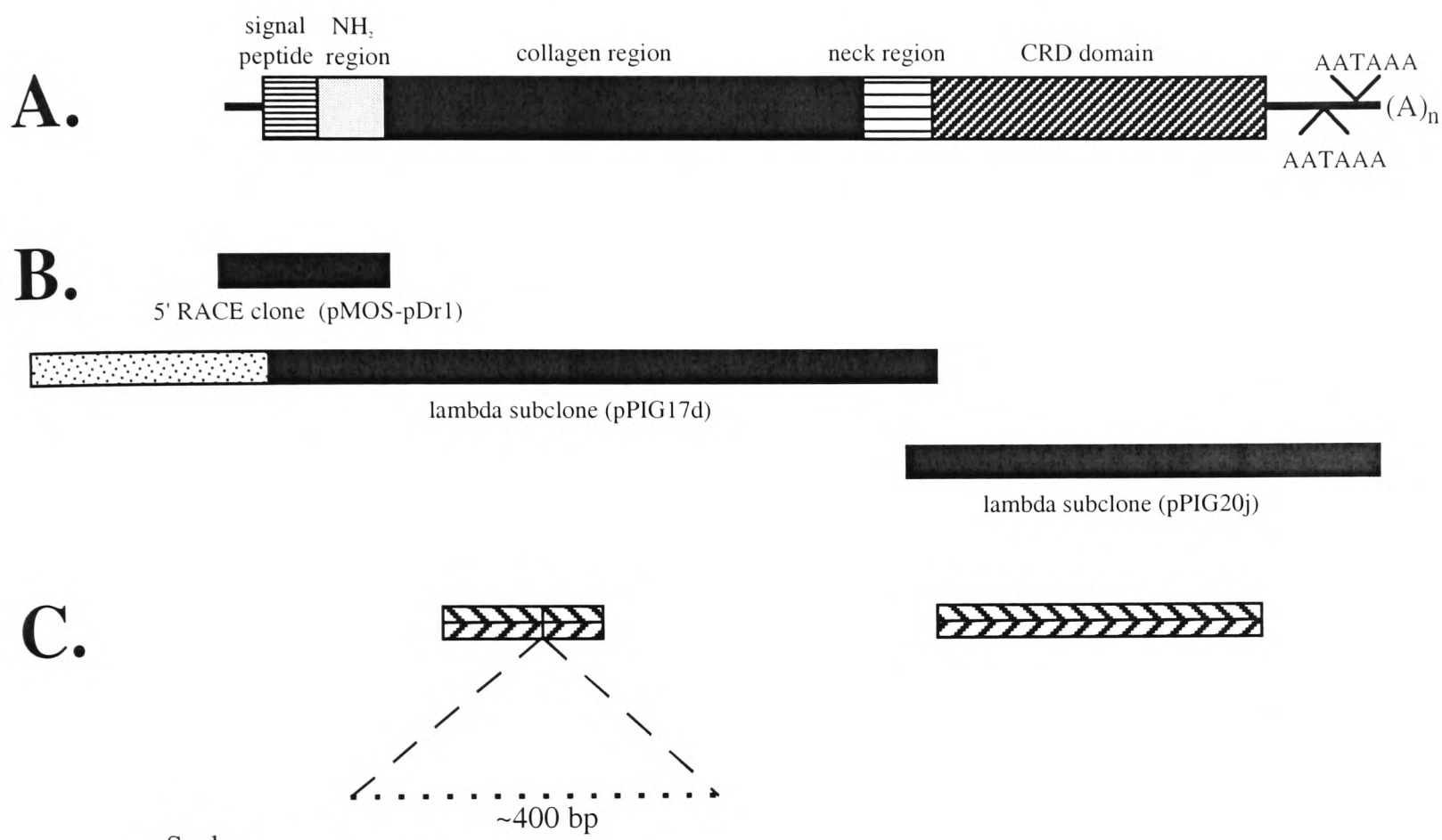


Figure 4.2. The arrangement of the clones used to obtain the full length porcine SP-D cDNA sequence. **(A)** A graphical representation of porcine SP-D, detailing the untranslated regions (sticks) and coding regions (boxes). Shading indicates different regions of the protein; signal peptide (dense horizontal bars), N-terminal cysteine rich region (dot shading), collagen-like region (solid bars), neck domain (light horizontal bars), carbohydrate recognition domain (dense diagonal bars). The polyadenosine tract ($(A)_n$) and the potential polyadenylation sites (AATAAA) are indicated. **(B)** The cDNA clones obtained by subcloning the two lambda clones and the 5' RACE clone (black bars). The speckled bar indicates the region of pPIG17d which is homologous with protein kinase C. **(C)** The genomic PCR products (hatched boxes) used in direct sequence confirmation of the three major variants in porcine SP-D. The genomic PCR covering part of the collagen encoding region contains an intron; this is indicated by an inserted dotted line.

(section 2.8.3) was performed with the anchor primer $XSCdG_{17}$ and PD-1R. A second PCR was carried out using a fraction of the first PCR as a template to introduce a higher level of stringency in the amplification, using the anchor specific oligonucleotide (XSC) and an anti-sense primer specific for nucleotides 137-167 of porcine SP-D (PD-5PE). The 5' RACE products formed a single band on agarose gel electrophoresis of the expected size, approximately 200 bp, as estimated from the sequence of human, rat and murine SP-D. The products were gel purified (section 2.2.5.4) and cloned into a TA overhang vector following the manufacturer's procedure (pMOSblue; Amersham Life Sciences).

A total of ten independent 5' RACE clones were sequenced using vector derived primers (Amersham Life Sciences). The sequences were aligned using Seqman (DNAstar) then edited and adjusted manually to generate a consensus sequence. The longest 5' extension products were used to demark the beginning of the transcript (+1). Seven of the RACE products extended to this position. The RACE product contained 96 bp of sequence identical to that derived from the 5' clone (pPIG17d) plus an additional 40 bp of novel sequence 5' of the point of chimerism. The RACE derived sequence completed the 5' sequence of porcine SP-D and also proved that the 5' lambda derived clone, pPIG17d, was a chimera of porcine PKC and SP-D. The sequences of the 5' RACE products revealed that the majority of the clones, seven out of ten, extended to position +1. However, a minor transcription start site can not be ruled out, since the remaining three clones extended to position +2. Indeed, the murine SP-D gene may also have an alternative minor transcription site (section 3.5). However, direct sequencing of the total, unpurified, 5' RACE product indicated that the majority of the cDNAs extended to the designated start of transcription as no appreciable drop in sequence strength was observed when sequencing toward the start of transcription. This implied that the start of transcription is consistent with the sequences of the 5' UTR of SP-D derived from human, rat and mouse.

4.7. Sequence Analysis of Porcine SP-D

The entire sequence of porcine SP-D constructed from the overlapping RACE and lambda derived clones has been submitted to GenBank with the accession number AF132496. Figure 4.1 shows the full length porcine SP-D sequence and Appendix 7 shows the GenBank database entry including the sequence annotations. All the clones overlapped by at least 96 bp and showed complete identity in the region of overlap. The longest open reading frame running from the first ATG methionine codon and terminating with a TGA opal stop codon encodes a 378 amino acid protein. The protein sequence displays the greatest similarity to bovine SP-D (78% identity) and is shown aligned with human, bovine, murine and rat SP-D in Figure 4.3.

4.7.1. Sequence Analysis of the Untranslated Region

The 3' UTR of porcine SP-D shows more identity to bovine SP-D than rodent and human SP-D. The 3' UTR of porcine SP-D is 170 bp long, similar in size to that of bovine SP-D, 151 bp. This is unlike the shorter 3' UTRs of human, rat and murine SP-D (101-121 bp). The long 3' UTRs of both porcine and bovine SP-D show greatest identity to another collectin, bovine conglutinin (81%), rather than human or rodent SP-D (less than 50%). The long 3' UTR of porcine SP-D contains two potential polyadenylation sites (AATAAA), 25 and 117 bp upstream of the polyadenosine tract that was sequenced. The dual polyadenylation feature is also shared by the long 3' UTRs of bovine SP-D and conglutinin, again quite distinct from the single polyadenylation site seen for rodent and human SP-D. In contrast, the sequence of the porcine 5' UTR displays the greatest similarity to both human and murine SP-D, at approximately 70% identity for both.

4.7.2. Analysis of the Predicted Protein Sequence of Porcine SP-D

The primary structure of porcine SP-D predicted from the cDNA sequence reveals a protein with the conserved modular organisation of the collectins; a cysteine-rich amino-terminal domain linked to a collagen-like region, a short neck region followed by a

```
          1                                                                                                  ||                  100
po SP-D   MLLLPLSVLILLTQPPRSLGAEMKTYSQRAVANACALVMCSPMENGLPGRDGRDGREGPRGEKGDPGLPGAVGRAGMPGLAGPVGPKGDNGSTGEPGAKG
hu SP-D   ...FL..A.V.....LGY.E.......H.TTPS..T.....SV.S.............................A.Q.....Q.............V....P..
bo SP-D   .........L.....W............KTM....T......P.D.....................S..PA.......P...I.L.....A....P..
mo SP-D   ..PF-..M.V..V..LGN......SL...S.P.T........T.............................PM.LS.LQ.PT.......E...A....P..
ra SP-D   ..HF-..M.V..V..LGD.......L...SIT.T.T..L...T.............................PM.LS.L..PR.......E...A....P..
                                                                                                   ||
con       ML...LS.L.LL.QP...L.AEMK..S.......C.LV.CS..E.GLPGRDGRDGREGPRGEKGDPG.PG..G..G..G..GP.G.KG.NGS.GEPG.KG
                                                                                                   ||

          101 ||                                        ||                                                ||          200
po SP-D   DIGPCGPPGPPGIPGPAGKEGPSGQQGNIGPPGTPGPKGETGPKGEVGALGMQGSTGARGPAGLKGERGAPGERGAPGSAGAAGPAGATGPQGPSGARGP
hu SP-D   .T..S.......V.....R...L.K......Q.K......A........P.....A....L..P.....V.....V..N.....S...M....SP.....
bo SP-D   .T---.......M.....R.....K..SM..........D.....G...P.I...P---..............P....R...P.....I...........
mo SP-D   ER.LS....L..............K......Q.K......A........P........K.ST.P........VQ....N........PA....AP.S...
ra SP-D   ER.LV....S...S..........K......Q.K......A........P.....A..K....P.........Q....N........PA....AP.S...
              ||                                        ||                                                ||
con       ..---GPPG.PG..GPAG.EGP.G.QG..GP.G.PGPKG..GPKG.VGA.G.QGS.---G..G.KGERG.PG..G.PG.AGA.G.AG..GPQG..G.RGP
              ||                                        ||                                                ||

          201                          ||                                ||                                          300
po SP-D   PGLKGDRGPPGERGAKGESGLPGITALRQQVETLQGQVQRLQKAFSQYKKVELFPNGRGVGEKIFKTGGFEKTFQDAQQVCTQAGGQMASPRSETENEAL
hu SP-D   ......K.I..DK.........DVAS......A......H..A..............QS........A..V.P.TE..LL.......L.....AA..A..
bo SP-D   ........T............AEVN....R.GI.E..L....N.......AM......S........V.S.........I.......LP....GA.....
mo SP-D   ........V..D..I.......DSA.....M.A.K.KL.. EV...H.Q.AA...D..S..D...R.ADS..P.E...EM.K.....L.....A...A.I
ra SP-D   ........A..D..I.......DSA.....M.A.N.KL...EA...R...AA...D.QS..D...RAANS.EP.E..KEM.R.....L.....A...A.V
                             ||                                ||
con       PGLKGD.G.PGE...KGESGL.....LRQ....L.G..Q.L..AFS.Y.K..LFP.G..VG.KIF........F..A...C.QAGGQ..SPRS..EN.A.
                             ||                                ||

          301                                                                                                  378
po SP-D   SQLVTAQNKAAFLSMTDIKTEGNFTYPTGEPLVYANWAPGEPNNNGGSSGAENCVEIFPNGKWNDKACGELRLVICEF
hu SP-D   Q...V.K.E........S....K.......S...S........DD...---.D.....T......R....K...V...
bo SP-D   T..A...........S.TRK..T.I.........S....Q....D...---.................V...Q.......
mo SP-D   Q..I..H..........VG...K...........S...........A---.......T..Q........Q.......
ra SP-D   Q.....HS.........VG...K.......A...S...........A---.......T..Q........Q.......

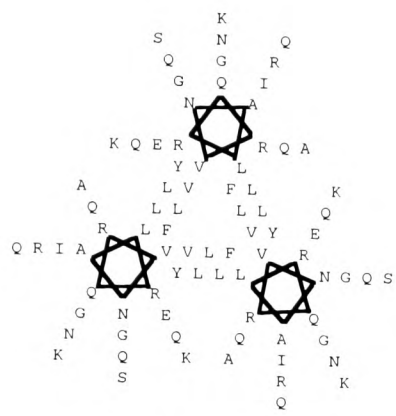
con       .QL..A...AAFLSM.D...EG.F.YPTGE.LVY.NWAP.EPN..GG.---E.CVEIF.NG.WND..CGE.RLV.CEF
                                                        * *         *            **
```

Figure 4.3. Alignment of the five mammalian protein sequences of SP-D. Porcine SP-D (po) aligned with four other mammalian protein sequences for SP-D; human (hu), bovine (bo), mouse (mo) and rat (ra). The three major differences in porcine SP-D are highlighted (bold); an extra cysteine residue, the presence of a potential glycosylation site in the CRD and a three amino acid insertion in the CRD. Dots indicate residues in common with the porcine sequence. Dashes indicate spaces that are inserted to maximise the identity across the alignment and underlining shows both conserved and non-conserved potential N-linked glycosylation sites. A consensus sequence of identity is constructed for the five mammalian SP-D sequences and shown below the alignment. The predicted key residues required for Ca^{2+} co-ordination and hydrogen bonding with hydroxyl groups on carbohydrate, are indicated with asterisk. The approximate position of exon-intron-exon boundaries are marked (| |), predicted from the human and mouse gene structures.

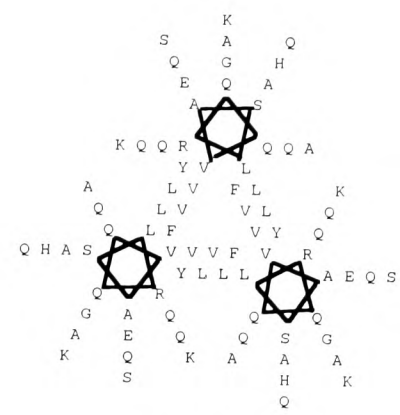
CRD domain. The 180 amino acid long collagen-like sequence is identical in length to that of human and rodent SP-D, unlike the collagen-like sequences of bovine SP-D and conglutinin which are two triplets shorter. The neck region conforms to the consensus sequence required for the formation of trimeric alpha-helical coiled coils, with hydrophobic residues (Leu, Phe, Met or Val) at positions +1 and +4 within a heptad repeated sequence. This is shown in the helical wheel diagrams of Figure 4.4.

4.8. The Major Sequence Variations Seen in Porcine SP-D

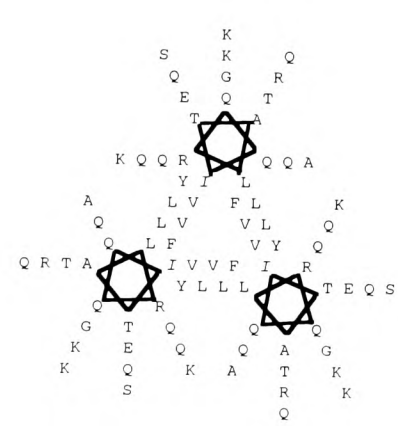
Comparison of the protein sequences of SP-D from different species by alignment reveals three substantial changes in porcine SP-D. The first is a cysteine residue in the collagen-like sequence of porcine SP-D at position 105, replacing a serine residue found in murine and human and a valine residue in the rat (there is a gap in the alignment at this position in the bovine sequence). In all cases this codon alteration requires two nucleotide changes. The codon in pigs is TGC, in humans and mice an AGT codon and in rats a GTT codon. The second feature that was revealed by alignment, is the insertion of three extra amino acids in the CRD. This kind of difference is not only unusual for SP-D but is also unreported for different species within any group of the collectins. Prior to this study, the length of the CRD from different collectins was known to vary but for each collectin from different species the length considered to be constant. Finally, porcine SP-D contains a potential N-linked glycosylation site within the CRD, unlike SP-D from other species. The potential N-linked glycosylation site results from an AAT codon (Asn) at position 323 that is changed from the AAG codons (Lys) of humans and rodents. In bovine SP-D the potential N-linked glycosylation site of porcine SP-D, Asn-Phe-Thr reads Thr-Phe-Ile. Again this requires two nucleotide changes from the bovine DNA sequence of ACT TTC ATC to the porcine sequence of A<u>A</u>T TTC A<u>C</u>C.



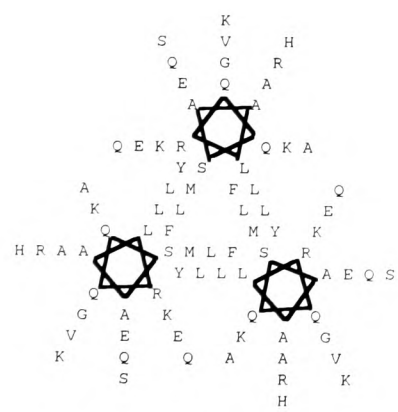
Bovine SP-D



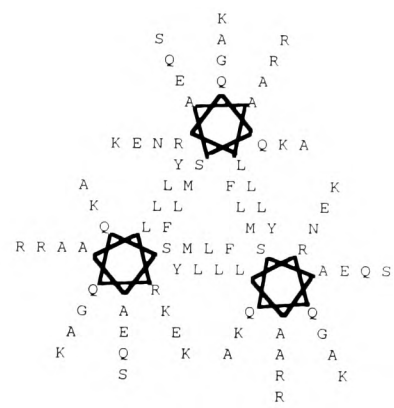
Human SP-D



Porcine SP-D



Mouse SP-D



Rat SP-D

Figure 4.4. Helical wheel diagrams. The heptad repeating sequence of the neck region in SP-D are shown as three helical wheels. Hydrophobic residues in all species to date can be seen at position 1 and 4 in the heptad repeat which are implicated in the trimerisation of three neck regions. Residues 224 to 250 are shown for the porcine SP-D neck region.

4.9. Confirmation of the Major Sequence Variations in Porcine SP-D

The three sequence variations in porcine SP-D were confirmed as being fixed within the porcine population (rather than mutations, polymorphisms or cloning artefacts) by sequencing SP-D from a number of different pig strains. It was deemed important to establish this point, as the variations within porcine SP-D may have important structural, functional and evolutionary consequences. In this study the most convenient route for sequence confirmation was PCR amplification of the fragments of the porcine SP-D gene from the different breeds, followed by direct sequencing of the PCR product.

4.9.1. Amplification of the Genomic Regions Containing Sequence Variation

Genomic DNA was purified using pig sperm from four different pig strains; Large White Sireline, Finnish Landrace, Duroc Sireline and Piëtrain. Pig semen was received as a generous gift from Dr H Feitsma (Varkens KI, Noord-Brabant, The Netherlands). To amplify the genomic regions of interest, primers were designed from the porcine SP-D cDNA sequence and positioned within exons, as predicted from the intron-exon boundaries of the human and murine SP-D genes (Figure 4.3). Two regions were amplified to assess the sequence variations. Firstly, to confirm the additional cysteine in the collagen region, a genomic fragment that encompassed the predicted first and second collagen encoding exons and the intervening intron was amplified using primers PD-CYS-F and PD-CYS-R. The two collagen encoding exons were amplified with the cysteine codon predicted to lie at the 3' end of the second collagen encoding exon. Secondly, to confirm the three amino acid insertion and the N-linked glycosylation site, the predicted single exon that encodes the CRD was amplified using the primers PD-CRD-F and PD-CRD-R. Standard PCR assays (section 2.8.1) were performed containing 1.5 mM magnesium chloride, an annealing temperature of 55°C and an extension time of 2 minutes.

4.9.2. Analysing the Different Pig Breeds For The Sequence Variations

The PCR assay yielded single discrete bands on agarose gel electrophoresis and were purified and sequenced directly (section 2.8.2) on both strands using the amplification primers. The PCR product covering part of the CRD was of the expected 381 bp, as predicted from the cDNA sequence. The PCR product from the collagen region was approximately 600 bp in size, compared to 219 bp product predicted from the cDNA sequence. This indicates the presence of an intron between the first and second collagen encoding exons of porcine SP-D, of approximately 400 bp (Figure 4.2C). The genomic sequences obtained from the four different pig breeds confirmed the variations seen in the porcine SP-D cDNA sequence. The variants seen in porcine SP-D appear fixed across different pig breeds, which means that they are not due to mutations, polymorphisms or cloning artefacts. Therefore, porcine SP-D contains three major sequence differences in comparison to SP-D from other species.

4.10. Confirming the Sequence of the Other Porcine Surfactant Collectin: SP-A

Porcine SP-A (accession number L41350) is the only other reported cDNA sequence of a porcine collectin or surfactant protein. The sequence was submitted directly to GenBank in 1995, by SmithKline Beecham, without prior publication in a scientific journal. Protein and cDNA alignment of SP-A from eight different species revealed that the porcine SP-A CRD is one amino acid longer and contains a region of low similarity (Figure 4.5A). Closer examination of the cDNA sequence revealed that this disturbance could be a result of sequencing errors. A dinucleotide insertion of TG at position 498 and a single nucleotide insertion at position 535 (see Figure 4.5B), would explain the frame change and the presence of the additional amino acid. This was confirmed in this study by genomic sequencing of the porcine SP-A CRD encoding exon.

A.

```
   125                                                                  194
ne:LSFQEFMLAV GEKVFSTNGQ SVAFDVIREL CARAG-GRIA APRSPEENEA IASIVKKHNT YAYLGLVEGP
po:LSFQEFMLAV GEKVFSTNGQ SVAFWMSLES CVPEQVGRIA APRSPEENEA IASIVKKHNT YAYLGLVEGP
gp:LSLQGSMLAV GDKVFATNGQ SVDFNAIKET CARAG-GDVA APRNSEENTA ISSIVKKYNI YSYLGLTEGH
hu:LSLQGSIMTV GEKVFSSNGQ SITFDAIQEA CARAG-GRIA VPRNPEENEA IASFVKKYNT YAYVGLTEGP
rb:LSLQGSMKAV GEKIFSTNGQ SVNFDAIREV CARAG-GRIA VPRSLEENEA IASIVKERNT YAYLGLAEGP
rt:LSLQGSMLSV GDKVFSTNGQ SVNFDTIKEM CTRAG-GNIA VPRTPEENEA IASIAKKYNN YVYLGMIEDQ
do:LSLHESLLVV GRKVFSSNAQ SINFNDIQEL CAGAG-GQIA APMSPEENEA VASIVKKYNT YAYLGLVESP
mo:LSLQGSMLSV GDKVFSTNGQ SVNFDTIREM CTRAG-GHIA APRNPEENEA IASITKKYNT YPYLGVIEGQ

   195                                                          249
ne:TAGDFFYLDG TPVNYTNWYP GEPRGRGKEK CVEMYTDGQW NDRNCQQYRL AICEF.
po:TAGDFFYLDG TPVNYTNWYP GEPRGRGKEK CVEMYTDGQW NDRNCQQYRL AICEF.
gp:TPGDFHYLDG SPLNYTNWYP GEPRGRGKEK CAEMYLDGTW NDKNCLQSRL TICEF.
hu:SPGDFRYSDG TPVNYTNWYR GEPAGRGKEQ CVEMYTDGQW NDRNCLYSRL TICEF.
rb:TAGDFYYLDG DPVNYTNWYP GEPRGQGREK CVEMYTDGKW NDKNCLQYRL VICEF.
rt:TPGDFHYLDG ASVNYTNWYP GEPRGQGKEK CVEMYTDGTW NDRGCLQYRL AVCEF.
do:DSGDFQYMDG APVNYTNWYP GEPRGRGKEQ CVEMYTDGQW NNKNCLQYRL AICEF.
mo:TPGDFHYLDG ASVNYTNWYP GEPRGRGKEK CVEMYTDGKW NDKGCLQYRL AICEF.
```

B.

```
   474                                                                  553
ne:ACCAATGGGCAGTCAGTCGCTTTT--GATGTCATTAGAGAGTTGTGTGCCAGAGCAGGTGG-ACGCATCGCTGCCCCAAG
po:ACCAATGGGCAGTCAGTCGCTTTTTGGATGTCATTAGAGAGTTGTGTGCCAGAGCAGGTGGGACGCATCGCTGCCCCAAG
gp:ACTAATGGCCAGTCGGTGGACTTT--AATGCCATTAAAGAGACTTGTGCCAGAGCAGGTGG-TGATGTTGCTGCCCCAAG
hu:AGCAATGGGCAGTCCATCACTTTT--GATGCCATTCAGGAGGCATGTGCCAGAGCAGGCGG-CCGCATTGCTGTCCCAAG
rb:ACCAATGGGCAGTCGGTCAATTTT--GATGCCATCAGAGAGGTATGCGCCAGAGCAGGTGG-CCGCATTGCCGTCCCGAG
rt:ACCAATGGGCAGTCAGTCAACTTT--GATACCATTAAAGAGATGTGTACCAGAGCAGGAGG-CAACATTGCTGTCCCGAG
do:AGCAATGCGCAGTCCATTAATTTC--AACGACATTCAAGAGTTATGTGCCGGGGCAGGCGG-CCAAATTGCTGCCCCGAT
mo:ACCAATGGGCAGTCAGTCAACTTT--GATACCATTAGAGAGATGTGTACCAGAGCCGGTGG-CCACATTGCTGCCCCGAG
```

C.

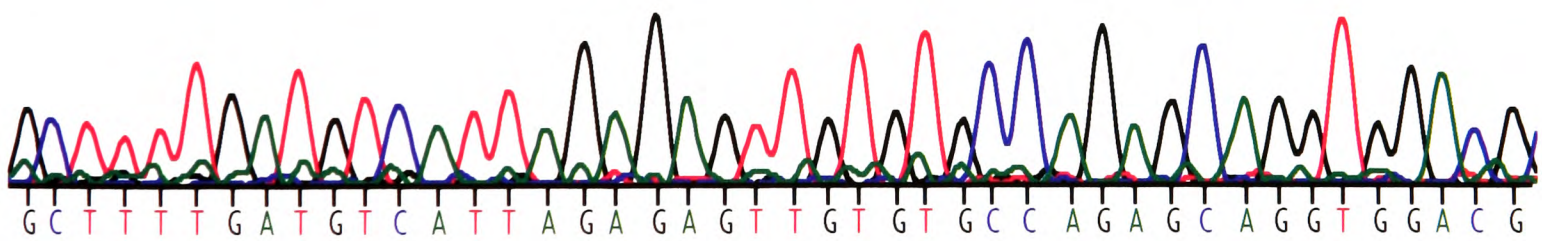


Figure 4.5. Correction of the porcine SP-A sequence. **(A)** Protein alignment of the CRD region, encompassed by the CRD encoding exon. The top line shows the corrected porcine sequence (ne) above the original porcine SP-A sequence (po). These are aligned against guinea pig (gp), human (hu), rabbit (rb), rat (rt), dog (do) and mouse (mo). The underlined region is aligned at the DNA level in **(B)**. **(C)** Part of the fluorescent dye DNA sequencing trace for porcine SP-A, covering the underlined region in (B). Amino acids and nucelotides conserved across four or more species are shown in bold. Dots represent sequence derived from the PCR primers. Numbering refers to the original porcine SP-A sequence.

A standard PCR assay (section 2.8.1) was performed containing 1.5 mM magnesium chloride, an annealing temperature of 55°C, an extension time of 2 minutes and using the primers PSPA-F and PSPA-R (see Appendix 5). Part of the porcine SP-A CRD was amplified from the four different pig breeds (as used for sequence variant confirmation of SP-D in section 4.9.1) and a consensus sequence was constructed. The PCR derived sequence of porcine SP-A is shown in alignment with the original GenBank sequence and those from other species (Figure 4.5). The corrected sequence of SP-A can be found in GenBank under the accession number AF133668 and is shown in Appendix 8 including the sequence annotations.

4.11. Chromosomal Localisation of Porcine SP-A and SP-D

The genes for SP-A and SP-D were mapped using a porcine/rodent somatic cell hybrid panel (Yerle *et al.*, 1996). Two PCR assays for SP-A and SP-D were developed to generate PCR fragments as sequence tag sites (STSs). The PCR assays were optimised to selectively amplify the porcine genes from the somatic cell hybrid panel over the background of hamster genes. This was achieved by optimising the temperature, magnesium concentration, number of amplification cycles and buffer conditions. For SP-A, a standard PCR assay was selected using 1.5 mM magnesium chloride, 33 cycles, an annealing temperature of 64°C and a 1 minute extension time, using the primers, PSPA-F and PSPA-R. For SP-D, the buffer conditions were changed to 16 mM $(NH_4)_2SO_4$, 67 mM Tris-Cl, pH 8.8, 0.01% (v/v) Tween-20 containing 1.5 mM magnesium chloride with the primers PD-4F and PD-4R. The SP-D PCR thermal cycling conditions were identical to those for SP-A except an annealing temperature of 61°C was used. PCR was performed on 50 ng of porcine/hamster somatic cell line genomic DNA or 50 ng of hamster and porcine genomic DNA as negative and positive control template, respectively.

The porcine SP-A STS amplified a 418 bp product from the CRD, covering the region sequenced (section 4.10). The STS for porcine SP-D amplified a 257 bp product from

the final exon encoding the CRD and 3' UTR, nucleotides 1002-1258. An additional fragment of approximately 600 bp was also amplified from most cell lines. However, this product was ignored because it was also present in the negative control of hamster genomic DNA. Chromosomal assignment to a porcine chromosome and statistical analysis was performed on a computer database operated over the Internet (*www.toulouse.inra.fr/lgc/pig/hybrid.htm*) (Chevalet *et al.*, 1997).

Both STS primer pairs, for SP-A and SP-D, gave the same pattern of positive and negative signals across the different somatic cell hybrid lines; hybrids 12, 20, 22 and 26 tested positive for both SP-A and SP-D (Figure 4.6). This pattern of positives reliably assigns both of the genes, SP-A and SP-D, to porcine chromosome 14 (SSC14) with a high degree of confidence (a correlation coefficient of 0.87 and a chromosomal probability greater than 98%). This chromosome assignment is consistent with the mapping of the surfactant collectins to human chromosome 10 (Hoover and Floros, 1998) and murine chromosome 14 (Motwani *et al.*, 1995; Moore *et al.*, 1992) because these chromosomes contain regions of synteny.

4.12. Discussion

This chapter describes the cloning and characterisation of the full length cDNA sequence for porcine SP-D. The presence of interesting sequence variations found in porcine SP-D, have been confirmed in other pig breeds. The sequence of porcine SP-A cDNA has been corrected. In addition, the porcine genes for both SP-A and SP-D have been chromosomally mapped.

4.12.1. The Cloning of Porcine SP-D cDNA

Porcine SP-D has been cloned and sequenced at the cDNA level. The full-length sequence was obtained from two overlapping lambda bacteriophage clones and 5' RACE technology. All the clones overlapped by at least 96 bp and showed complete identity across the overlapping regions. The bacteriophage clones were

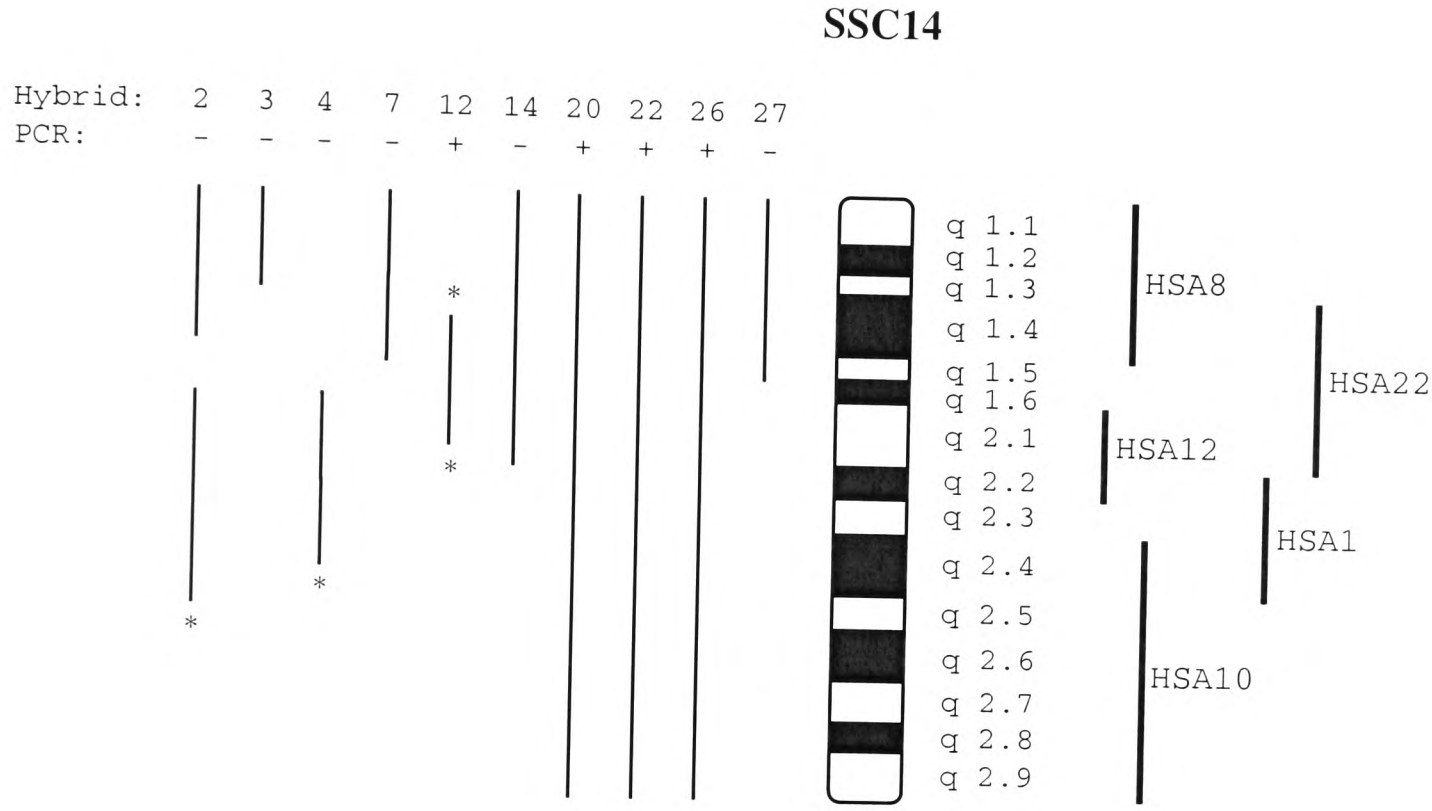


Figure 4.6. Chromosomal localisation of the porcine SP-A and SP-D genes. Solid bars on the left show the regions of SSC14 found in the hybrid cell line indicated by the number above. Only those hybrid lines that contain part of SSC14 are shown, asterisks indicate SSC14 regions in the hybrid cell lines that do not correspond with G-band limits (adapted from Yerle *et. al.* 1996). The "+" or "-" sign above each bar denotes whether that hybrid tested positive or negative for SFTPA and SFTPD. Solid bars on the right of the ideogram represent regions of SSC14 which are homologous to human chromosomes (adapted from Rettenberger *et al.* 1995)

sequenced on both strands and confirmed with at least three independently primed sequencing reactions. Ten different 5' RACE clones were sequenced on both strands to confirm the sequence of the 5' region of the cDNA missing from the bacteriophage library clones.

The full length cDNA for porcine SP-D is composed of a 43 bp 5' UTR, a 1134 bp open reading frame and 172 bp of 3' UTR including the opal stop codon, TGA. Two potential polyadenylation sites are present in the 3' UTR, 25 and 117 bp upstream of the polyadenosine tract (see Figure 4.1).

The coding region of porcine SP-D contains a 20 amino acid hydrophobic leader peptide. This is followed by a 25 amino acid cysteine rich amino-terminal region believed to be important for higher order oligomerisation of the basic trimeric unit of the collectin peptide chain (Brown Augsburger *et al.*, 1996a). The collagen-like region immediately follows the N-terminal region is composed of 60 repeating Gly-Xxx-Yyy collagen subunits, the same length in human, murine and rat SP-D. The collagen-like region in porcine SP-D probably shares the same exon-intron boundaries as those known for murine and human SP-D (Lawson *et al.*, 1999; Crouch *et al.*, 1993b), which are composed of four exons, each of 117 bp. While, the intron-exon boundary of bovine SP-D probably resembles the gene structure of bovine conglutinin with two shorter collagen-encoding exons of 108 bp, as judged by sequence alignment (Lim *et al.*, 1993).

The trimerisation of the collagen region of the collectins is believed to be initiated by the neck region, an alpha-helical coiled coil forming protein sequence (Hoppe *et al.*, 1994; Brown Augsburger *et al.*, 1996a). The neck region of porcine SP-D conforms to the requirements for forming this structure (see Figure 4.4). It is a 28 amino acid region composed of heptad repeats with hydrophobic amino acids positioned at residues +1 and +4.

The crystal structures of neck-CRD fragments of MBL and SP-D (Weis *et al.*, 1992; Hakansson *et al.*, 1999) together with alignments of the collectins, clearly indicate conserved residues within the CRD that are essential for the co-ordination of Ca^{2+}-ions and hydroxyl groups from carbohydrate, for imparting lectin function (shown in Figure 4.3). The CRD of porcine SP-D conforms to these primary structure requirements, despite the presence of an additional potential N-linked glycosylation site and an insertion of three extra amino acids. The insertion of the extra sequence does not disturb the single exon structure of the CRD, as judged by PCR sequencing. This is in agreement with the single CRD exon structure of MBL, conglutinin, SP-A and potentially CL-43 (see section 3.6.2).

4.12.2. The Differences in SP-D from Porcine compared to other Species

There are three prominent sequence differences between porcine SP-D and the SP-D from other species. Each of these differences may have important functional and structural consequences for porcine SP-D. Despite these sequence differences, this molecule can be regarded as the porcine homologue of SP-D as observed from the high sequence identity to bovine SP-D, rather than a closely related collectin, such as pig conglutinin or CL-43. It remains to be shown if porcine SP-D has any functional differences in comparison to SP-D derived from other species. The differences in porcine SP-D may reflect divergence of SP-D from other species or convergent evolution of porcine SP-D towards other members of the collectins.

4.12.2.1. The Extra Cysteine

The first difference in porcine SP-D is the presence of an extra cysteine residue. This results in an odd number of cysteine residues within each polypeptide chain. The extra cysteine residue within the collagen region is a feature not shared with the other known sequences of SP-D. Instead this feature draws parallels with other collectins. Indeed,

the extra cysteine is reminiscent of that found in one of the two polypeptide chains of human and baboon SP-A (Katyal *et al.*, 1992; Gao *et al.*, 1996) and bovine conglutinin (Lee *et al.*, 1991).

The extra cysteine residues in conglutinin and SP-A are hypothesised to have structurally important consequences for these collectins. The extra cysteine in conglutinin is at the beginning of the collagen region and appears by electronmicroscopy to form disulphide bands between trimers creating an elongated central hub at the point of radiation for the collagen stalks (Lu *et al.*, 1993b). The extra cysteine in human and baboon SP-A has been proposed to form interchain disulphide bonds which either stabilise or may be essential for the heteropolymeric chain structure of SP-A in these species (Malhotra *et al.*, 1994b). Only one polypeptide chain exists for porcine SP-D, it will be interesting to determine the effect that the extra cysteine will have upon the structure and function of this molecule in comparison to SP-D from other species.

4.12.2.2. The CRD Insertion

The second distinguishing feature of porcine SP-D from other species is an insertion of three amino acids in the CRD. This type of variation is not seen across the different species within any group of the collectins: all the CRDs across different species for each type of collectin have the same length. However, the CRDs between different members of the collectins are of different sizes. For example the CRD of SP-A is 124 amino acids long while the CRD of MBL is 123 amino acids. From the crystal structure of MBL (Weis *et al.*, 1992) and SP-D (Hakansson *et al.*, 1999) the CRD insertion lies within a loop region, loop 4 (Figure 4.7). By alignment of the collectins, SP-A has the shortest loop region. In comparison to SP-A, both porcine SP-D and CL-43 have an extra four amino acids inserted in loop 4, conglutinin has an extra three amino acids, while MBL and SP-D from all other species have only one extra amino acid within this loop. This is illustrated in Figure 4.8.

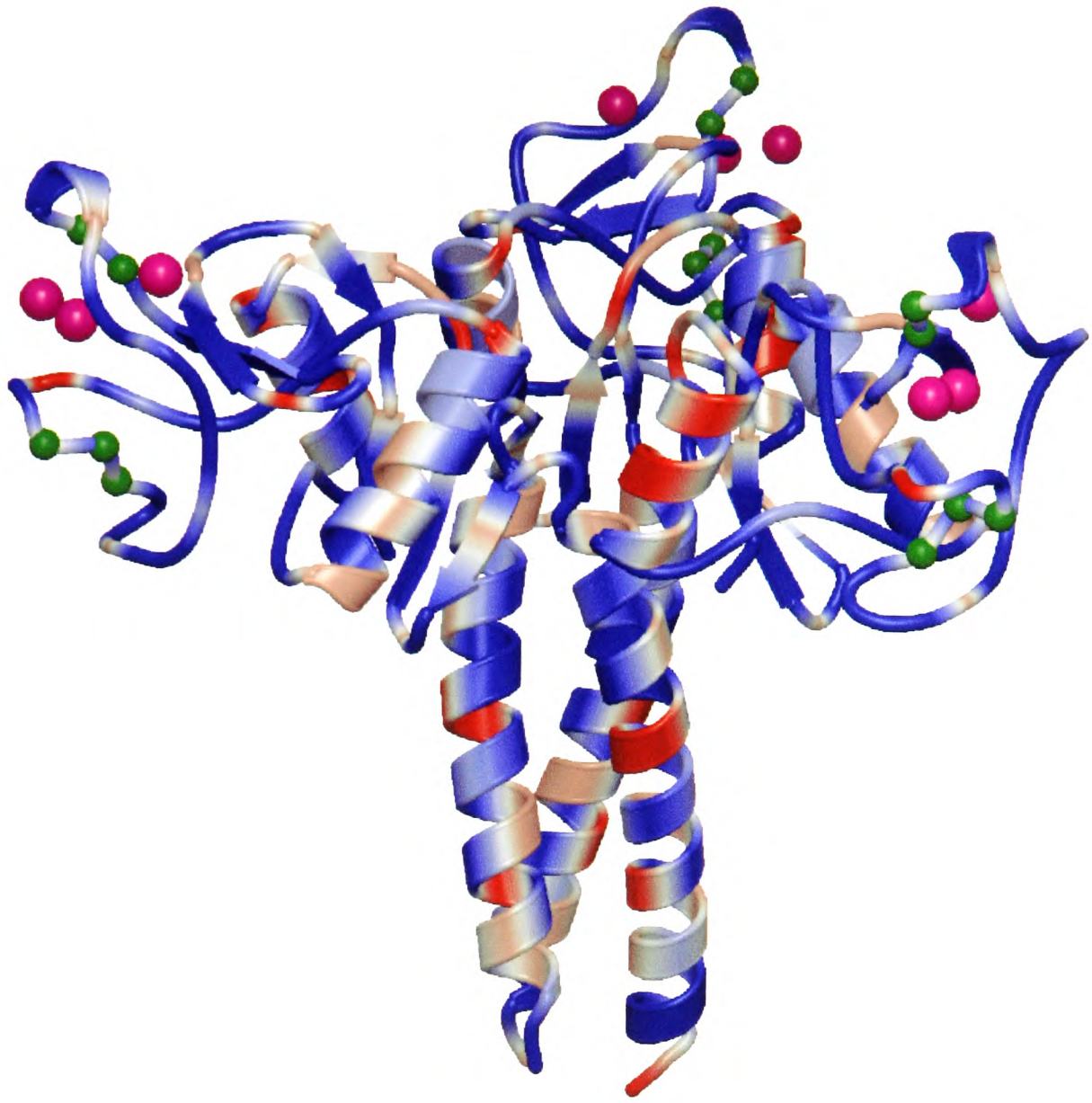


Figure 4.7. The X-ray crystal structure of the trimeric alpha-helical coiled-coil neck-CRD of human SP-D indicating sequence conservation and the sequence variations in porcine SP-D. Only the alpha carbon backbone of the neck-CRD is displayed. The colour scheme represents sequence conservation across five different species (human, rat, mouse, bovine and porcine). Blue represents conservation of sequence across all species, through white to red which indicates sequence conservation in only two species. Magenta balls represent the co-ordinated calcium ions. The two green balls on the alpha carbon backbone indicate the position of the three amino acid insertion in porcine SP-D (not modelled). The three green balls represent the N-linked glycosylation site novel to porcine SP-D. The crystal structure co-ordinates are taken from Hakansson *et al.*, 1999.

The insertion within the CRD initially cast doubt upon assigning this porcine cDNA as the porcine homologue of SP-D. However, porcine cDNA shows the highest homology to SP-D among the different collectins, with the greatest identity to bovine SP-D at both the cDNA and protein levels; across the CRD and the entire length of the molecule. The tissue distribution of porcine SP-D, as revealed by Northern blotting (performed by Martin van Eijk, Utrecht University, The Netherlands) showed an almost exclusive expression of a 1.3 kb transcript within the lungs, with a considerably reduced expression present in the small intestinal region of the digestive tract (Van Eijk *et al.*, 2000). Together this data suggests that the cDNA for porcine SP-D has been isolated.

4.12.2.3. The Additional N-linked Glycosylation Site

Porcine SP-D shows a third feature which is absent from SP-D from other species but shared with other collectins. Porcine SP-D shares a potential N-linked glycosylation site within the CRD with the C form of murine MBL and all the known sequences of SP-A. Alignment reveals that these N-linked glycosylation sites lie within the same region of the CRD, to lie within loop region 1 from the crystal structure of MBL (Weis *et al.*, 1992). The crystal structure of human SP-D is shown in Figure 4.7 with the potential glycosylation site (Asn-Phe-Thr) indicated by three green atoms. The N-linked glycosylation site and an amino acid insertion within the loop regions of CRD may impart functional disturbances on carbohydrate binding, especially considering the variation in loop 4 which contains the variable inserts amongst different collectins. Consequently, these variations within the CRD of porcine SP-D may affect the types of pathogens bound.

4.12.3. The Porcine Surfactant Collectin Cluster

The porcine genes for SP-A and SP-D were chromosomally localised by somatic cell hybrid mapping (Yerle *et al.*, 1996). The porcine chromosomal contribution to the

SP-D:

Porcine SP-D

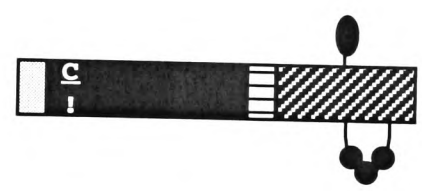


Other SP-D's:
Human, Mouse, Rat
and Bovine

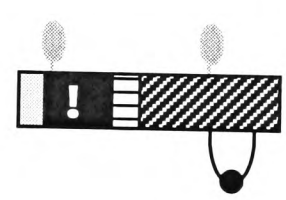


Other Collectins:

Conglutinin



MBL



CL-43



SP-A



100 Amino
Acids

Figure 4.8. A graphical comparison of the sequence of porcine SP-D with other collectins. Boxes indicate domains of the protein, as described in figure 4.2. The ball and sticks indicate potential N-linked glycosylation sites. The 'C' refers to the presence of a cysteine in the collagen-like region. The exclamation mark indicates a disruption in the collagen region. The loop indicates loop 4 in the MBL X-ray crystal structure, the beads on the loop refer to the number of extra amino acids contained in loop 4, as compared to the smallest predicted loop 4 region of only 7 amino acids in SP-A. Shaded symbols for cysteines and glycosylation sites indicate motifs which are not conserved across all species examined.

porcine/hamster hybrid panel was originally defined by the gross anatomy using G-banding and FISH followed by PCR based assignment of genes. Smaller fragments of porcine chromosome within the hybrid cell panel which escaped G-banding and FISH analysis have been defined using PCR based gene assignment. The two porcine surfactant collectin genes, SP-A and SP-D show the same mapping pattern across the somatic cell hybrids. The porcine surfactant collectins have been assigned to porcine chromosome 14 with a correlation coefficient of 0.87 and a chromosomal probability greater than 98% (analysed by Thomas Skinner and Alan Archibald, Roslin Institute, Scotland). The chromosomal localisation of the porcine surfactant collectins is limited to chromosome 14 rather than a specific sub-region of chromosome because of at least one discordant result in the positive mapping panel (see hybrid 12 in Figure 4.6). The discordant result is probably caused by hybrid 12 containing a small fragment of the distal region of chromosome 14 which escaped G-banding or FISH characterisation. Zoo-FISH studies between porcine and human chromosomes (Rettenberger *et al.*, 1995) have revealed the syntenic regions between porcine chromosome 14 and human chromosomes 1, 8, 10, 12 and 22 (Figure 4.6). The human surfactant collectin gene cluster has been localised to HSA10q22.2-23.1 (Hoover and Floros, 1998). Porcine SP-A and SP-D can be predicted to be localised at SSC14q2.3-qter as this region is syntenic with the long arm of human chromosome 10 (HSA10q) (Goureau *et al.*, 1996).

4.12.4. Porcine Surfactant Protein A

Alignment of the original porcine sequence of SP-A (accession number L41350) against SP-A from different species revealed that the predicted protein sequence contains an extra amino acid in the CRD and a region of 12 amino acids in the CRD with little similarity. Examining the cDNA sequence revealed two frame insertions that may account for a stretch of sequence being out of frame plus the extra amino acid present. By sequencing part of gene encoding SP-A from four different pig breeds, the insertions were confirmed as sequencing errors. The corrected sequence has been

submitted to GenBank under the accession number AF133668, a copy is shown in Appendix 8. The corrected sequence of porcine SP-A will be important for constructing recombinant proteins of this innate immune protein.

4.13 Perspective

The 9 kDa increase in molecular weight seen for porcine SP-D in comparison with SP-D from other species could not be completely explained by examining the cDNA sequence. The insertion of three amino acids would attribute approximately 0.23 kDa and an additional N-linked carbohydrate could account for approximately 2 kDa, based upon the size of a typical N-linked carbohydrates. To identify the source of the molecular weight increase in porcine SP-D, the protein should be examined. The molecular weight variation can be attributed to combination of post-translation modifications, such as N- or O-linked glycosylation or other amino acid modifications. The molecular weight variation could also be attributed to a small post-translational modification that leads to an anomalous electrophoretic migration on SDS-PAGE.

Porcine surfactant is a valuable therapeutic agent for human surfactant disorders. This is the first detailed molecular description for two of the surfactant protein constituents of porcine surfactant. The initial characterisation of the proteins of porcine surfactant at the cDNA level will be useful for designing recombinant molecules that could easily be used in existing porcine animal models of human respiratory infectious diseases (Kamp *et al.*, 1997; Van Reeth *et al.*, 1998).

SP-A and SP-D have both been recognised as potential markers for monitoring particular pulmonary disease in humans (Kuroki *et al.*, 1998). Given that respiratory infections are the most frequent health problem in the pig breeding industry (Christensen and Mousing, 1992), it is hoped that information pertaining to the lung defence systems which influence lung infection and inflammation, such as the collectins, may lead to the development of potential markers for disease and immunity.

Furthermore, it is coming to light that the innate immune system plays an important contribution towards delayed xenograft rejection (Hancock, 1997). A detailed knowledge of porcine innate immunity and non-self recognition systems of the lung may be a useful key step towards improved xenotransplantation.

CHAPTER 5: CHROMOSOMAL ASSIGNMENT OF MURINE COMPLEMENT COMPONENTS USING A NOVEL APPROACH

5.1. Introduction and Aims

This chapter describes a novel and rapid approach for genetic mapping of murine genes that was developed to localise genes that had not been cloned at the cDNA or gene. This novel approach uses murine EST data, intron prediction and restriction fragment length polymorphism (RFLP) to localise murine genes. This method is applicable to any murine gene where the cDNA or gene have not been cloned but where suitable murine EST information exists. This new approach for genetic mapping has been used for the chromosomal mapping of three complement genes (MASP2, C1r and C1s). The chromosomal localisation of these genes has shed light upon the molecular evolution of the MASPs/C1s/C1r family, from the chromosomal distribution of these genes in the mouse. The murine gene for C3aR was also mapped using a simplified version of the technique, relying upon known gene structure instead of EST information and has uncovered an unexpected complement gene cluster.

Current techniques for gene localisation all require the gene of interest to be cloned at the gene or cDNA level. Chromosomal mapping also requires highly specialised expertise for fluorescent *in situ* hybridisation (FISH) and equipment. While genetic mapping requires the identifying polymorphic markers associated with the gene by characterising the gene at the sequence level or by using the gene to find a restriction fragment length polymorphism. This novel technique circumvents these problems and in addition, maps known expressed genes to be mapped onto genetic maps that are mainly composed of anonymous nucleotide repeats that generally have no known associated gene.

The starting point for this novel technique involves identifying murine ESTs where the gene that has not been cloned or mapped but where gene or cDNA homologues

exist in other species. Murine ESTs homologues of the genes from other species were identified in the murine EST computer database at NCBI (*www.ncbi.nlm.nih.gov*, see section 2.9.2). The murine ESTs for the gene of interest were aligned and the exon-intron boundaries were predicted from the gene structures of other species or related genes. A PCR assay was designed to amplify a fragment of the murine gene that included two exons and the intervening intron, using primers designed on the murine EST alignment. The fragment was amplified from both parental mouse strains used for genetic mapping and screened for potential RFLPs. The RFLP was discovered using a series of restriction enzymes and used to define the two different parental alleles. By examining the inheritance of the alleles across a panel of backcross mice from the European Collaborative Intraspecific Backcross (EUCIB) used for genetic mapping, a chromosomal localisation could be assigned for the gene.

A description of the method and its use in mapping complement genes has been published in the journal, International Immunology (Lawson and Reid, 2000a).

5.1.1 Testing the Approach: Genetic Mapping of Murine Complement Genes

To test this approach, either protein systems or protein families were regarded as ideal candidates for investigation. Because of the familiarity of the laboratory with the complement system, the complement system was examined as a potential model system to test this approach. The murine mapping database (*www.informatics.jax.org*), literature and the murine EST database were used to identify several murine complement proteins that had neither been cloned nor mapped, but for which murine ESTs were known for these components.

At the beginning of the study, out of approximately 40 genes which encode for the complement system proteins in mice, nine had not been chromosomally mapped in mice (see Table 5.1). Of the unmapped complement genes, only six murine complement genes could be mapped using this technique. At the time of conducting

the research murine ESTs for C8α were not available, there are no murine ESTs for C1qRp and the only EST for the alpha chain of complement receptor 3 (CR3) does not cover an intron-exon boundary. Four genes, *C1r*, *C1s*, *C3aR*, and *Masp2* were mapped using this technique[1] and the location of *Masp1*[2] was confirmed.

Murine Complement Component	Murine ESTs	Human Localisation	Potential Murine Localisation
C1qRp[3]	None	Unknown	Unknown
C1r	Yes	12p13	6
C1s	Yes	12p13	6
C3aR[4]	Yes	12p12-p13.3	6
C5aR[1]	Yes	19q13.3-q13.4	17
C8, α subunit	None	1p32	4
C9[1]	Yes	5p13	15
CR3, α chain (Cd11a)	Yes	16q11.2	7
MASP2[4]	Yes	1p36.3-p36.2	4

Table 5.1. Murine complement genes that have not cloned or mapped, prior to this study. The existence of potential ESTs are indicated. The possible murine chromosomal assignment is indicated, based upon human chromosomal mapping.

5.2. Identification of the Murine Complement Homologues

Murine ESTs for C1r, C1s and MASP2 were found by searching the murine EST database (dbEST at the National Centre for Biotechnology Information) with the human homologues of these cDNAs using BLAST 2.0 at NCBI (section 2.9.2). The ESTs were aligned against the human homologues using Seqman (DNAstar, London, UK) then adjusted by eye (section 2.9.3). All alignments of the partial cDNA sequences derived from murine ESTs showed between 74.6% to 80.2% identity at the

[1]Because of set backs in developing a reliable PCR for C5aR and C9 and due to the success and interesting finding that emerged when localising the other genes, it was decided not to continue with attempts at localising the murine C5aR and C9 genes.
[2]The original or alternative gene name for murine MASP1 is *Crarf*, which stands for C4/C2 activating component of Ra-reactive factor. In the mouse, MASP1 protein was originally termed P100.
[3]At the time of conducting the research, the murine gene for C1qRp was not sequenced (Norsworthy *et al.*, 1999) otherwise this could have been mapped in a simalar manner to *C3aR*.
[4]It is important to note that at the time of the investigation, neither the murine gene for MASP2 (Stover *et al.*, 1999b) was sequenced nor were the human genes for MASP2 and C3aR localised (Stover *et al.*, 1999a; Hollmann *et al.*, 1998).

nucleotide level with the human counterparts. The murine MASP2 EST alignments was shown to be identical to the murine cDNA sequence published later (Stover *et al.*, 1999b; Endo *et al.*, 1998)

The murine gene for MASP1 was placed upon the linkage map for comparison to the location of *Masp2*, *C1r* and *C1s*. The murine cDNA for MASP1 is known (Takahashi *et al.*, 1993) and the intron-exon boundaries were predicted from the human gene structure of MASP1 (Endo *et al.*, 1996a). The prediction of introns in murine C1s were determined from the human gene structure (Tosi *et al.*, 1989). In the case of *Masp2* and *C1r*, introns were predicted from the relatively conserved intron-exon boundaries within protein modules. Alternatively, gene structures from other species could also be used for this prediction.

5.3. Primer Design

Primers were designed to generate PCR products which spanned two predicted exons and the intervening intron. Introns that were between one and two kb in length were chosen for amplification to facilitate amplification of a single PCR product with sufficient yield for visualisation. ESTs from dbEST that covered suitable exon-intron boundaries of the genes for MASP2, C1r and C1s were AA895155, AA427242 and AA798057, respectively (see approximate primer location in Figure 5.1). For MASP1, primers were designed from the known murine cDNA sequence, accession number D16492 (Takahashi *et al.*, 1993). For C3aR, the murine gene sequence was known (Tornetta *et al.*, 1997; Hsu *et al.*, 1998; Hollmann *et al.*, 1998). The only intron in the gene is over 4 kb in length, so primers were designed to amplify a 663 bp region from the 5' end of exon 2 of the murine C3aR gene including ~200 bp of exonic sequence (accessions U77461, U97537 and AF053757), running from nucleotide number 31 to 693 from U77461.

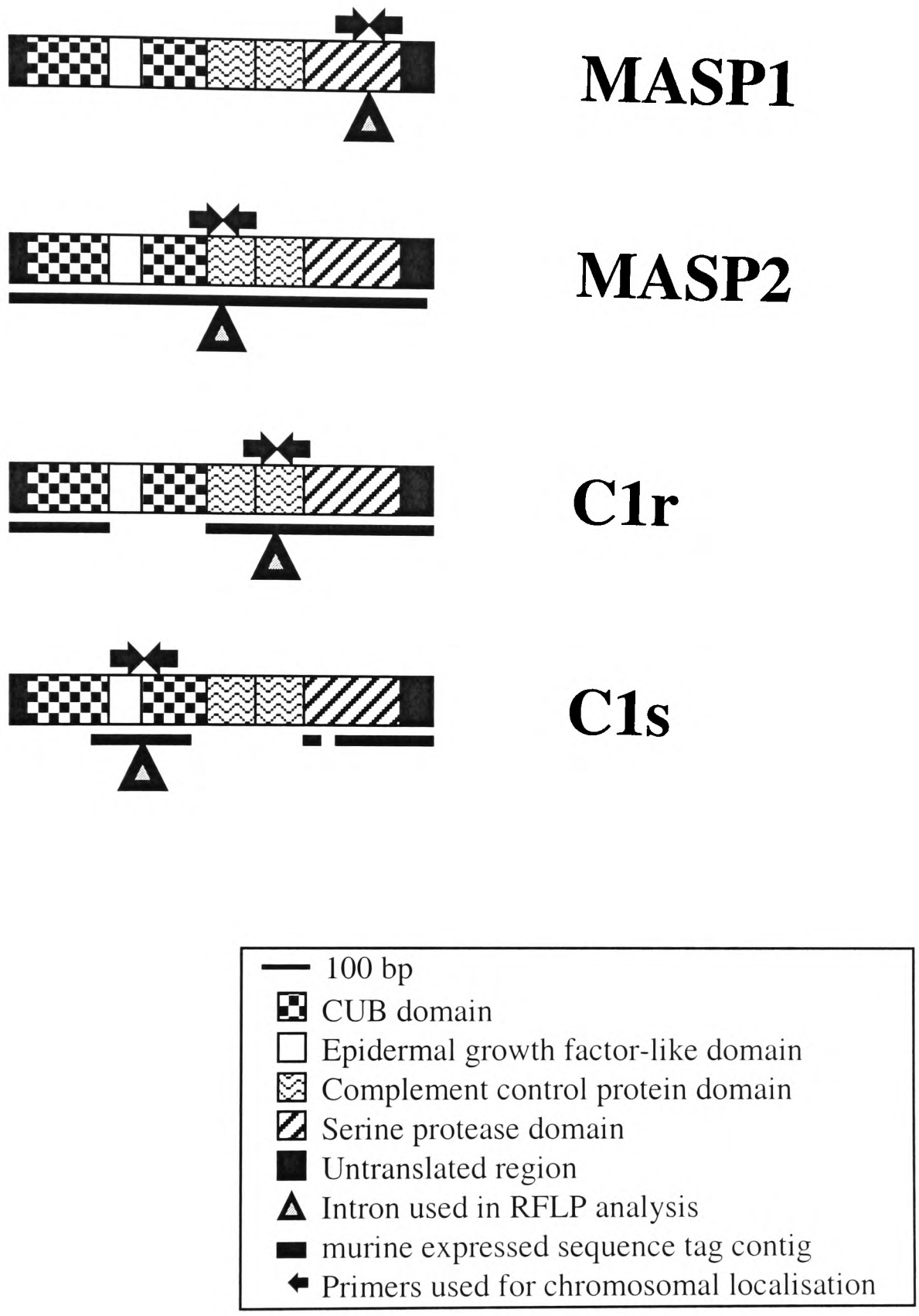


Figure 5.1. The domain structures of the MASP1, MASP2, C1r and C1s. The protein domains of the MASPs, C1r and C1s are superimposed onto the mRNA (shaded boxes), including the position of primers (arrows) used for chromosomal localisation, position of the introns identified by PCR (arrow head) and the extent of the murine EST contigs (horizontal bars). A key for the figure is shown, including the scale.

5.4. Amplification of the Murine Complement Gene

Each PCR assay was optimised (see Table 5.2) in order to generate a single distinct product. The PCR assay was performed on genomic DNA samples from the two different inbred mouse species (C57BL/6 and *Mus spretus*). Standard PCR assays (section 2.8.1) were carried out at the optimised annealing temperature for each set of primers (Appendix 9), with an extension time of 1 or 2 minutes using either 1.25 mM or 1.5 mM $MgCl_2$ (shown in Table 5.2). Parental Genomic DNA (25 ng) from either of the inbred strains was used as template. A fraction of the reaction product (10 µl) was run on a 1% (w/v) agarose gel and visualised by ethidium bromide staining. The PCR products for each gene gave a discrete single band (see the top of Figures 5.2 and Figure 5.3) and these were subcloned and sequenced (section 5.5).

Intronic DNA was chosen for amplification, since it is more divergent than exonic DNA between different species and strains, hence it was assumed to be a potent source of RFLPs between the two species of mice used in backcross analysis (C57BL/6 and *Mus spretus*). Amplification products were designed to include at least 200 bp of exonic DNA. This was used in sequence analysis to confirm that the correct gene was being amplified from the murine genome.

5.5. Sequence Confirmation of the PCR Products used for Localisation

The PCR products used in genetic mapping were confirmed as the desired gene by sequencing. The PCR product was first cloned into a 3' deoxythymidine residue overhang plasmid (pCRII-TOPO, Invitrogen. See section 2.8.4). For each gene of interest, one plasmid bearing the amplified section of the gene was selected for sequence analysis. Fluorescent dye-primer cycle sequencing was used to confirm the identity of the PCR product (section 2.7.2).

The PCR fragment of the murine C3aR gene was in agreement with the known gene sequence (Hollmann *et al.*, 1998; Tornetta *et al.*, 1997; Hsu *et al.*, 1998). The exonic

sequences of the PCR fragments of MASP2, C1r and C1s were in agreement with the EST alignments and the cDNA sequences (Stover *et al.*, 1999b; Endo *et al.*, 1998).

Gene	Tm (°C)	Primer	Mg^{2+} (mM)	Elongation Time (seconds)	Restriction Enzyme	Allele Size (bp) MS	B6
Masp1	61	Crarf-1f Crarf-1r	1.5	60	*Bsu*36 I	2000	1500 + 500
C1s	61	C1s-1f C1s-1r	1.5	60	*Bsu*36 I	1000	800 + 200
C1r	61	C1r-2f C1r-2r	1.5	60	*Ava* II	1300 + 500	700 + 600 + 500
C3aR	61	C3aR-1f C3aR-1r	1.5	60	*Apa*L I	450 + 250	700
Masp2	55	Masp2-3f Masp2-3r	1.25	120	*Xho* I	2300	1400 + 900

Table 5.2. Conditions used for chromosomal localisation of each complement gene. The conditions of the PCR assay are listed, including the length of the elongation step (in seconds) and the annealling temperature of each primer pair (T_m). The sequences for the primers can be found in Appendix 9. The enzyme used for detecting the polymorphism is shown, along with approximate allele sizes. MS and B6 indicates the mouse strains *Mus spretus* and C57BL/6, respectively.

5.6. Screening for Restriction Fragment Length Polymorphisms

To map the gene of interest using backcross mice, the inheritance of the different parental alleles of the gene were scored on a panel of backcross mice. The alleles were identified by the RFLP contained within the PCR product derived from the gene of interest.

The RFLPs between the two mouse species were discovered by screening with forty one different restriction enzymes that each had six or five base pair recognition sites (see Table 5.3). For speed and efficiency, 10 µl of the PCR product generated from the different mice was used directly from the PCR assay without purification in a 20 µl restriction digestion using 10 U of each enzyme (section 2.2.9.1). The presence of restriction polymorphisms was then judged by running the digestion products on a 1% (w/v) agarose gel and comparing the two different alleles.

Acc I	*Apa*L I	*Asp*718 I	*Ava* I	*Ava* II
*Bam*H I	*Bcl* I	*Bgl* I	*Bgl* II	*Bst*B I(*Sfu* I)
*Bst*E II	*Bst*Y I	*Bsu*36 I	*Cla* I	*Eco*0109 I
*Eco*R I	*Eco*R V	*Hinc* II	*Hind* III	*Kpn* I
Mlu I	*Nar* I	*Nco* I	*Nde* I	*Nhe* I
Not I	*Pst* I	*Pvu* II	*Sac* I (*Sst* I)	*Sac* II
Sal I	*Sca* I	*Sma* I	*Spe* I	*Sph* I
Ssp I	*Stu* I	*Sty* I	*Tth*111 I	*Xba* I
Xho I				

Table 5.3. Restriction endonucleases used in screening for restriction fragment length polymorphisms.

The use of forty one different restriction enzymes with either five or six base pair recognition sites generally produced at least one RFLP between the two alleles. An RFLP was chosen if it could be clearly distinguished on a 1% (w/v) agarose gel. Details of the enzymes used for each gene and the allele sizes for each species are detailed in Table 5.2 and examples of the restriction pattern observed for parental and heterozygous animals are shown at the top of figures 5.2 and 5.3.

Between the mouse strains C57BL/6 and *Mus spretus* an *Apa*L I polymorphism was found in the C3aR gene, *Masp2* has a *Xho* I polymorphism, *C1r* has an *Ava* II polymorphism and both *C1s* and *Masp1* have *Bsu*36 I polymorphisms. The same restriction fragmentation pattern was seen with both the purified PCR product and the

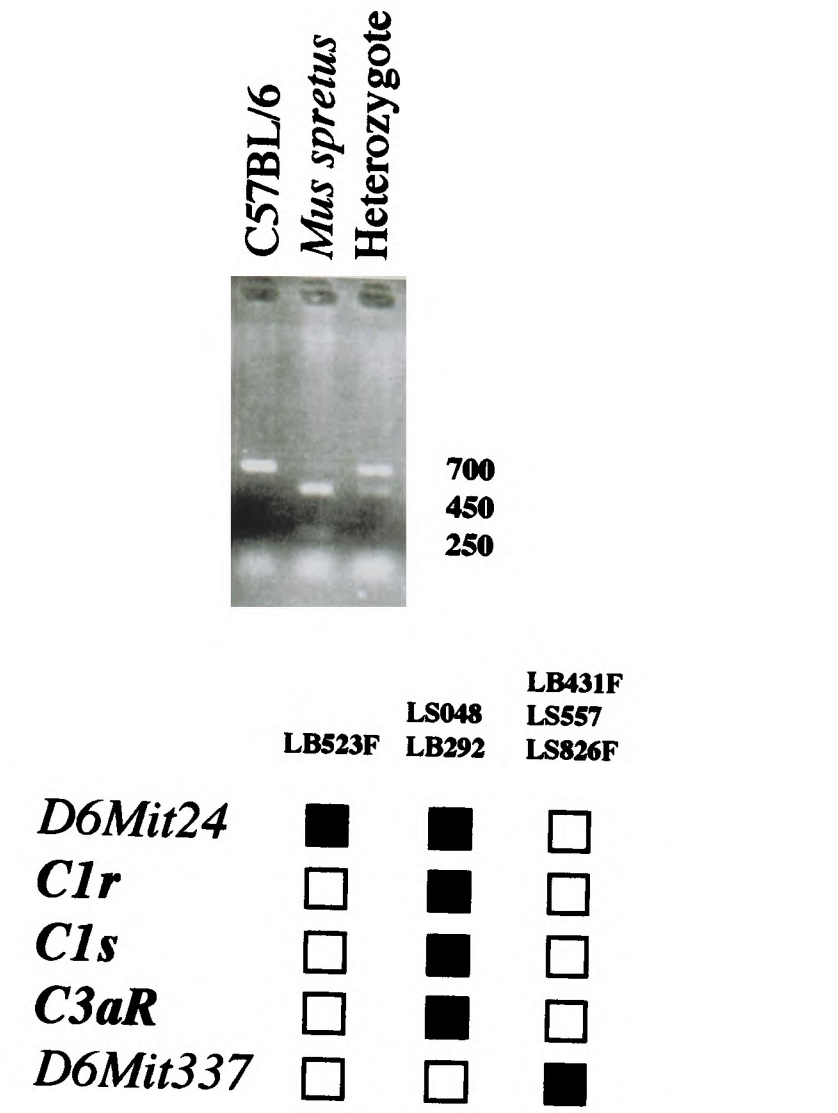


Figure 5.2. Chromosomal localisation of the genes for C3aR, C1r and C1s. (TOP) A 1% (w/v) agarose gel of the digested PCR products from the C3aR gene is shown for the wildtype mice (C57BL/6 and *Mus spretus*) and a heterozygote. Molecular sizes (bp) of the alleles are shown to the right. (MIDDLE) Haplotyping of informative recombinant mice for the genes and flanking markers are shown. EUCIB mice numbers are listed above the columns of haplotype boxes, mice heterozgous for the parental alleles are shown as filled boxes, open boxes represent a homozygous haplotype. (BOTTOM) A graphical representation of the chromosomal location of the genes, shown in centimorgans (cM) to the left, an expansion of the region is shown to the right, including the neighbouring markers.

unpurified PCR product. The polymorphisms identified in the PCR products of these complement genes were then used for genetic mapping of those loci.

5.7. The Intron-Exon Boundary of the Serine Protease Domain of MASP2

Alignment of murine ESTs with the human cDNA sequence for MASP2, revealed several ESTs which spanned the serine protease domain (summarised in Figure 5.1). Murine ESTs from dbEST were aligned across the part of the cDNA sequence encoding the serine protease domain of human MASP2. All ESTs identified by database screening were aligned using the human sequence as reference. The murine MASP2 serine protease domain protein sequence shared an identity of 79.9% and 76.1% with that of human MASP2 protein and cDNA, respectively. The murine MASP2 EST alignment showed identity to the later published cDNA sequence (Stover *et al.*, 1999b; Endo *et al.*, 1998).

The resulting coding sequence for the serine protease domain of murine *Masp2* was initially used for developing a PCR-RFLP assay by predicting intron-exon boundaries from the human MASP1 gene (Endo *et al.*, 1996a), see Figure 5.1. PCR assay across this region revealed a surprising feature of the murine MASP2 gene. PCR assays with two independent primer pairs across two predicted *Masp2* introns within the serine protease domain, yielded products from genomic DNA that were of the same size as the EST assembled murine cDNA sequence (results not shown). This suggested that the serine protease domain of murine MASP2 is encoded by a single uninterrupted exon.

To further understand this finding, the entire serine protease encoding region of *Masp2* was cloned by PCR from murine genomic DNA. PCR primers were designed at the 5' and 3' end of the murine MASP2 serine protease domain from murine ESTs (AA244853 and AA530523). A standard PCR assay (section 2.8.1) was optimised. PCR assays were carried out on ~25 ng of genomic DNA from a *Mus spretus* mouse,

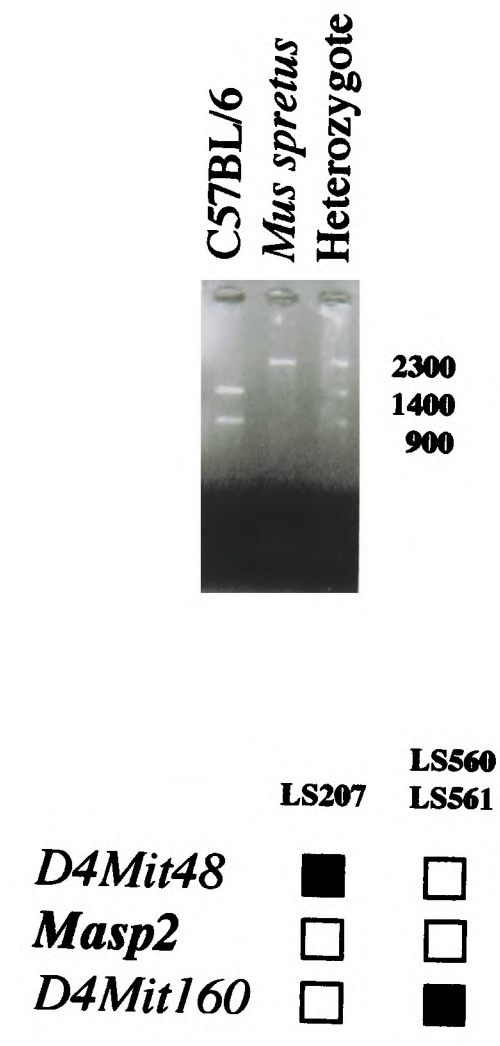


Figure 5.3. Chromosomal localisation of the genes for MASP2. (TOP) A 1 % (w/v) agarose gel of the digested PCR products from the MASP2 gene for the wildtype mice (C57BL/6 and *Mus spretus*) and a heterozygote. Molecular sizes (bp) of the alleles are shown on the right. (MIDDLE) Haplotyping of informative recombinant mice for the gene and flanking markers are shown. EUCIB mice numbers are listed above the columns of haplotype boxes, mice heterozgous for the parental alleles are shown as filled boxes, open boxes represent a homozygous haplotype. (BOTTOM) A graphical representation of the chromosomal location of the gene, shown in centimorgans (cM) to the left, an expansion of the region is shown to the right, including the neighbouring markers.

using the PCR primers MM2-1f and MM2-2r (See appendix 9), 1 mM magnesium chloride, an annealing temperature of 59°C and an extension time of 90 seconds. A single discrete PCR product of the expected size (approximately 625 bp) was obtained, as predicted from the MASP2 cDNA and the EST alignment. This product was subcloned into a 3' deoxythymidine residue overhang vector (pCRII-TOPO; see section 2.8.4.). The sequence from two independent clones showed complete identity to the DNA sequence of *Masp2* generated from ESTs and the later published murine MASP2 cDNA sequence. This confirmed that the PCR product was the murine homologue of human MASP2 and that the genomic region encoding the serine protease domain of murine gene for MASP2 did not contain introns. An intron-less encoded serine protease domain is consistent with the gene structures of *C1s* and *C1r* (Tosi *et al.*, 1989) and for recently published human MASP2 gene (Stover *et al.*, 1999c). This is quite distinct from the six exon arrangement of the serine protease domain of human MASP1 (Endo *et al.*, 1996a), see Figure 5.4.

5.8. Chromosomal Localisation

A random selection of genomic DNA from 50 backcross mice of the EUCIB panel (Breen *et al.*, 1994) were used for mapping. All the animals used were from a C57BL/6 x *Mus spretus* F1 generation backcrossed with either parental species. This gave a chromosomal and approximate sub-chromosomal localisation between two well characterised genetic markers. For fine scale mapping, a panel of mice with informative recombinations between the two well characterised marks were haplotyped. The inheritance of each allele was assessed using the PCR based restriction fragmentation length polymorphism assay (section 5.6). The results were used to determine both the chromosomal location and the fine scale mapping position of the gene, with respect to other known EUCIB markers on the mouse genome.

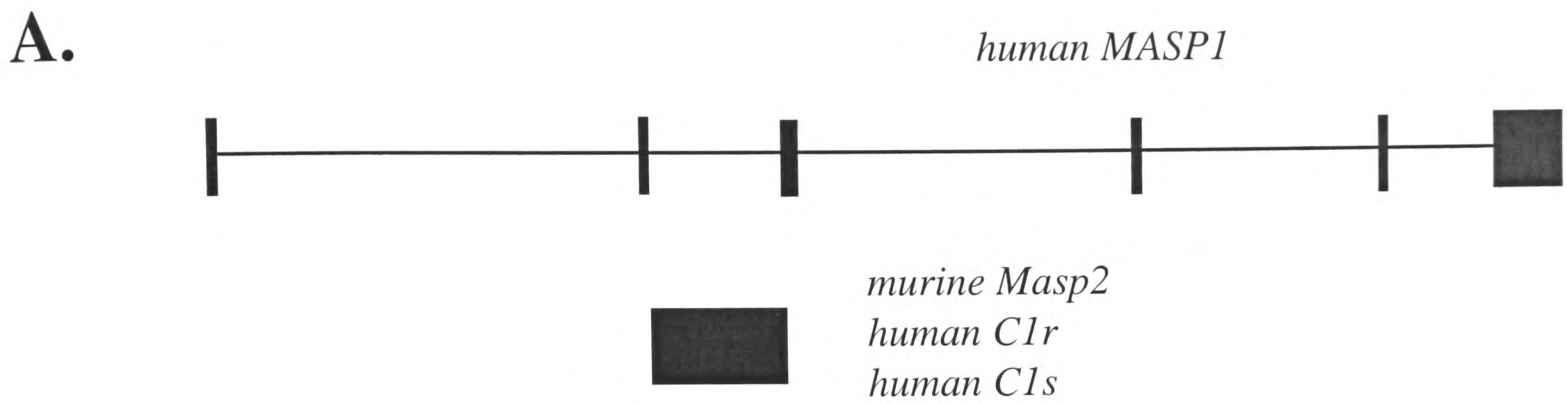


Figure 5.4. The genomic and protein organisation of the serine protease domains of the MASPs, C1r and C1s. (A) The genomic structure encoding the serine protease domain of human *MASP1* is shown above, the single exon that encodes the serine protease domains of human C1r and C1s and murine gene for MASP2. (B) The protein structure of the serine protease domain (boxes) of murine MASP1, above those for MASP2, C1r and C1s. The murine sequences were obtained from cDNA sequence for MASP1 and EST data for MASP2 (accession numbers; AA244853, AA764061, AA530523, AA1266911), C1s (AA895834, AA763800,) and C1r (AA427242, AA175226, AA733400, AA871944). Lines crossing the boxes indictate the intron/exon boundaries predicted from the analogous human genes (*C1r*, *C1s* and *MASP1*) and data from this paper (*Masp2*). The three amino acids essential for serine protease specificity are indicated (His, Asp and Ser), the cysteine bond formation, along with the codon usage for the active site serine residue. The extra disulphide bond in MASP1 creates the histidine loop. The aspartic acid residue three amino acids up from the serine residue, shown circled, which is important for typsin-like substrate specificity. The residue three amino acids C-terminal to the active site serine residue differentiates C1s from C1r and MASP1 from MASP2. Amino acid numbering is based on the murine MASP1 cDNA and MASP2 EST-based sequences.

5.8.1. Chromosomal Localisation of C1r and C1s

From a random panel of 50 backcross mice, *C1r* and *C1s* were shown to map to the telomeric end of chromosome 6 with a LOD scores of 6 and 9.7, respectively. Fine scale mapping of mice with informative recombinants in the region *D6Mit24* to *D6Mit14* revealed that *C1s* and *C1r* co-segregated between the markers *D6Mit24* and *D6Mit337*, at a position approximately 65 cM from the centromere (Figure 5.2). This region of murine chromosome 6, is syntenic with human chromosomal region 12p12-p13.3 which contains the human genes for C1r and C1s (Nguyen *et al.*, 1988). The human genes have been shown to be close physical linkage, arranged in a "tail-to-tail" fashion (Kusumoto *et al.*, 1988).

5.8.2. Chromosomal Localisation of C3aR

The murine gene for the complement C3a anaphylatoxin receptor, *C3aR*, has been localised to region F1 on chromosome 6 (Hollmann *et al.*, 1998). This region is syntenic to an area on the short arm of human chromosome 12, which contains the human genes for C1r and C1s, 12p12-p13.3 (Nguyen *et al.*, 1988). To examine whether these genes formed an unexpected cluster at the genetic level or merely localised to the same chromosomal band, murine *C3aR* was also mapped onto the EUCIB genetic map. Initial mapping showed that *C3aR* localised to a region of chromosome 6, with a LOD score of 7.12, which confirmed *in situ* hybridisation data (Hollmann *et al.*, 1998). The genetic mapping data of this study also suggested a localisation for *C3aR*, which was close to *C1r* and *C1s*. Fine scale mapping showed that *C3aR* co-segregates with *C1s* and *C1r*, to a segment of DNA that spans approximately 0.17 cM, between the markers *D6Mit24* and *D6Mit337* towards the telomeric end of chromosome 6, at approximately 65 cM from the centromere (Figure 5.2).

5.8.3. Chromosomal Localisation of murine *Masp2*

A PCR-RFLP assay was developed which extended across the two exons that encode the first complement control protein (CCP) domain in murine *Masp2* (see Figure 5.1). In the human gene for MASP2 the first CCP domain does not contain an intron (Stover *et al.*, 1999c). The murine MASP2 gene was mapped to chromosome 4 with a LOD score of 6.28. Fine scale mapping with mice recombinant between *D4Mit179* and *D4Nds16*, showed that *Masp2* maps to the 0.11 cM region between *D4Mit48* and *D4Mit160,* at the telomeric end of chromosome 4, 75 cM from the centromere (Figure 5.3). The chromosomal localisation of *Masp2* is distinct from that observed for *Masp1*, where *Masp1* has been shown by FISH to map to the B2-B3 region of mouse chromosome 16 (Takada *et al.*, 1995). This was confirmed by localising the gene onto the EUCIB genetic map using a PCR-RFLP assay that was developed to *Masp1*. Using a random selection of only 50 mice, *Masp1* mapped to the telomeric tip of chromosome 16 (results not shown). Unlike the C1q-associated serine proteases (C1r and C1s) the MBL-associated serine proteases do not form a linkage group and in fact map to different chromosomes in both mice and humans.

5.9. Discussion

5.9.1 Development of a Novel Approach to Genetic Mapping

This chapter describes the use of a novel and rapid approach for localising murine genes onto a backcross genetic map by utilising data available from the murine EST projects. This method has been successfully used to rapidly assign high resolution chromosomal localisations for three complement genes. *C3aR* was localised using the known gene sequence. This method does not rely upon highly specialised technical expertise, reagents or equipment. This approach can be applied in most molecular biology laboratories. More importantly it does not require cloning of the genes or cDNAs. Instead, this technique relies upon the availability of murine EST generated from this large-scale sequencing project.

The large scale murine EST sequencing effort has already amassed over 900,000 sequences (dbEST release date 7/1/2000) and over the next few years it will provide sequence information on the vast majority expressed genes (Marra *et al.*, 1998). Therefore, one of the potential drawbacks of finding the relevant EST information rapidly becomes a diminishing problem. The second difficulty in using this method is the prediction of introns, but with the constant release of improving genome-EST alignment software (Mott, 1997) and the expanding genetic knowledge of protein modules and gene structures from humans and other species, this decision process becomes simpler. A third limitation is uncovering polymorphisms within single exon genes or genes with small introns, but less frequent untranslated and coding region polymorphisms can be used to find polymorphisms (Takahashi and Ko, 1993).

5.9.2. Discussion of the Data

The complement system was identified as an ideal candidate for testing this new method upon (see Table 5.4) because a number complement genes had not been cloned but were identified in the murine EST database (see Table 5.1). Herein, the chromosomal localisation of three complement genes has been assigned using this rapid localisation approach and one has been assigned using the available gene sequence. From this an unexpected clustering of these genes has been revealed.

5.9.2.1. The C1r, C1s and C3aR gene cluster

The murine genes for C1r and C1s map to the telomeric end of chromosome 6, at approximately 65 cM, with LOD scores of 6.0 and 9.7, respectively (Figure 5.2). This region of the mouse genome is syntenic with human chromosome 12 which contains the two linked human genes for C1r and C1s, 12p12-p13.3 (Takada *et al.*, 1993). Mapping of the murine C3aR gene to chromosome 6 band F1 (Hollmann *et al.*, 1998), a broad region that encompasses *C1r* and *C1s*, provoked the idea that *C3aR* may map to this cluster.

Murine Chromosome	Murine Position	Murine Gene Symbol and Name	Human Chromosome
2	ND	*C1nh*, complement component 1 inhibitor	11q11-q13.1
4	66.1	*C1qa*, complement component 1, q subcomponent, a chain	1p36.3-p34.1
4	66.1	*C1qb*, complement component 1, q subcomponent, b chain	1p36.3-p34.1
4	66.1	*C1qc*, complement component 1, q subcomponent, c chain	1p36.3-p34.1
ND		*C1qRp,* C1q/MBL/surfactant protein-A receptor	ND
6	**64.81**	***C1r*, complement component 1, r subcomponent**	**12p13**
6	**64.81**	***C1s*, complement component 1, s subcomponent**	**12p13**
*17	18.86	*C2*, complement component 2	6p21.3
17	34.3	*C3*, complement component 3	19p13.3-p13.2
6	**64.81**	***C3ar1*, complement component 3a receptor 1**	**12p12-p13.3**
*17	18.8	*C4*, complement component 4	6p21.3
†1	67.6	*C4bp*, complement component 4 binding protein	1q32
2	23.5	*Hc*, complement component 5 (hemolytic complement)	9q34.1
ND		*C5ar1*, complement component 5a, receptor 1, *Cd88*	19q13.3-q13.4
¥15	3.0	*C6*, complement component 6	5p13
¥15	3.0	*C7*, complement component 7	5p13
ND		*C8a*, complement component 8, α subunit	1p32
4	49.4	*C8b*, complement component 8, β subunit	1p32
2	12.3	*C8g*, complement component 8, γ subunit	9q34.3
¥ND		*C9*, complement component 9	5p13
2	55	*Cd59*, protectin or membrane inhibitor of reactive lysis	11p13
†1	74.1	*Cfh*, complement component factor H	1q32
3	66.6	*Cfi*, complement component factor I	4q25
14	28.5	*Clu*, clusterin	8p21-p12
†1	106.6	*Cr2*, complement receptor 2, *Cd21*	1q32
†1	106.6	*Crry*, complement receptor related protein	see below
†1	67.6	*Daf1* and *Daf2*, decay accelerating factor 1 and 2, *Cd55*	1q32
10	43.0	*Dfc*, complement factor D or *Adn*, Adipsin	19pTel
*17	18.85	*H2-Bf*, complement component factor B	6p21.3
ND		*Itgam*, αM-integrin, complement receptor 3 α chain (*Cr3a*, *Cd11a*)	16p11.2
10	41.5	*Itgb2,* β_2-integrin, complement receptor 3 β chain (*Cd18*)	21q22.3
16	B2-B3	*Masp1* or *Crarf*, MBL associated serine protease	3q27-q28
4	**75.07**	***Masp2*, MBL associated serine protease 2**	**1p36.3-p36.2**
14	15	*Mbl1*, mannose binding lectin, liver (A)	10q11.2-q21
19	25	*Mbl2*, mannose binding lectin, serum (C)	Absent
†1	106.6	*Mcp,* membrane cofactor protein, *Cd46*	1q32
X	6.2	*Pfc*, properdin	Xp11.4-p11.23
11	45.09	*Vtn*, vitronectin or s-protein	17q11

Table 5.4. A list of known gene loci for murine and human complement components. Genes belonging to the three complement gene clusters are identified by the following superscript symbols; *Major Histocompatibility complex class III complement gene cluster, † Regulators of complement activation (RCA) gene cluster (however, mice have two *Daf* genes which are located outside the RCA gene cluster), ¥Membrane attack complex (MAC) gene cluster. ND indicate gene localisations not determined. Entries shown in bold highlight the murine genes which were mapped by this study. Alternative splicing of the gene *Cr2*, generates two gene products with homology to the human complement receptors; CR1 and CR2, while *Crry* is considered to be the evolutionary homologue of human CR1.

This idea was tested quickly by developing a PCR-RFLP assay to *C3aR*. Mapping showed that *C3aR* localised to the same region, as *C1r* and *C1s* on chromosome 6. Fine scale mapping revealed that all three genes were linked. *C3aR* co-segregates with *C1r* and *C1s* in an approximately 0.17 cM segment of the murine genome.

This is an unexpected complement gene cluster compared with the other complement gene clusters in which all the genes share high sequence similarity with each other (Table 5.4: and section 5.1.6.3). C3aR is very distinct from C1r and C1s at the protein sequence level. Even at the gene level, *C3aR* has a single structural exon while *C1r* and *C1s* have dispersed structural genes covering several exons. Comparison of the recent FISH mapping for the human C3aR gene to 12p13.2-12p13.3 (Hollmann *et al.*, 1998; Paral *et al.*, 1998) and the radiation human-hamster hybrid mapping (Nguyen *et al.*, 1988) of *C1s* and *C1r* implies that a similar clustering of genes may exist for humans.

5.9.2.2. The Localisation of C3aR

The murine gene for C3aR is the only seven transmembrane receptor gene so far found on this region of chromosome 6. In humans, the other complement anaphylatoxin receptor gene (C5aR) maps to the cluster of seven transmembrane spanning receptors of the chemotactic family (Murphy *et al.*, 1992; Bao *et al.*, 1992) on chromosome 19p13.3-13.4 (Gerard *et al.*, 1993). The murine gene for C5aR was not mapped due to both being unsuccessful in amplifying two independent segments of the gene and the likelihood of accurately predicting the location of the gene. From the mapping of this cluster of chemotaxic receptors in mice (Stubbs *et al.*, 1996) and by examining syntenic chromosomal regions, murine *C5aR* would be expected to map to chromosome 17, at approximately 10 cM from the centromere. So, despite the homology of these complement anaphylatoxin receptors at both the functional and protein family level, these two complement anaphylatoxin genes probably do not form a gene cluster.

5.9.2.3 Genetic Mapping of *Masp2*

Genetic mapping of *Masp2* revealed that it maps to chromosome 4 (Figure 5.3, top) with a LOD score of 6.28. Fine scale mapping of informative recombinants (Figure 5.3, middle and bottom) indicated that the murine gene lies in an approximately 0.11 cM region of DNA towards the telomeric end of chromosome 4, between *D4Mit48* and *D4Mit160*. This region on chromosome 4 is syntenic with the short arm of human chromosome 1, where human *MASP2* localises (Stover *et al.*, 1999a).

The gene for murine MASP2 clearly localises to a different position to its partner MBL-associated serine protease, *Masp1*, which localises to the telomeric region of murine chromosome 16 by genetic mapping described here and shown by FISH (Takada *et al.*, 1995). The lack of genetic linkage between the two murine gene MBL-associated serine proteases genes is in agreement with the localisation of the human MASP1 and MASP2 genes on different chromosomes (Takada *et al.*, 1995; Stover *et al.*, 1999a). This is in marked contrast to the C1r and C1s genes in humans (Kusumoto *et al.*, 1988) and mice (shown here) that are closely linked.

5.9.2.4. The Non-Homologous Functions Between the MASPs and C1r/s

The chromosomal positions of the MASP genes discounts the idea that *C1r* and *C1s* arose through a simple duplication of the possibly more ancient MASP genes. This is highlighted at the functional level where human MASP1 and MASP2 do not appear to be functional analogues of C1r and C1s, despite the elegant symmetry seen between the MBL and classical pathways (see (Reid and Law, 1995; Thiel *et al.*, 1997)). After auto-activation of C1r, via C1q binding to clustered antibodies, C1r activates C1s. C1s proceeds to cleave the first two components of the classical pathway, C4 followed by C2. The resulting complex, composed of activated C4 and C2, is the classical pathway C3 convertase. In contrast, despite the MBL activation pathway

impinging on the same key components and forming an identical classical pathway C3 convertase, the functions of the MASPs can not be superimposed onto those of C1r and C1s. The activation process involving the MASPs appears far more complex. MASP2 has been reported to have a similar substrate specificity to that of C1s, cleaving C4 (Thiel *et al.*, 1997), but may also possess an auto-activation capacity (Vorup Jensen *et al.*, 1998). Whereas MASP1 has been shown to cleave C2 (Matsushita and Fujita, 1992) and perhaps C3, thus circumventing the action of the classical pathway C3 convertase (Matsushita and Fujita, 1995; Ogata *et al.*, 1995; Matsushita *et al.*, 1998). The cleavage of C3 by MASP1 remains controversial, as it can not be confirmed by others (Wong *et al.*, 1998).

5.9.2.5 The Evolutionary Theory from the Protein and cDNA Level

It is clear from phylogenetic analysis that MASP1 and MASP2/C1r/C1s form divergent arms of serine protease family of the complement system (Ji *et al.*, 1997; Endo *et al.*, 1998; Matsushita *et al.*, 1998). This is echoed at many different levels (see Figure 5.4). The functional elements of the serine protease domain, which supports the evolutionary theory of Endo and colleagues (Endo *et al.*, 1998) are important for clarifying that the ESTs identified are the murine counterparts of the human sequences. MASP1 is a TCN serine protease, since the active site serine is encoded by a TCN codon, which is TCT in mice, while MASP2/C1r/C1s are AGY types, with the active site serine being encoded respectively by AGT, AGT and AGC in mice. The histidine loop, typical of MASP1 is present in murine MASP1 and absent in murine MASP2, C1r and C1s. The residue, three amino acids C-terminal to the active site serine, is a proline in murine MASP1, which is generally conserved in chymotrypsinogen-like proteases but replaced with a valine or an alanine in MASP2/C1r/C1s (see Figure 5.4B). These features are confirmed in the later published murine cDNA sequence for MASP2 (Stover *et al.*, 1999b). The idea of the two lineages of MASP/C1r/C1s (Endo *et al.*, 1998) is also reflected at the gene level.

5.9.2.6. The Evolutionary Theory from the Genetic Viewpoint

The intron-exon organisation of *Masp1* displays a type of split exon arrangement for the serine protease domain (Endo *et al.*, 1996a), which is not seen in other serine protease genes (Rogers, 1985; Irwin *et al.*, 1988). However, the serine protease domains of *Masp2, C1r* and *C1s* are encoded by single exons, like the non-functional serine protease domain of haptoglobin (Maeda *et al.*, 1984). The exon arrangement of the serine protease domains taken together with the chromosomal localisations, allows speculation as to the evolution of these proteins (Figure 5.5).

The parsimonious answer, based on this evidence, would suggest that the ancestral MASP/C1s/C1r gene was composed of a split exon serine protease domain, with an active site serine encoded by a TCN codon (see Figure 5.5). This is confirmed by the tunicate genes which have been shown to have a split exon encoded serine protease domain of the TCN type (Ji *et al.*, 1997; Matsushita *et al.*, 1998; Endo *et al.*, 1998). The model proposed here can explain how this ancestral gene with a split exon encoded serine protease domain could lead to the current gene and exon arrangements of four complement activating serine protease genes.

Duplication of the ancestral serine protease domain gene occurs after the divergence of cyclostomes (jawless fish, for example lamprey) from urochordata (or tunicata, for example sea squirts), probably by a RNA-mediated event as suggested by Endo and colleagues (Endo *et al.*, 1998). This generates a second gene with a single exon encoded serine protease domain with a mutation in the active site serine codon from TCN to AGY, which is speculated to be through a retrotransposition of a partially processed mRNA (Endo *et al.*, 1998). This type of mutation in the serine protease domain from a TCN to AGY type has occurred at least twice in evolutionary history (Irwin, 1988) and two different mechanisms have been proposed (Irwin, 1988; Brenner, 1988). The idea of RNA-mediated gene duplication or intron loss occurring during gene family evolution is well documented (Weiner *et al.*, 1986; Rogers, 1985).

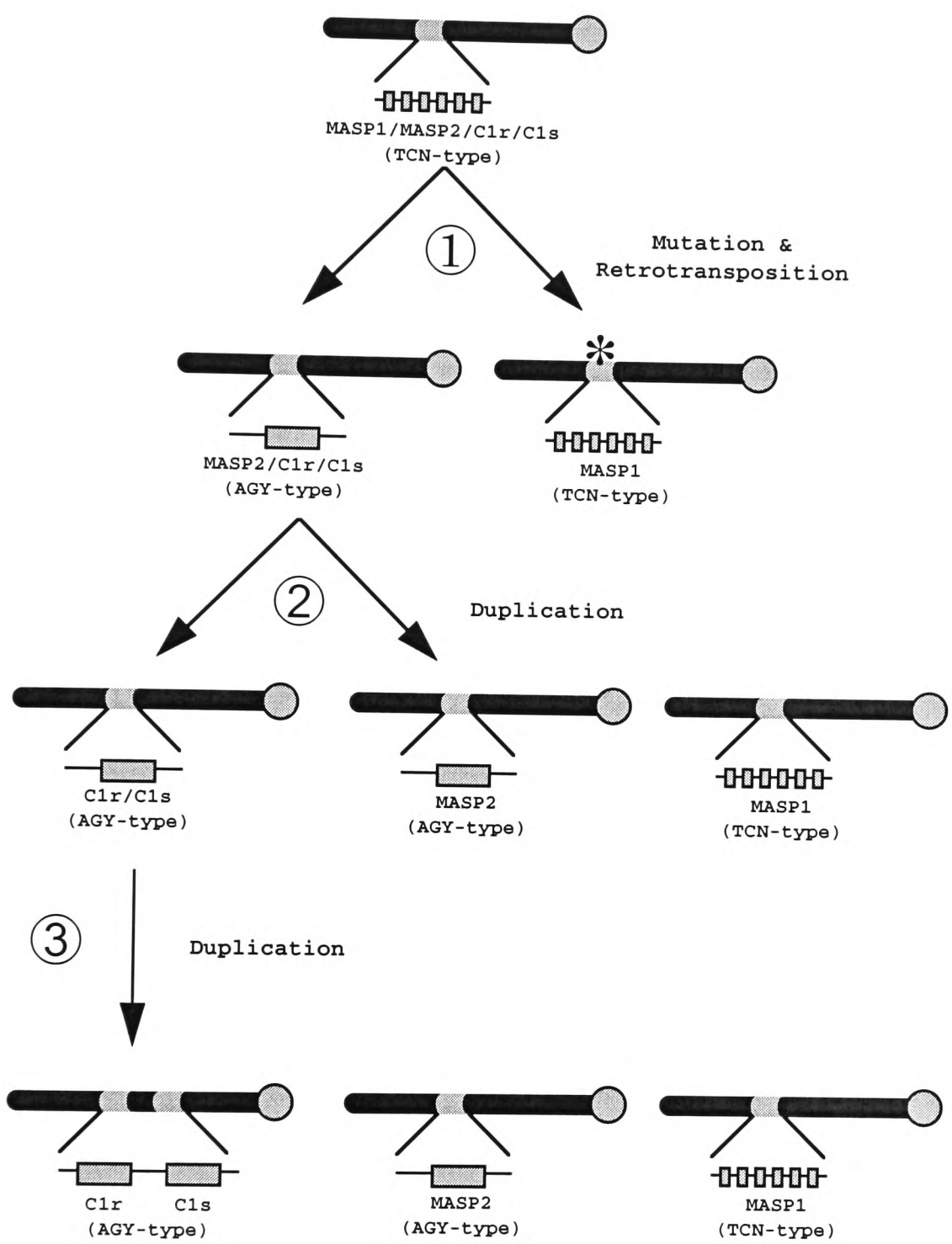


Figure 5.5. A graphical representation of the molecular theory of evolution for MASP1, MASP2, C1r and C1s. The exon-intron arrangement of only the serine protease encoding region (boxes and lines) of the genes is shown expanded below the chromosome (lollipops). The asterisk represents a putative gene silencing mutation in the TCN-type MASP1, as proposed in an evolutionary model (Endo *et al.*, 1998).

Therefore, this would leave an ancestral MASP2/C1r/C1s-like gene and an ancestral MASP1 gene, residing on two different chromosomes (Figure 5.5). Incidentally, the two MASP genes of tunicates are TCN type serine proteases, which probably arose from an independent gene duplication event occurring within their own lineage. In addition, the apparent absence of protein for the TCN-type serine protease (Masp1) in carp (bony fish), sharks (cartilagenous fish) and lamprey (jawless fish) (Matsushita *et al.*, 1998; Endo *et al.*, 1998) is believed to be due to a gene silencing effect at the retrotransposition step (Endo *et al.*, 1998) (see Figure 5.5, asterisk).

The second step of the model is a duplication of the ancestral MASP2/C1r/C1s gene, which would generate two single exon encoded serine protease genes, MASP2 and the ancestral C1r/C1s, residing on two different genomic regions. This event probably occurred prior to the divergence of cartilaginous fish, as demonstrated by the presence of a C1 like complex in the nurse shark (Smith, 1999).

The final evolutionary event occurring prior to the divergence of amphibians, involving the duplication of the C1r/C1s-like gene, would lead to the generation of the tandomly arranged, closely linked, C1 associated serine proteases genes; *C1*r and *C1s*. A similar model of gene family evolution and chromosomal dispersal, through RNA-mediated transposition and gene duplications has been proposed for other systems, such as the neuropeptide Y receptors (Darby *et al.*, 1997).

This model, built from chromosomal localisation data gene structures and phylogenetic studies of others, helps explain the different chromosomal and exon arrangements of the two lineages of MBL/C1q associated serine proteases.

5.10. Summary

This work describes the development of a novel approach for mapping genes onto a panel of backcross mice. The technique circumvents the laborious cDNA cloning and genomic cloning steps typically employed in other techniques. This novel and rapid PCR-based RFLP assay approach relies upon murine EST data, the prediction of exon-intron boundaries and unearthing restriction length polymorphism within these introns.

This method can be easily applied to the genes of other systems, ranging from the interests of the individual researcher to large-scale gene localisation projects.

This approach was tested upon the murine complement system of genes that encodes a cascade of serum proteins and receptors which forms a vital arm of innate immunity and enhances the adaptive immune response. Using this approach the chromosomal localisations of three genes of the murine complement system have been established and *C3aR* was localised using the known gene structure. It was shown that the C3a receptor, C1r and C1s genes form an unexpected complement gene cluster towards the telomeric end of chromosome 6. The second mannan binding lectin associated serine protease gene was mapped to the telomeric end of chromosome 4, which is distinct from other complement activating serine proteases. These results provide new insights into the evolution of this group of proteins.

CHAPTER 6: SUMMARY

The murine gene encoding surfactant protein D (SP-D) was cloned in order to generate a transgenic mouse that lacked SP-D. It was hoped that a mouse deficient in SP-D would determine whether or not this protein was engaged in immune defence or participated in maintenance of surfactant function. In addition, if SP-D did function as an immune molecule, the knock-out mouse could identify the specific types of pathogens it was forming an immunological barrier against. During the work for this thesis, it was discovered that two independent groups had generated SP-D transgenic mice, so the course of this project was subsequently changed.

The work in chapter 3 describes the isolation and characterisation of the mouse SP-D gene (*Sftpd*), which spans 8 exons over 14 kb of sequence and shows an overall organisation that is strikingly similar to the other collectin genes. The 5' untranslated region of the mRNA was cloned to identify the start of transcription and obtain the full length sequence of murine SP-D.

Among the human collectins, the regulation of SP-D expression is least understood. By selecting the murine SP-D clone containing the longest 5' region, a detailed analysis of the promoter region could be performed. Analysis of 3.5 kb of 5' flanking nucleotide sequence for *Sftpd* is described. When the sequence was compared to that of human SP-D gene (SFTPD) and the bovine conglutinin gene (CGN1) it revealed positional conservation of a number of transcription factor binding sites. Comparison of the SP-D with bovine conglutinin was performed since conglutinin is believed to have arisen from SP-D by gene duplication and conservation of sequences involved in transcription control may still exist. Although functional transcription analysis was not performed upon the murine SP-D gene, the detailed theoretical analysis will form the basis of future experiments by others.

Since two independent groups were generating SP-D deficient mice, the obvious route was to generate mice that lacked SP-A and SP-D, or triple collectin knockouts. The simplest route was to cross-breed single collectin null animals. Since the genes for SP-A, SP-D and one of the two MBL genes (MBL-A) were known to be clustered on murine chromosome 14, genetic mapping was used to determine if crossbreeding was feasible route. The three loci were genetically mapped using the European collaborative intraspecific backcross of 1000 mice. This panel would not only provide the highest genetic map of the gene cluster, it would also indicate the approximate number of transgenic mice required to obtain a recombination between two different collectin genes for the generation of double null transgenic animals. Unfortunately, the three loci co-segregate across 1000 mice, making the idea of cross-breeding an unlikely route to generate double or triple collectin gene null animals.

The alteration in surfactant homeostasis and cellular morphology of mice lacking SP-D, makes the use of such animals, or indeed double surfactant collectin knockouts, unfavourable when examining both immune defence clearance and defence of pathogens. A more constructive approach at present would be to investigate if the phenotype seen in null SP-D animals is a direct cause of SP-D deficiency or an artefact from the construct used for generating the transgenic animal. One group has reported the presence of a non-sense mRNA being transcribed from the null SP-D gene. Since both groups used a similar design for the null SP-D gene construct other routes have to be explored to confirm the observed phenotype is due to the lack of SP-D and not to a deleterious effect of the null allele. One approach to avoid this would be to rescue the null gene phenotype with the SP-D gene at a different locus. This could be achieved by crossing the SP-D knockout mouse with a SP-D knockin mouse; to generate a mouse that contains the wildtype SP-D gene at a different locus. Alternatively, a SP-D knockin mouse could be made with the SP-D gene driven by the SP-A or SP-C promoter to achieve a higher level of expression of SP-D. A high expressing transgenic

mouse could be used in a converse but similar manner to a SP-D knockout mouse, to address if SP-D imparts immune protection or if surfactant maintenance improved.

The adaptive immune system is built upon the innate immune system both in the evolutionary and immunological sense. The adaptive immune system is composed of a vast array of antigen receptors that are formed by gene rearrangements. In contrast, the innate immune system which is composed of a limited set of pattern recognition molecules protects the host while the adaptive immune response is mounted in higher organisms. The examination of the immune system of other species and simpler organisms has been a successful method of investigation into the roles of innate immune components. Within the collectins, MBL has been subjected to the greatest phylogenetic analysis. Only limited phylogenetic analysis has been conducted for the surfactant collectins. The work in chapter 3 describes the use a cDNA probe derived from the CRD encoding region of murine SP-D that has been used to determine the presence of SP-D like gene in different species. The probe hybridised with single DNA fragments from mammals, birds and amphibians but not fish. The key point is the presence of only a single hybridising DNA fragment.

In an attempt to glean more information about potential SP-D like genes, low stringency PCR assays were performed on genomic DNA from a range of species. A fragment of a SP-D-like gene, identified in sheep and most likely to be CL-43. This represents the second species, other than cattle, to have the CL-43 gene. This dates the gene duplication of SP-D to generate CL-43 prior to the divergence of sheep from cattle within bovidae. This fragment will be used to clone the full-length sequence of sheep CL-43. Alternatively, the consensus sequence of CL-43 will assist the cloning of the gene from other members of the bovidea (goats, bison) or other closely related species (camels or horses). This would generate a detailed consensus sequence for CL-43, providing a structural insight as to why two closely related collectins, SP-D and CL-43, have quite distinct carbohydrate specificity. The small genetic distance between sheep

and cattle is reflected by 11 substitutions across 110 amino acids. By contrast, the cloning of porcine SP-D, which was undertaken because of the medical relevance of porcine surfactant, revealed startling sequence changes.

The cDNA for porcine SP-D was cloned for two reasons. First, to investigate the components of porcine surfactant as both porcine surfactant and organs have medically important applications in the therapy of respiratory disease and use in xenotransplantation. Second, to gain a greater understanding of SP-D from the analysis of a new sequence from a different species (as mentioned above). This work was conducted in collaboration with a Dutch research group, who are establishing a porcine animal model of human lung infectious diseases.

The full-length cDNA sequence of porcine SP-D including the 5' and 3' untranslated regions, was obtained from two overlapping bacteriophage clones and by PCR cloning. From the analysis of the predicted amino acid sequence, three unique features were revealed from the porcine sequence in comparison to SP-D from other species. The collagen region contained an extra cysteine residue, which may have important structural consequences in the formation of higher order oligomers by disulphide bonds. The two other differences lie within a loop region in the carbohydrate recognition domain, close to the carbohydrate binding region that may have functional implications; a potential N-linked glycosylation site and an insertion of three amino acids. These variations were shown not to be polymorphisms or mutations as they were present in four different breeds of pig.

The known sequence of porcine SP-A also showed a potential insertion in the CRD. However, by PCR based cloning the CRD from the genome of four different pig breeds the extra amino acid and lack of conservation over a 12 amino acid stretch was revealed as a sequencing error. Because of the sequence variations in porcine SP-D in comparison to SP-D from other species, both the genes for SP-D and SP-A were

chromosomally mapped to determine whether the gene localisation varied. The porcine genes for SP-D and SP-A were mapped using a porcine-hamster somatic cell hybrid mapping panel. Both genes co-localised to porcine chromosome 14, which is syntenic with the regions on human chromosome 10 that contains the surfactant collectin loci. However, the limited resolution of the porcine cell hybrid mapping panel precluded obtaining a sub-chromosomal localisation.

There seemed a deficit in the current mapping approaches. Somatic cell hybrid panels of high resolution are only available for mapping the human genome. In mice, the high resolving power of genetic mapping has made the development of high resolution somatic cell hybrid panels unnecessary. However, somatic cell hybrid panels are extraordinarily useful for mapping short segments of cDNA: most notably expressed sequence tags (EST). Despite the abundance of murine EST information, there was no high resolution method for chromosomally mapping these transcribed gene fragments. A total of nine murine complement genes have not been mapped, yet EST information existed for seven of these genes. However, once the gene structure is known it is very easy to genetically map murine genes as I have shown for the murine collectin gene, using PCR based restriction fragment length polymorphisms within genes. Indeed the gene structure does not need to be known; the predicted exon-intron boundaries is simple to estimate once the structure of a homologous gene is known, for example work in chapter 4 demonstrates that an intron-exon boundary within porcine SP-D could be readily predicted. In light of this, I realised that it was possible to map ESTs onto a high resolution genetic map, by predicting exon-intron-exon boundaries in ESTs through alignment and using homologous gene structures. This meant the two exons could be amplified, including the intervening intron. The intronic sequence could then be used as a potential source of polymorphic restriction sites between the two inbred mouse strains used for genetic mapping.

This method of mapping is both novel and rapid. It exploits the murine expressed sequence tag database and maps genes using PCR based restriction fragment length polymorphism assays. This technique circumvents the laborious cDNA or genomic cloning steps of other mapping methods by relying on EST data and the prediction of exon-intron boundaries. This intuitive method for genetically mapping ESTs was used to map three complement genes. The gene for C3aR was mapped using a simplified version of this method, which relied upon the known gene sequence. From this, it was revealed that the murine C3a receptor, C1r and C1s genes formed an unexpected complement gene cluster towards the telomeric end of chromosome 6. In addition, the second mannose binding lectin associated serine protease (MASP2) gene was mapped to the telomeric end of chromosome 4, which is distinct from other complement activating serine proteases. The clustering of the C1q and MASP associated serine protease genes form new insights into the evolution of this group of proteins. This method can be easily applied to the genes of any system, ranging from the interests of the individual researcher to large scale gene localisation projects.

REFERENCES

Acton, S., Resnick, D., Freeman, M., Ekkel, Y., Ashkenas, J., and Krieger, M. (1993). The collagenous domains of macrophage scavenger receptors and complement component C1q mediate their similar, but not identical, binding specificities for polyanionic ligands. J Biol Chem *268*, 3530-7.

Akinbi, H. T., Breslin, J. S., Ikegami, M., Iwamoto, H. S., Clark, J. C., Whitsett, J. A., Jobe, A. H., and Weaver, T. E. (1997). Rescue of SP-B knockout mice with a truncated SP-B proprotein. Function of the C-terminal propeptide. J Biol Chem *272*, 9640-7.

Akiyama, J., Volik, S. V., Plajzer Frick, I., Prince, A., Sago, H., Weier, H. U., Vanderbilt, J. N., Hawgood, S., and Poulain, F. R. (1999). Characterization of the mouse collectin gene locus. Am J Respir Cell Mol Biol *21*, 193-9.

Altschul, S. F., Boguski, M. S., Gish, W., and Wootton, J. C. (1994). Issues in searching molecular sequence databases. Nat Genet *6*, 119-29.

Andersson, M., Hanson, A., Englund, G., and Dahlback, B. (1991). Inhibition of complement components C3 and C4 by cadralazine and its active metabolite. Eur J Clin Pharmacol *40*, 261-5.

Arbibe, L., Koumanov, K., Vial, D., Rougeot, C., Faure, G., Havet, N., Longacre, S., Vargaftig, B. B., Bereziat, G., Voelker, D. R., Wolf, C., and Touqui, L. (1998). Generation of lyso-phospholipids from surfactant in acute lung injury is mediated by type-II phospholipase A2 and inhibited by a direct surfactant protein A-phospholipase A2 protein interaction. J Clin Invest *102*, 1152-60.

Arlaud, G. J., Thielens, N. M., and Illy, C. (1990). Arrangement of the C1 complex of complement. Biochem Soc Trans *18*, 1148-51.

Ausubel, F. A., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1998). Current Protocols in Molecular Biology (New York: John Wiley & Sons).

Baker, C. S., Evans, T. W., Randle, B. J., and Haslam, P. L. (1999). Damage to surfactant-specific protein in acute respiratory distress syndrome. Lancet *353*, 1232-7.

Bao, L., Gerard, N. P., Eddy, R. L., Jr., Shows, T. B., and Gerard, C. (1992). Mapping of genes for the human C5a receptor (C5AR), human FMLP receptor (FPR), and two FMLP receptor homologue orphan receptors (FPRH1, FPRH2) to chromosome 19. Genomics *13*, 437-40.

Benne, C. A., Benaissa Trouw, B., van Strijp, J. A., Kraaijeveld, C. A., and van Iwaarden, J. F. (1997). Surfactant protein A, but not surfactant protein D, is an opsonin for influenza A virus phagocytosis by rat alveolar macrophages. Eur J Immunol *27*, 886-90.

Benne, C. A., Kraaijeveld, C. A., van Strijp, J. A., Brouwer, E., Harmsen, M., Verhoef, J., van Golde, L. M., and van Iwaarden, J. F. (1995). Interactions of surfactant protein A with influenza A viruses: binding and neutralization. J Infect Dis *171*, 335-41.

Benson, D. A., Boguski, M. S., Lipman, D. J., and Ostell, J. (1997). GenBank. Nucleic Acids Res *25*, 1-6.

Bezouska, K., Crichlow, G. V., Rose, J. M., Taylor, M. E., and Drickamer, K. (1991). Evolutionary conservation of intron position in a subfamily of genes encoding carbohydrate-recognition domains. J Biol Chem *266*, 11604-9.

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res *7*, 1513-23.

Blau, H., Riklis, S., Van Iwaarden, J. F., McCormack, F. X., and Kalina, M. (1997). Nitric oxide production by rat alveolar macrophages can be modulated in vitro by surfactant protein A. Am J Physiol *272*, L1198-204.

Bohinski, R. J., Di Lauro, R., and Whitsett, J. A. (1994). The lung-specific surfactant protein B gene promoter is a target for thyroid transcription factor 1 and hepatocyte nuclear factor 3, indicating common factors for organ-specific gene expression along the foregut axis. Mol Cell Biol *14*, 5671-81.

Borron, P., McCormack, F. X., Elhalwagi, B. M., Chroneos, Z. C., Lewis, J. F., Zhu, S., Wright, J. R., Shepherd, V. L., Possmayer, F., Inchley, K., and Fraher, L. J. (1998a). Surfactant protein A inhibits T cell proliferation via its collagen-like tail and a 210-kDa receptor. Am J Physiol *275*, L679-86.

Borron, P., Veldhuizen, R. A., Lewis, J. F., Possmayer, F., Caveney, A., Inchley, K., McFadden, R. G., and Fraher, L. J. (1996). Surfactant associated protein-A inhibits human lymphocyte proliferation and IL-2 production. Am J Respir Cell Mol Biol *15*, 115-21.

Borron, P. J., Crouch, E. C., Lewis, J. F., Wright, J. R., Possmayer, F., and Fraher, L. J. (1998b). Recombinant rat surfactant-associated protein D inhibits human T lymphocyte proliferation and IL-2 production. J Immunol *161*, 4599-603.

Botas, C., Poulain, F., Akiyama, J., Brown, C., Allen, L., Goerke, J., Clements, J., Carlson, E., Gillespie, A. M., Epstein, C., and Hawgood, S. (1998). Altered surfactant homeostasis and alveolar type II cell morphology in mice lacking surfactant protein D. Proc Natl Acad Sci U S A *95*, 11869-74.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem *50*, 349-83.

Breen, M., Deakin, L., Macdonald, B., Miller, S., Sibson, R., Tarttelin, E., Avner, P., Bourgade, F., Guenet, J. L., Montagutelli, X., Poirier, C., Simon, D., Tailor, D., Bishop, M., Kelly, M., Rysavy, F., Rastan, S., Norris, D., Shepherd, D., Abbott, C., Pilz, A., Hodge, S., Jackson, I., Boyd, Y., Blair, H., Maslen, G., Todd, J. A., Reed, P. W., Stoye, J., Ashworth, A., McCarthy, L., Cox, R., Schalkwyk, L., Lehrach, H., Klose, J., Gangadharan, U., and Brown, S. (1994). Towards high resolution maps of the mouse and human genomes--a facility for ordering markers to 0.1 cM resolution. European Backcross Collaborative Group. Hum Mol Genet *3*, 621-627.

Brenner, S. (1988). The molecular evolution of genes and proteins: a tale of two serines. Nature *334*, 528-30.

Brown Augsburger, P., Chang, D., Rust, K., and Crouch, E. C. (1996a). Biosynthesis of surfactant protein D. Contributions of conserved NH2-terminal cysteine residues and collagen helix formation to assembly and secretion. J Biol Chem *271*, 18912-9.

Brown Augsburger, P., Hartshorn, K., Chang, D., Rust, K., Fliszar, C., Welgus, H. G., and Crouch, E. C. (1996b). Site-directed mutagenesis of Cys-15 and Cys-20 of

pulmonary surfactant protein D. Expression of a trimeric protein with altered anti-viral properties. J Biol Chem *271*, 13724-30.

Bruno, M. D., Bohinski, R. J., Huelsman, K. M., Whitsett, J. A., and Korfhagen, T. R. (1995). Lung cell-specific expression of the murine surfactant protein A (SP-A) gene is mediated by interactions between the SP-A promoter and thyroid transcription factor-1 [published erratum appears in J Biol Chem 1995 Jul 7;270(27):16482]. J Biol Chem *270*, 6531-6.

Bruns, G., Stroh, H., Veldman, G. M., Latt, S. A., and Floros, J. (1987). The 35 kd pulmonary surfactant-associated protein is encoded on chromosome 10. Hum Genet *76*, 58-62.

Burnette, W. N. (1981). "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal Biochem *112*, 195-203.

Caddle, M. S., Lussier, R. H., and Heintz, N. H. (1990). Intramolecular DNA triplexes, bent DNA and DNA unwinding elements in the initiation region of an amplified dihydrofolate reductase replicon [published erratum appears in J Mol Biol 1990 Apr 20;212(4):865]. J Mol Biol *211*, 19-33.

Cai, G. Z., Griffin, G. L., Senior, R. M., Longmore, W. J., and Moxley, M. A. (1999). Recombinant SP-D carbohydrate recognition domain is a chemoattractant for human neutrophils. Am J Physiol *276*, L131-6.

Carroll, M. C. (1998). The role of complement and complement receptors in induction and regulation of immunity. Annu Rev Immunol *16*, 545-68.

Carroll, M. C., Campbell, R. D., Bentley, D. R., and Porter, R. R. (1984). A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. Nature *307*, 237-41.

Carroll, M. C., and Fischer, M. B. (1997). Complement and the immune response. Curr Opin Immunol *9*, 64-9.

Chailley-Heu B, Rubio S, Rougier JP, Ducroc R, Barlier-Mur AM, Ronco P, and JR, B. (1997). Expression of hydrophilic surfactant proteins by mesentery cells in rat and man. Biochem J *328*, 251-256.

Chen, A., Gaddipati, S., Hong, Y., Volkman, D. J., Peerschke, E. I., and Ghebrehiwet, B. (1994). Human T cells express specific binding sites for C1q. Role in T cell activation and proliferation. J Immunol *153*, 1430-40.

Cheng, G., Ueda, T., Nakajima, H., Nakajima, A., Kinjyo, S., Motojima, S., and Fukuda, T. (1998). Suppressive effects of SP-A on ionomycin-induced IL-8 production and release by eosinophils. Int Arch Allergy Immunol *117 Suppl 1*, 59-62.

Chevalet, C., Gouzy, J., and SanCristobal Gaudy, M. (1997). Regional assignment of genetic markers using a somatic cell hybrid panel: a WWW interactive program available for the pig genome. Comput Appl Biosci *13*, 69-73.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry *18*, 5294-9.

Christensen, G., and Mousing, J. (1992). Respiratory System. In Diseases of swine, A. D. Leman, B. E. Straw, W. L. Mengeling, S. D'Allaire and D. J. Taylor, eds. (London: Wolfe Publishing), pp. 138-62.

Clark, J. C., Weaver, T. E., Iwamoto, H. S., Ikegami, M., Jobe, A. H., Hull, W. M., and Whitsett, J. A. (1997). Decreased lung compliance and air trapping in heterozygous SP-B-deficient mice. Am J Respir Cell Mol Biol *16*, 46-52.

Clark, J. C., Wert, S. E., Bachurski, C. J., Stahlman, M. T., Stripp, B. R., Weaver, T. E., and Whitsett, J. A. (1995). Targeted disruption of the surfactant protein B gene disrupts surfactant homeostasis, causing respiratory failure in newborn mice. Proc Natl Acad Sci U S A *92*, 7794-8.

Cohen, S. N., Chang, A. C., and Hsu, L. (1972). Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA. Proc Natl Acad Sci U S A *69*, 2110-4.

Coto, E., Martinez Naves, E., Dominguez, O., DiScipio, R. G., Urra, J. M., and Lopez Larrea, C. (1991). DNA polymorphisms and linkage relationship of the human complement component C6, C7, and C9 genes. Immunogenetics *33*, 184-7.

Creuwels, L. A., Demel, R. A., van Golde, L. M., Benson, B. J., and Haagsman, H. P. (1993). Effect of acylation on structure and function of surfactant protein C at the air-liquid interface. J Biol Chem *268*, 26752-8.

Creuwels, L. A., van Golde, L. M., and Haagsman, H. P. (1997). The pulmonary surfactant system: biochemical and clinical aspects. Lung *175*, 1-39.

Crouch, E., Persson, A., and Chang, D. (1993a). Accumulation of surfactant protein D in human pulmonary alveolar proteinosis. Am J Pathol *142*, 241-8.

Crouch, E., Persson, A., Chang, D., and Heuser, J. (1994). Molecular structure of pulmonary surfactant protein D (SP-D). J Biol Chem *269*, 17311-9.

Crouch, E., Rust, K., Marienchek, W., Parghi, D., Chang, D., and Persson, A. (1991). Developmental expression of pulmonary surfactant protein D (SP-D). Am J Respir Cell Mol Biol *5*, 13-8.

Crouch, E., Rust, K., Veile, R., Donis Keller, H., and Grosso, L. (1993b). Genomic organization of human surfactant protein D (SP-D). SP-D is encoded on chromosome 10q22.2-23.1. J Biol Chem *268*, 2976-83.

Crouch, E. C. (1998). Collectins and pulmonary host defense. Am J Respir Cell Mol Biol *19*, 177-201.

Crouch, E. C., Persson, A., Griffin, G. L., Chang, D., and Senior, R. M. (1995). Interactions of pulmonary surfactant protein D (SP-D) with human blood leukocytes. Am J Respir Cell Mol Biol *12*, 410-5.

Darby, K., Eyre, H. J., Lapsys, N., Copeland, N. G., Gilbert, D. J., Couzens, M., Antonova, O., Sutherland, G. R., Jenkins, N. A., and Herzog, H. (1997). Assignment of the Y4 receptor gene (PPYR1) to human chromosome 10q11.2 and mouse chromosome 14. Genomics *46*, 513-5.

Dico, A. S., Taneva, S., Morrow, M. R., and Keough, K. M. (1997). Effect of calcium on phospholipid interaction with pulmonary surfactant protein C. Biophys J *73*, 2595-602.

Dobbie, J. W. (1996). Surfactant protein A and lamellar bodies: a homologous secretory function of peritoneum, synovium, and lung. Perit Dial Int *16*, 574-81.

Dobbie, J. W., Tasiaux, N., Meijers, P., Anderson, J. D., Bodart, C., Hind, C., Bourguet, C., and Perret, J. (1994). Lamellar bodies in synoviocytes, mesothelium and specific epithelia as possible site of auto-antigen in rheumatoid disease [see comments]. Br J Rheumatol *33*, 508-19.

Dodds, A. W., and Petry, F. (1993). The phylogeny and evolution of the first component of complement, C1. Behring Inst Mitt *93*, 87-102.

Downing, J. F., Pasula, R., Wright, J. R., Twigg, H. L. d., and Martin, W. J. d. (1995). Surfactant protein a promotes attachment of Mycobacterium tuberculosis to alveolar macrophages during infection with human immunodeficiency virus. Proc Natl Acad Sci U S A *92*, 4848-52.

Drickamer, K. (1992). Engineering galactose-binding activity into a C-type mannose-binding protein. Nature *360*, 183-6.

Drickamer, K., and McCreary, V. (1987). Exon structure of a mannose-binding protein gene reflects its evolutionary relationship to the asialoglycoprotein receptor and nonfibrillar collagens. J Biol Chem *262*, 2582-9.

Eda, S., Suzuki, Y., Kawai, T., Ohtani, K., Kase, T., Fujinaga, Y., Sakamoto, T., Kurimura, T., and Wakamiya, N. (1997). Structure of a truncated human surfactant protein D is less effective in agglutinating bacteria than the native structure and fails to inhibit haemagglutination by influenza A virus. Biochem J *323*, 393-9.

Eliakim, R., DeSchryver Kecskemeti, K., Nogee, L., Stenson, W. F., and Alpers, D. H. (1989). Isolation and characterization of a small intestinal surfactant-like particle containing alkaline phosphatase and other digestive enzymes. J Biol Chem *264*, 20614-9.

Eliakim, R., Goetz, G. S., Rubio, S., Chailley Heu, B., Shao, J. S., Ducroc, R., and Alpers, D. H. (1997). Isolation and characterization of surfactant-like particles in rat and human colon. Am J Physiol *272*, G425-34.

Emi, M., Asaoka, H., Matsumoto, A., Itakura, H., Kurihara, Y., Wada, Y., Kanamori, H., Yazaki, Y., Takahashi, E., Lepert, M., and et al. (1993). Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene. J Biol Chem *268*, 2120-5.

Endo, Y., Sato, T., Matsushita, M., and Fujita, T. (1996a). Exon structure of the gene encoding the human mannose-binding protein-associated serine protease light chain: comparison with complement C1r and C1s genes. Int Immunol *8*, 1355-8.

Endo, Y., Sato, Y., Matsushita, M., and Fujita, T. (1996b). Cloning and characterization of the human lectin P35 gene and its related gene. Genomics *36*, 515-521.

Endo, Y., Takahashi, M., Nakao, M., Saiga, H., Sekine, H., Matsushita, M., Nonaka, M., and Fujita, T. (1998). Two lineages of mannose-binding lectin-assosciated serine protease (MASP) in vertebrates. J Immunol *161*, 4924-30.

Epstein, J., Eichbaum, Q., Sheriff, S., and Ezekowitz, R. A. (1996). The collectins in innate immunity. Curr Opin Immunol *8*, 29-35.

Exposito, J. Y., and Garrone, R. (1990). Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearance of two collagen gene families. Proc Natl Acad Sci U S A *87*, 6669-73.

Exposito, J. Y., Le Guellec, D., Lu, Q., and Garrone, R. (1991). Short chain collagens in sponges are encoded by a family of closely related genes. J Biol Chem *266*, 21923-8.

Faisst, S., and Meyer, S. (1992). Compilation of vertebrate-encoded transcription factors. Nucleic Acids Res *20*, 3-26.

Fearon, D. T. (1998). The complement system and adaptive immunity. Semin Immunol *10*, 355-61.

Fearon, D. T., and Locksley, R. M. (1996). The instructive role of innate immunity in the acquired immune response. Science *272*, 50-3.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem *132*, 6-13.

Ferguson, J. S., Voelker, D. R., McCormack, F. X., and Schlesinger, L. S. (1999). Surfactant protein D binds to Mycobacterium tuberculosis bacilli and lipoarabinomannan via carbohydrate-lectin interactions resulting in reduced phagocytosis of the bacteria by macrophages. J Immunol *163*, 312-21.

Fisher, J. H., Kao, F. T., Jones, C., White, R. T., Benson, B. J., and Mason, R. J. (1987). The coding sequence for the 32,000-dalton pulmonary surfactant-associated protein A is located on chromosome 10 and identifies two separate restriction-fragment-length polymorphisms. Am J Hum Genet *40*, 503-11.

Fisher, J. H., and Mason, R. (1995). Expression of pulmonary surfactant protein D in rat gastric mucosa. Am J Respir Cell Mol Biol *12*, 13-8.

Floros, J., and Hoover, R. R. (1998). Genetics of the hydrophilic surfactant proteins A and D. Biochim Biophys Acta *1408*, 312-22.

Floros, J., Veletza, S. V., Kotikalapudi, P., Krizkova, L., Karinch, A. M., Friedman, C., Buchter, S., and Marks, K. (1995). Dinucleotide repeats in the human surfactant protein-B gene and respiratory-distress syndrome. Biochem J *305*, 583-90.

Frohman, M. A., Dush, M. K., and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A *85*, 8998-9002.

Gallagher, D. S., Jr., Ryan, A. M., Liou, L. S., Sastry, K. N., and Womack, J. E. (1993). Somatic cell mapping of conglutinin (CGN1) to cattle syntenic group U29 and fluorescence in situ localization to Chromosome 28. Mamm Genome *4*, 716-9.

Gao, E., Alcorn, J. L., and Mendelson, C. R. (1993). Identification of enhancers in the 5'-flanking region of the rabbit surfactant protein A (SP-A) gene and characterization of their binding proteins. J Biol Chem *268*, 19697-709.

Gao, E., Wang, Y., McCormick, S. M., Li, J., Seidner, S. R., and Mendelson, C. R. (1996). Characterization of two baboon surfactant protein A genes. Am J Physiol *271*, L617-30.

Garred, P., Madsen, H. O., Balslev, U., Hofmann, B., Pedersen, C., Gerstoft, J., and Svejgaard, A. (1997). Susceptibility to HIV infection and progression of AIDS in relation to variant alleles of mannose-binding lectin. Lancet *349*, 236-40.

Garred, P., Madsen, H. O., Hofmann, B., and Svejgaard, A. (1995). Increased frequency of homozygosity of abnormal mannan-binding-protein alleles in patients with suspected immunodeficiency. Lancet *346*, 941-3.

Gaynor, C. D., McCormack, F. X., Voelker, D. R., McGowan, S. E., and Schlesinger, L. S. (1995). Pulmonary surfactant protein A mediates enhanced phagocytosis of Mycobacterium tuberculosis by a direct interaction with human macrophages. J Immunol *155*, 5343-51.

Geertsma, M. F., Nibbering, P. H., Haagsman, H. P., Daha, M. R., and van Furth, R. (1994). Binding of surfactant protein A to C1q receptors mediates phagocytosis of Staphylococcus aureus by monocytes. Am J Physiol *267*, L578-84.

Gerard, N. P., Bao, L., Xiao Ping, H., Eddy, R. L., Jr., Shows, T. B., and Gerard, C. (1993). Human chemotaxis receptor genes cluster at 19q13.3-13.4. Characterization of the human C5a receptor gene. Biochemistry *32*, 1243-50.

Gericke, A., Flach, C. R., and Mendelsohn, R. (1997). Structure and orientation of lung surfactant SP-C and L-alpha-dipalmitoylphosphatidylcholine in aqueous monolayers. Biophys J *73*, 492-9.

Ghildyal, R., Hartley, C., Varrasso, A., Meanger, J., Voelker, D. R., Anders, E. M., and Mills, J. (1999). Surfactant protein A binds to the fusion glycoprotein of respiratory syncytial virus and neutralizes virion infectivity. J Infect Dis *180*, 2009-13.

Gibbs, M., Collick, A., Kelly, R. G., and Jeffreys, A. J. (1993). A tetranucleotide repeat mouse minisatellite displaying substantial somatic instability during early preimplantation development. Genomics *17*, 121-8.

Glasser, S. W., Korfhagen, T. R., Weaver, T., Pilot Matias, T., Fox, J. L., and Whitsett, J. A. (1987). cDNA and deduced amino acid sequence of human pulmonary surfactant-associated proteolipid SPL(Phe). Proc Natl Acad Sci U S A *84*, 4007-11.

Glasser, S. W., Korfhagen, T. R., Weaver, T. E., Clark, J. C., Pilot Matias, T., Meuth, J., Fox, J. L., and Whitsett, J. A. (1988). cDNA, deduced polypeptide structure and chromosomal assignment of human pulmonary surfactant proteolipid, SPL(pVal). J Biol Chem *263*, 9-12.

Goerke, J. (1998). Pulmonary surfactant: functions and molecular composition. Biochim Biophys Acta *1408*, 79-89.

Goss, K. L., Kumar, A. R., and Snyder, J. M. (1998). SP-A2 gene expression in human fetal lung airways. Am J Respir Cell Mol Biol *19*, 613-21.

Goureau, A., Yerle, M., Schmitz, A., Riquet, J., Milan, D., Pinton, P., Frelat, G., and Gellin, J. (1996). Human and porcine correspondence of chromosome segments using bidirectional chromosome painting. Genomics *36*, 252-62.

Gunther, A., Siebert, C., Schmidt, R., Ziegler, S., Grimminger, F., Yabut, M., Temmesfeld, B., Walmrath, D., Morr, H., and Seeger, W. (1996). Surfactant alterations in severe pneumonia, acute respiratory distress syndrome, and cardiogenic lung edema. Am J Respir Crit Care Med *153*, 176-84.

Guo, N., Mogues, T., Weremowicz, S., Morton, C. C., and Sastry, K. N. (1998). The human ortholog of rhesus mannose-binding protein-A gene is an expressed pseudogene that localizes to chromosome 10. Mamm Genome *9*, 246-9.

Gurr, S. J., McPherson, M. J., Scollan, C., Atkinson, H. J., and Bowles, D. J. (1991). Gene expression in nematode-infected plant roots. Mol Gen Genet *226*, 361-6.

Hakansson, K., Lim, N. K., Hoppe, H. J., and Reid, K. B. (1999). Crystal structure of the trimeric alpha-helical coiled-coil and the three lectin domains of human lung surfactant protein D. Structure *7*, 255-64.

Halliday, H. L. (1997). Synthetic or natural surfactants. Acta Paediatr *86*, 233-7.

Hamada, H., Seidman, M., Howard, B. H., and Gorman, C. M. (1984). Enhanced gene expression by the poly(dT-dG).poly(dC-dA) sequence. Mol Cell Biol *4*, 2622-30.

Hancock, W. W. (1997). Delayed xenograft rejection. World J Surg *21*, 917-23.

Hardon, E. M., Frain, M., Paonessa, G., and Cortese, R. (1988). Two distinct factors interact with the promoter regions of several liver-specific genes. EMBO J *7*, 1711-9.

Harrod, K. S., Mounday, A. D., Stripp, B. R., and Whitsett, J. A. (1998). Clara cell secretory protein decreases lung inflammation after acute virus infection. Am J Physiol *275*, L924-30.

Hartshorn, K., Chang, D., Rust, K., White, M., Heuser, J., and Crouch, E. (1996a). Interactions of recombinant human pulmonary surfactant protein D and SP-D multimers with influenza A. Am J Physiol *271*, L753-62.

Hartshorn, K. L., Crouch, E., White, M. R., Colamussi, M. L., Kakkanatt, A., Tauber, B., Shepherd, V., and Sastry, K. N. (1998). Pulmonary surfactant proteins A and D enhance neutrophil uptake of bacteria. Am J Physiol *274*, L958-69.

Hartshorn, K. L., Crouch, E. C., White, M. R., Eggleton, P., Tauber, A. I., Chang, D., and Sastry, K. (1994). Evidence for a protective role of pulmonary surfactant protein D (SP-D) against influenza A viruses. J Clin Invest *94*, 311-9.

Hartshorn, K. L., Reid, K. B., White, M. R., Jensenius, J. C., Morris, S. M., Tauber, A. I., and Crouch, E. (1996b). Neutrophil deactivation by influenza A viruses: mechanisms of protection after viral opsonization with collectins and hemagglutination-inhibiting antibodies. Blood *87*, 3450-61.

Hartshorn, K. L., Sastry, K., White, M. R., Anders, E. M., Super, M., Ezekowitz, R. A., and Tauber, A. I. (1993). Human mannose-binding protein functions as an opsonin for influenza A viruses. J Clin Invest *91*, 1414-20.

Haurum, J. S., Thiel, S., Haagsman, H. P., Laursen, S. B., Larsen, B., and Jensenius, J. C. (1993a). Studies on the carbohydrate-binding characteristics of human pulmonary surfactant-associated protein A and comparison with two other collectins: mannan-binding protein and conglutinin. Biochem J *293*, 873-8.

Haurum, J. S., Thiel, S., Jones, I. M., Fischer, P. B., Laursen, S. B., and Jensenius, J. C. (1993b). Complement activation upon binding of mannan-binding protein to HIV envelope glycoproteins. Aids *7*, 1307-13.

Hawgood, S., Benson, B. J., Schilling, J., Damm, D., Clements, J. A., and White, R. T. (1987). Nucleotide and amino acid sequences of pulmonary surfactant protein SP

18 and evidence for cooperation between SP 18 and SP 28-36 in surfactant lipid adsorption. Proc Natl Acad Sci U S A *84*, 66-70.

Heery, D. M., Gannon, F., and Powell, R. (1990). A simple method for subcloning DNA fragments from gel slices. Trends Genet *6*, 173.

Heine Suner, D., Diaz Guillen, M. A., de Villena, F. P., Robledo, M., Benitez, J., and Rodriguez de Cordoba, S. (1997). A high-resolution map of the regulator of the complement activation gene cluster on 1q32 that integrates new genes and markers. Immunogenetics *45*, 422-7.

Hellqvist, M., Mahlapuu, M., Samuelsson, L., Enerback, S., and Carlsson, P. (1996). Differential activation of lung-specific genes by two forkhead proteins, FREAC-1 and FREAC-2. J Biol Chem *271*, 4482-90.

Helms, C., Graham, M. Y., Dutchik, J. E., and Olson, M. V. (1985). A new method for purifying lambda DNA from phage lysates. Dna *4*, 39-49.

Hickling, T. P., Bright, H., Wing, K., Gower, D., Martin, S. L., Sim, R. B., and Malhotra, R. (1999). A recombinant trimeric surfactant protein D carbohydrate recognition domain inhibits respiratory syncytial virus infection in vitro and in vivo. Eur J Immunol *29*, 3478-84.

Hickling, T. P., Malhotra, R., and Sim, R. B. (1998a). Human lung surfactant protein A exists in several different oligomeric states: oligomer size distribution varies between patient groups. Mol Med *4*, 266-75.

Hickling, T. P., Sim, R. B., and Malhotra, R. (1998b). Induction of TNF-alpha release from human buffy coat cells by Pseudomonas aeruginosa is reduced by lung surfactant protein A. FEBS Lett *437*, 65-9.

Hickman Davis, J., Gibbs Erwin, J., Lindsey, J. R., and Matalon, S. (1999). Surfactant protein A mediates mycoplasmacidal activity of alveolar macrophages by production of peroxynitrite. Proc Natl Acad Sci U S A *96*, 4953-8.

Hickman Davis, J. M., Lindsey, J. R., Zhu, S., and Matalon, S. (1998). Surfactant protein A mediates mycoplasmacidal activity of alveolar macrophages. Am J Physiol *274*, L270-7.

Hickman Davis, J. M., Michalek, S. M., Gibbs Erwin, J., and Lindsey, J. R. (1997). Depletion of alveolar macrophages exacerbates respiratory mycoplasmosis in mycoplasma-resistant C57BL mice but not mycoplasma-susceptible C3H mice. Infect Immun *65*, 2278-82.

Hollmann, T. J., Haviland, D. L., Kildsgaard, J., Watts, K., and Wetsel, R. A. (1998). Cloning, expression, sequence determination, and chromosome localization of the mouse complement C3a anaphylatoxin receptor gene. Mol Immunol *35*, 137-48.

Holmskov, U., Lawson, P., Teisner, B., Tornoe, I., Willis, A. C., Morgan, C., Koch, C., and Reid, K. B. (1997). Isolation and characterization of a new member of the scavenger receptor superfamily, glycoprotein-340 (gp-340), as a lung surfactant protein-D binding molecule. J Biol Chem *272*, 13743-9.

Holmskov, U., Malhotra, R., Sim, R. B., and Jensenius, J. C. (1994). Collectins: collagenous C-type lectins of the innate immune defense system. Immunol Today *15*, 67-74.

Holmskov, U., Mollenhauer, J., Madsen, J., Vitved, L., Gronlund, J., Tornoe, I., Kliem, A., Reid, K. B., Poustka, A., and Skjodt, K. (1999). Cloning of gp-340, a putative opsonin receptor for lung surfactant protein D. Proc Natl Acad Sci U S A *96*, 10794-9.

Holmskov, U., Teisner, B., Willis, A. C., Reid, K. B., and Jensenius, J. C. (1993). Purification and characterization of a bovine serum lectin (CL-43) with structural homology to conglutinin and SP-D and carbohydrate specificity similar to mannan-binding protein. J Biol Chem *268*, 10120-5.

Holton, T. A., and Graham, M. W. (1991). A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res *19*, 1156.

Hoover, R. R., and Floros, J. (1998). Organization of the human SP-A and SP-D loci at 10q22-q23. Physical and radiation hybrid mapping reveal gene order and orientation. Am J Respir Cell Mol Biol *18*, 353-62.

Hoover, R. R., and Floros, J. (1999). SP-A 3'-UTR is involved in the glucocorticoid inhibition of human SP-A gene expression. Am J Physiol *276*, L917-24.

Hopken, U. E., Lu, B., Gerard, N. P., and Gerard, C. (1996). The C5a chemoattractant receptor mediates mucosal defence to infection. Nature *383*, 86-9.

Hopken, U. E., Lu, B., Gerard, N. P., and Gerard, C. (1997). Impaired inflammatory responses in the reverse arthus reaction through genetic deletion of the C5a receptor. J Exp Med *186*, 749-56.

Hoppe, H. J., Barlow, P. N., and Reid, K. B. (1994). A parallel three stranded alpha-helical bundle at the nucleation site of collagen triple-helix formation. FEBS Lett *344*, 191-5.

Hsu, M. H., Ember, J. A., Wang, M., Prossnitz, E. R., Hugli, T. E., and Ye, R. D. (1998). Cloning and functional characterization of the mouse C3a anaphylatoxin receptor gene. Immunogenetics *47*, 64-72.

Iannuzzi, L., Gallagher, D. S., Liou, L. S., Di Meo, G. P., Ryan, A. M., Sastry, K. N., and Womack, J. E. (1994). Chromosomal localization of conglutinin (CGN1) gene to river buffalo by sequential RBH-banding and FISH. Hereditas *120*, 283-6.

Ihara, I., Ueda, H., Suzuki, A., and Kawakami, M. (1982). Physicochemical properties of a new bactericidal factor, Ra-reactive factor. Biochem Biophys Res Commun *107*, 1185-90.

Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T., and Yamashina, I. (1987). Serum lectin with known structure activates complement through the classical pathway. J Biol Chem *262*, 7451-4.

Ikegami, M., Korfhagen, T. R., Bruno, M. D., Whitsett, J. A., and Jobe, A. H. (1997). Surfactant metabolism in surfactant protein A-deficient mice. Am J Physiol *272*, L479-85.

Ikegami, M., Korfhagen, T. R., Whitsett, J. A., Bruno, M. D., Wert, S. E., Wada, K., and Jobe, A. H. (1998). Characteristics of surfactant from SP-A-deficient mice. Am J Physiol *275*, L247-54.

Iobst, S. T., Wormald, M. R., Weis, W. I., Dwek, R. A., and Drickamer, K. (1994). Binding of sugar ligands to Ca(2+)-dependent animal lectins. I. Analysis of mannose binding by site-directed mutagenesis and NMR. J Biol Chem *269*, 15505-11.

Irwin, D. M. (1988). Evolution of an active-site codon in serine proteases [letter]. Nature *336*, 429-30.

Irwin, D. M., Robertson, K. A., and MacGillivray, R. T. (1988). Structure and evolution of the bovine prothrombin gene. J Mol Biol *200*, 31-45.

Jacobs, K. A., Phelps, D. S., Steinbrink, R., Fisch, J., Kriz, R., Mitsock, L., Dougherty, J. P., Taeusch, H. W., and Floros, J. (1987). Isolation of a cDNA clone encoding a high molecular weight precursor to a 6-kDa pulmonary surfactant-associated protein [published erratum appears in J Biol Chem 1988 Jan 15;263(2):1093]. J Biol Chem *262*, 9808-11.

Ji, X., Azumi, K., Sasaki, M., and Nonaka, M. (1997). Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding protein-associated serine protease from a urochordate, the Japanese ascidian, Halocynthia roretzi. Proc Natl Acad Sci U S A *94*, 6340-5.

Ji, Y. H., Fujita, T., Hatsuse, H., Takahashi, A., Matsushita, M., and Kawakami, M. (1993). Activation of the C4 and C2 components of complement by a proteinase in serum bactericidal factor, Ra reactive factor. J Immunol *150*, 571-8.

Jobe, A. H., and Ikegami, M. (1998). Surfactant and acute lung injury. Proc Assoc Am Physicians *110*, 489-95.

Johansson, J., and Curstedt, T. (1997). Molecular structures and interactions of pulmonary surfactant components. Eur J Biochem *244*, 675-93.

Johansson, J., Szyperski, T., and Wuthrich, K. (1995). Pulmonary surfactant-associated polypeptide SP-C in lipid micelles: CD studies of intact SP-C and NMR secondary structure determination of depalmitoyl-SP-C(1-17). FEBS Lett *362*, 261-5.

Kabha, K., Schmegner, J., Keisari, Y., Parolis, H., Schlepper Schaeffer, J., and Ofek, I. (1997). SP-A enhances phagocytosis of Klebsiella by interaction with capsular polysaccharides and alveolar macrophages. Am J Physiol *272*, L344-52.

Kala, P., Ten Have, T., Nielsen, H., Dunn, M., and Floros, J. (1998). Association of pulmonary surfactant protein A (SP-A) gene and respiratory distress syndrome: interaction with SP-B. Pediatr Res *43*, 169-77.

Kalina, M., Blau, H., Riklis, S., and Kravtsov, V. (1995). Interaction of surfactant protein A with bacterial lipopolysaccharide may affect some biological functions. Am J Physiol *268*, L144-51.

Kamp, E. M., Stockhofe Zurwieden, N., van Leengoed, L. A., and Smits, M. A. (1997). Endobronchial inoculation with Apx toxins of Actinobacillus pleuropneumoniae leads to pleuropneumonia in pigs. Infect Immun *65*, 4350-4.

Karinch, A. M., Deiter, G., Ballard, P. L., and Floros, J. (1998). Regulation of expression of human SP-A1 and SP-A2 genes in fetal lung explant culture. Biochim Biophys Acta *1398*, 192-202.

Karinch, A. M., deMello, D. E., and Floros, J. (1997). Effect of genotype on the levels of surfactant protein A mRNA and on the SP-A2 splice variants in adult humans. Biochem J *321*, 39-47.

Karinch, A. M., and Floros, J. (1995). 5' splicing and allelic variants of the human pulmonary surfactant protein A genes. Am J Respir Cell Mol Biol *12*, 77-88.

Kashi, Y., King, D., and Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. Trends Genet *13*, 74-78.

Katyal, S. L., Singh, G., and Locker, J. (1992). Characterization of a second human pulmonary surfactant-associated protein SP-A gene. Am J Respir Cell Mol Biol *6*, 446-52.

Kawakami, M., Ihara, I., Ihara, S., Suzuki, A., and Fukui, K. (1984). A group of bactericidal factors conserved by vertebrates for more than 300 million years. J Immunol *132*, 2578-81.

Kawasaki, N., Itoh, N., and Kawasaki, T. (1994). Gene organization and 5'-flanking region sequence of conglutinin: a C-type mammalian lectin containing a collagen-like domain. Biochem Biophys Res Commun *198*, 597-604.

Kelly, R., Gibbs, M., Collick, A., and Jeffreys, A. J. (1991). Spontaneous mutation at the hypervariable mouse minisatellite locus Ms6-hm: flanking DNA sequence and analysis of germline and early somatic mutation events. Proc R Soc Lond B Biol Sci *245*, 235-45.

Kelly, R. G. (1994). Similar origins of two mouse minisatellites within transposon-like LTRs. Genomics *24*, 509-15.

Kelly, S. E., Bachurski, C. J., Burhans, M. S., and Glasser, S. W. (1996). Transcription of the lung-specific surfactant protein C gene is mediated by thyroid transcription factor 1. J Biol Chem *271*, 6881-8.

Kerr, M. H., and Paton, J. Y. (1999). Surfactant protein levels in severe respiratory syncytial virus infection. Am J Respir Crit Care Med *159*, 1115-8.

Koop, B. F., and Hood, L. (1994). Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. Nat Genet *7*, 48-53.

Korfhagen, T. R., Bruno, M. D., Glasser, S. W., Ciraolo, P. J., Whitsett, J. A., Lattier, D. L., Wikenheiser, K. A., and Clark, J. C. (1992). Murine pulmonary surfactant SP-A gene: cloning, sequence, and transcriptional activity. Am J Physiol *263*, L546-54.

Korfhagen, T. R., Bruno, M. D., Ross, G. F., Huelsman, K. M., Ikegami, M., Jobe, A. H., Wert, S. E., Stripp, B. R., Morris, R. E., Glasser, S. W., Bachurski, C. J., Iwamoto, H. S., and Whitsett, J. A. (1996). Altered surfactant function and structure in SP-A gene targeted mice. Proc Natl Acad Sci U S A *93*, 9594-9.

Korfhagen, T. R., Glasser, S. W., Bruno, M. D., McMahan, M. J., and Whitsett, J. A. (1991). A portion of the human surfactant protein A (SP-A) gene locus consists of a pseudogene. Am J Respir Cell Mol Biol *4*, 463-9.

Korfhagen, T. R., Sheftelyevich, V., Burhans, M. S., Bruno, M. D., Ross, G. F., Wert, S. E., Stahlman, M. T., Jobe, A. H., Ikegami, M., Whitsett, J. A., and Fisher, J. H. (1998). Surfactant protein-D regulates surfactant phospholipid homeostasis in vivo. J Biol Chem *273*, 28438-43.

Kuan, S. F., Persson, A., Parghi, D., and Crouch, E. (1994). Lectin-mediated interactions of surfactant protein D with alveolar macrophages. Am J Respir Cell Mol Biol *10*, 430-6.

Kuan, S. F., Rust, K., and Crouch, E. (1992). Interactions of surfactant protein D with bacterial lipopolysaccharides. Surfactant protein D is an Escherichia coli-binding protein in bronchoalveolar lavage. J Clin Invest *90*, 97-106.

Kuhlman, M., Joiner, K., and Ezekowitz, R. A. (1989). The human mannose-binding protein functions as an opsonin. J Exp Med *169*, 1733-45.

Kuroki, Y., Gasa, S., Ogasawara, Y., Shiratori, M., Makita, A., and Akino, T. (1992). Binding specificity of lung surfactant protein SP-D for glucosylceramide. Biochem Biophys Res Commun *187*, 963-9.

Kuroki, Y., McCormack, F. X., Ogasawara, Y., Mason, R. J., and Voelker, D. R. (1994). Epitope mapping for monoclonal antibodies identifies functional domains of pulmonary surfactant protein A that interact with lipids. J Biol Chem *269*, 29793-800.

Kuroki, Y., Takahashi, H., Chiba, H., and Akino, T. (1998). Surfactant proteins A and D: disease markers. Biochim Biophys Acta *1408*, 334-45.

Kusumoto, H., Hirosawa, S., Salier, J. P., Hagen, F. S., and Kurachi, K. (1988). Human genes for complement components C1r and C1s in a close tail-to-tail arrangement. Proc Natl Acad Sci U S A *85*, 7307-11.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-5.

LaForce, F. M., Kelly, W. J., and Huber, G. L. (1973). Inactivation of staphylococci by alveolar macrophages with preliminary observations on the importance of alveolar lining material. Am Rev Respir Dis *108*, 784-90.

Laursen, S. B., Hedemand, J. E., Thiel, S., Willis, A. C., Skriver, E., Madsen, P. S., and Jensenius, J. C. (1995). Collectin in a non-mammalian species: isolation and characterization of mannan-binding protein (MBP) from chicken serum. Glycobiology *5*, 553-61.

Lawson, P. R., Perkins, V. C., Holmskov, U., and Reid, K. B. M. (1999). Genomic organisation of the mouse gene for lung surfactant protein D (Sftpd). Am J Respir Cell Mol Biol *20*, 953-63.

Lawson, P. R., and Reid, K. B. M. (2000a). Genetic mapping of the murine anaphylatoxin receptor for C3a and three activating serine proteases of the complement system (Masp-2, C1r and C1s) by a new rapid PCR based method. Int Immunol *12*, 231-40.

Lawson, P. R., and Reid, K. B. M. (2000b). The roles of surfactant protein A and D in innate immunity. Immunol Rev *173*, 66-78.

Lee, Y. M., Leiby, K. R., Allar, J., Paris, K., Lerch, B., and Okarma, T. B. (1991). Primary structure of bovine conglutinin, a member of the C-type animal lectin family. J Biol Chem *266*, 2715-23.

LeVine, A. M., Bruno, M. D., Huelsman, K. M., Ross, G. F., Whitsett, J. A., and Korfhagen, T. R. (1997). Surfactant protein A-deficient mice are susceptible to group B streptococcal infection. J Immunol *158*, 4336-40.

LeVine, A. M., Gwozdz, J., Stark, J., Bruno, M., Whitsett, J., and Korfhagen, T. (1999a). Surfactant protein-A enhances respiratory syncytial virus clearance in vivo. J Clin Invest *103*, 1015-21.

LeVine, A. M., Kurak, K. E., Bruno, M. D., Stark, J. M., Whitsett, J. A., and Korfhagen, T. R. (1998). Surfactant protein-A-deficient mice are susceptible to Pseudomonas aeruginosa infection. Am J Respir Cell Mol Biol *19*, 700-8.

LeVine, A. M., Kurak, K. E., Wright, J. R., Watford, W. T., Bruno, M. D., Ross, G. F., Whitsett, J. A., and Korfhagen, T. R. (1999b). Surfactant protein-A binds group B streptococcus enhancing phagocytosis and clearance from lungs of surfactant protein-A-deficient mice. Am J Respir Cell Mol Biol *20*, 279-86.

Liepinsh, E., Andersson, M., Ruysschaert, J. M., and Otting, G. (1997). Saposin fold revealed by the NMR structure of NK-lysin [letter]. Nat Struct Biol *4*, 793-5.

Lim, B. L., Lu, J., and Reid, K. B. (1993). Structural similarity between bovine conglutinin and bovine lung surfactant protein D and demonstration of liver as a site of synthesis of conglutinin. Immunology *78*, 159-65.

Lim, B. L., Wang, J. Y., Holmskov, U., Hoppe, H. J., and Reid, K. B. (1994a). Expression of the carbohydrate recognition domain of lung surfactant protein D and demonstration of its binding to lipopolysaccharides of gram-negative bacteria. Biochem Biophys Res Commun *202*, 1674-80.

Lim, B. L., Willis, A. C., Reid, K. B., Lu, J., Laursen, S. B., Jensenius, J. C., and Holmskov, U. (1994b). Primary structure of bovine collectin-43 (CL-43). Comparison with conglutinin and lung surfactant protein-D. J Biol Chem *269*, 11820-4.

Limper, A. H., Crouch, E. C., O'Riordan, D. M., Chang, D., Vuk Pavlovic, Z., Standing, J. E., Kwon, K. Y., and Adlakha, A. (1995). Surfactant protein-D modulates interaction of Pneumocystis carinii with alveolar macrophages [see comments]. J Lab Clin Med *126*, 416-22.

Lin, S., Na, C. L., Akinbi, H. T., Apsley, K. S., Whitsett, J. A., and Weaver, T. E. (1999). Surfactant protein B (SP-B) -/- mice are rescued by restoration of SP-B expression in alveolar type II cells but not Clara cells. J Biol Chem *274*, 19168-74.

Liou, L. S., Sastry, R., Hartshorn, K. L., Lee, Y. M., Okarma, T. B., Tauber, A. I., and Sastry, K. N. (1994a). Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization of the BC cDNA reveals strong homology to surfactant protein-D. Gene *141*, 277-81.

Liou, L. S., Sastry, R., Hartshorn, K. L., Lee, Y. M., Okarma, T. B., Tauber, A. I., and Sastry, K. N. (1994b). Bovine conglutinin gene exon structure reveals its evolutionary relationship to surfactant protein-D. J Immunol *153*, 173-80.

Loos, M., and Colomb, M. (1993). C1, the first component of complement: structure-function-relationship of C1q and collectins (MBP, SP-A, SP-D, conglutinin), C1-esterases (C1r and C1s), and C1-inhibitor in health and disease. Behring Inst Mitt *93*, 1-5.

Lu, J. (1997). Collectins: collectors of microorganisms for the innate immune system. Bioessays *19*, 509-18.

Lu, J., Laursen, S. B., Thiel, S., Jensenius, J. C., and Reid, K. B. (1993a). The cDNA cloning of conglutinin and identification of liver as a primary site of synthesis of conglutinin in members of the Bovidae. Biochem J *292*, 157-62.

Lu, J., Wiedemann, H., Timpl, R., and Reid, K. B. (1993b). Similarity in structure between C1q and the collectins as judged by electron microscopy. Behring Inst Mitt *93*, 6-16.

Lu, J., Willis, A. C., and Reid, K. B. (1992). Purification, characterization and cDNA cloning of human lung surfactant protein D. Biochem J *284*, 795-802.

Lu, J. H., Thiel, S., Wiedemann, H., Timpl, R., and Reid, K. B. (1990). Binding of the pentamer/hexamer forms of mannan-binding protein to zymosan activates the proenzyme C1r2C1s2 complex, of the classical pathway of complement, without involvement of C1q. J Immunol *144*, 2287-94.

Madan, T., Eggleton, P., Kishore, U., Strong, P., Aggrawal, S. S., Sarma, P. U., and Reid, K. B. (1997a). Binding of pulmonary surfactant proteins A and D to Aspergillus fumigatus conidia enhances phagocytosis and killing by human neutrophils and alveolar macrophages. Infect Immun *65*, 3171-9.

Madan, T., Kishore, U., Shah, A., Eggleton, P., Strong, P., Wang, J. Y., Aggrawal, S. S., Sarma, P. U., and Reid, K. B. (1997b). Lung surfactant proteins A and D can inhibit specific IgE binding to the allergens of Aspergillus fumigatus and block allergen-induced histamine release from human basophils. Clin Exp Immunol *110*, 241-9.

Madsen, H. O., Garred, P., Kurtzhals, J. A., Lamm, L. U., Ryder, L. P., Thiel, S., and Svejgaard, A. (1994). A new frequent allele is the missing link in the structural polymorphism of the human mannan-binding protein. Immunogenetics *40*, 37-44.

Madsen, H. O., Garred, P., Thiel, S., Kurtzhals, J. A., Lamm, L. U., Ryder, L. P., and Svejgaard, A. (1995). Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein. J Immunol *155*, 3013-20.

Madsen, H. O., Satz, M. L., Hogh, B., Svejgaard, A., and Garred, P. (1998). Different molecular events result in low protein levels of mannan-binding lectin in populations from southeast Africa and South America. J Immunol *161*, 3169-75.

Maeda, N., Yang, F., Barnett, D. R., Bowman, B. H., and Smithies, O. (1984). Duplication within the haptoglobin Hp2 gene. Nature *309*, 131-5.

Malhotra, R., Haurum, J., Thiel, S., Jensenius, J. C., and Sim, R. B. (1993). Pollen grains bind to lung alveolar type II cells (A549) via lung surfactant protein A (SP-A). Biosci Rep *13*, 79-90.

Malhotra, R., Haurum, J., Thiel, S., and Sim, R. B. (1992). Interaction of C1q receptor with lung surfactant protein A. Eur J Immunol *22*, 1437-45.

Malhotra, R., Haurum, J. S., Thiel, S., and Sim, R. B. (1994a). Binding of human collectins (SP-A and MBP) to influenza virus. Biochem J *304*, 455-61.

Malhotra, R., Lu, J., Holmskov, U., and Sim, R. B. (1994b). Collectins, collectin receptors and the lectin pathway of complement activation. Clin Exp Immunol *97 Suppl 2*, 4-9.

Manz Keinke, H., Egenhofer, C., Plattner, H., and Schlepper Schafer, J. (1991). Specific interaction of lung surfactant protein A (SP-A) with rat alveolar macrophages. Exp Cell Res *192*, 597-603.

Manz Keinke, H., Plattner, H., and Schlepper Schafer, J. (1992). Lung surfactant protein A (SP-A) enhances serum-independent phagocytosis of bacteria by alveolar macrophages. Eur J Cell Biol *57*, 95-100.

Marra, M., Hillier, L., Kucaba, T., Allen, M., Barstead, R., Beck, C., Blistain, A., Bonaldo, M., Bowers, Y., Bowles, L., Cardenas, M., Chamberlain, A., Chappell, J., Clifton, S., Favello, A., Geisel, S., Gibbons, M., Harvey, N., Hill, F., Jackson, Y., Kohn, S., Lennon, G., Mardis, E., Martin, J., Waterston, R., and et al. (1999). An encyclopedia of mouse genes. Nat Genet *21*, 191-4.

Marra, M. A., Hillier, L., and Waterston, R. H. (1998). Expressed sequence tags--ESTablishing bridges between genomes. Trends Genet *14*, 4-7.

Matsushita, M. (1996). The lectin pathway of the complement system. Microbiol Immunol *40*, 887-93.

Matsushita, M., Endo, Y., Nonaka, M., and Fujita, T. (1998). Complement-related serine proteases in tunicates and vertebrates. Curr Opin Immunol *10*, 29-35.

Matsushita, M., Ezekowitz, R. A., and Fujita, T. (1995). The Gly-54-->Asp allelic form of human mannose-binding protein (MBP) fails to bind MBP-associated serine protease. Biochem J *311*, 1021-3.

Matsushita, M., and Fujita, T. (1992). Activation of the classical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease. J Exp Med *176*, 1497-502.

Matsushita, M., and Fujita, T. (1995). Cleavage of the third component of complement (C3) by mannose-binding protein-associated serine protease (MASP) with subsequent complement activation. Immunobiology *194*, 443-48.

McCormack, F. X., Calvert, H. M., Watson, P. A., Smith, D. L., Mason, R. J., and Voelker, D. R. (1994a). The structure and function of surfactant protein A. Hydroxyproline- and carbohydrate-deficient mutant proteins. J Biol Chem *269*, 5833-41.

McCormack, F. X., Festa, A. L., Andrews, R. P., Linke, M., and Walzer, P. D. (1997a). The carbohydrate recognition domain of surfactant protein A mediates binding to the major surface glycoprotein of Pneumocystis carinii. Biochemistry *36*, 8092-9.

McCormack, F. X., Kuroki, Y., Stewart, J. J., Mason, R. J., and Voelker, D. R. (1994b). Surfactant protein A amino acids Glu195 and Arg197 are essential for receptor binding, phospholipid aggregation, regulation of secretion, and the facilitated uptake of phospholipid by type II cells. J Biol Chem *269*, 29801-7.

McCormack, F. X., Pattanajitvilai, S., Stewart, J., Possmayer, F., Inchley, K., and Voelker, D. R. (1997b). The Cys6 intermolecular disulfide bond and the collagen-like region of rat SP-A play critical roles in interactions with alveolar type II cells and surfactant lipids. J Biol Chem *272*, 27971-9.

McCormack, F. X., Stewart, J., Voelker, D. R., and Damodarasamy, M. (1997c). Alanine mutagenesis of surfactant protein A reveals that lipid binding and pH-dependent liposome aggregation are mediated by the carbohydrate recognition domain. Biochemistry *36*, 13963-71.

McIntosh, J. C., Mervin Blake, S., Conner, E., and Wright, J. R. (1996a). Surfactant protein A protects growing cells and reduces TNF-alpha activity from LPS-stimulated macrophages. Am J Physiol *271*, L310-9.

McIntosh, J. C., Swyers, A. H., Fisher, J. H., and Wright, J. R. (1996b). Surfactant proteins A and D increase in response to intratracheal lipopolysaccharide. Am J Respir Cell Mol Biol *15*, 509-19.

McNeely, T. B., and Coonrod, J. D. (1993). Comparison of the opsonic activity of human surfactant protein A for Staphylococcus aureus and Streptococcus pneumoniae with rabbit and human macrophages. J Infect Dis *167*, 91-7.

McNeely, T. B., and Coonrod, J. D. (1994). Aggregation and opsonization of type A but not type B Hemophilus influenzae by surfactant protein A. Am J Respir Cell Mol Biol *11*, 114-22.

Medzhitov, R., and Janeway, C. A., Jr. (1997). Innate immunity: The virtues of a nonclonal system of recognition. Cell *91*, 295-298.

Mendelson, C. R., Acarregui, M. J., Odom, M. J., and Boggaram, V. (1991). Developmental and hormonal regulation of surfactant protein A (SP-A) gene expression in fetal lung. J Dev Physiol *15*, 61-9.

Mermer, B., Colb, M., and Krontiris, T. G. (1987). A family of short, interspersed repeats is associated with tandemly repetitive DNA in the human genome. Proc Natl Acad Sci U S A *84*, 3320-4.

Miyamura, K., Leigh, L. E., Lu, J., Hopkin, J., Lopez Bernal, A., and Reid, K. B. (1994). Surfactant protein D binding to alveolar macrophages. Biochem J *300*, 237-42.

Mogues, T., Ota, T., Tauber, A. I., and Sastry, K. N. (1996). Characterization of two mannose-binding protein cDNAs from rhesus monkey (Macaca mulatta): structure and evolutionary implications. Glycobiology *6*, 543-50.

Moore, K. J., D'Amore Bruno, M. A., Korfhagen, T. R., Glasser, S. W., Whitsett, J. A., Jenkins, N. A., and Copeland, N. G. (1992). Chromosomal localization of three pulmonary surfactant protein genes in the mouse. Genomics *12*, 388-93.

Morio, H., Kurata, H., Katsuyama, R., Oka, S., Kozutsumi, Y., and Kawasaki, T. (1997). Renal expression of serum-type mannan-binding protein in rat. Eur J Biochem *243*, 770-4.

Morrissey, J. H. (1981). Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. Anal Biochem *117*, 307-10.

Mott, R. (1997). EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. Comput Appl Biosci *13*, 477-8.

Motwani, M., White, R. A., Guo, N., Dowler, L. L., Tauber, A. I., and Sastry, K. N. (1995). Mouse surfactant protein-D. cDNA cloning, characterization, and gene localization to chromosome 14. J Immunol *155*, 5671-7.

Murphy, P. M., Ozcelik, T., Kenney, R. T., Tiffany, H. L., McDermott, D., and Francke, U. (1992). A structural homologue of the N-formyl peptide receptor. Characterization and chromosome mapping of a peptide chemoattractant receptor family. J Biol Chem *267*, 7637-43.

Nag, K., Taneva, S. G., Perez Gil, J., Cruz, A., and Keough, K. M. (1997). Combinations of fluorescently labeled pulmonary surfactant proteins SP-B and SP-C in phospholipid films. Biophys J *72*, 2638-50.

Nepomuceno, R. R., Henschen Edman, A. H., Burgess, W. H., and Tenner, A. J. (1997). cDNA cloning and primary structure analysis of C1qR(P), the human C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro. Immunity *6*, 119-29.

Ng, K. K., Drickamer, K., and Weis, W. I. (1996). Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. J Biol Chem *271*, 663-74.

Nguyen, V. C., Tosi, M., Gross, M. S., Cohen Haguenauer, O., Jegou Foubert, C., de Tand, M. F., Meo, T., and Frezal, J. (1988). Assignment of the complement serine protease genes C1r and C1s to chromosome 12 region 12p13. Hum Genet *78*, 363-8.

Nogee, L. M. (1998). Genetics of the hydrophobic surfactant proteins. Biochim Biophys Acta *1408*, 323-33.

Nogee, L. M., de Mello, D. E., Dehner, L. P., and Colten, H. R. (1993). Brief report: deficiency of pulmonary surfactant protein B in congenital alveolar proteinosis. N Engl J Med *328*, 406-10.

Nogee, L. M., Garnier, G., Dietz, H. C., Singer, L., Murphy, A. M., deMello, D. E., and Colten, H. R. (1994). A mutation in the surfactant protein B gene responsible for fatal neonatal respiratory disease in multiple kindreds. J Clin Invest *93*, 1860-3.

Nonaka, M., and Azumi, K. (1999). Opsonic complement system of the solitary ascidian, Halocynthia roretzi. Dev Comp Immunol *23*, 421-7.

Norsworthy, P. J., Taylor, P. R., Walport, M. J., and Botto, M. (1999). Cloning of the mouse homolog of the 126-kDa human C1q/MBL/SP-A receptor, C1qR(p). Mamm Genome *10*, 789-93.

O'Riordan, D. M., Standing, J. E., Kwon, K. Y., Chang, D., Crouch, E. C., and Limper, A. H. (1995). Surfactant protein D interacts with Pneumocystis carinii and mediates organism adherence to alveolar macrophages. J Clin Invest *95*, 2699-710.

Ogasawara, Y., Kuroki, Y., and Akino, T. (1992). Pulmonary surfactant protein D specifically binds to phosphatidylinositol. J Biol Chem *267*, 21244-9.

Ogasawara, Y., Kuroki, Y., Shiratori, M., Shimizu, H., Miyamura, K., and Akino, T. (1991). Ontogeny of surfactant apoprotein D, SP-D, in the rat lung. Biochim Biophys Acta *1083*, 252-6.

Ogasawara, Y., McCormack, F. X., Mason, R. J., and Voelker, D. R. (1994). Chimeras of surfactant proteins A and D identify the carbohydrate recognition domains as essential for phospholipid interaction. J Biol Chem *269*, 29785-92.

Ogasawara, Y., and Voelker, D. R. (1995). Altered carbohydrate recognition specificity engineered into surfactant protein D reveals different binding mechanisms for phosphatidylinositol and glucosylceramide. J Biol Chem *270*, 14725-32.

Ogata, R. T., Low, P. J., and Kawakami, M. (1995). Substrate specificities of the protease of mouse serum Ra-reactive factor. J Immunol *154*, 2351-7.

Ohkuro, M., Kobayashi, K., Takahashi, K., and Nagasawa, S. (1994). Effect of C1q on the processing of immune complexes by human neutrophils. Immunology *83*, 507-11.

Ohmer Schrock, D., Schlatterer, C., Plattner, H., and Schlepper Schafer, J. (1995). Lung surfactant protein A (SP-A) activates a phosphoinositide/calcium signaling pathway in alveolar macrophages. J Cell Sci *108*, 3695-702.

Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Yamazaki, H., Shimada, T., Keshi, H., Sakai, Y., Fukuoh, A., Sakamoto, T., and Wakamiya, N. (1999). Molecular cloning of a novel human collectin from liver (CL-L1). J Biol Chem *274*, 13681-9.

Oka, S., Kawasaki, T., and Yamashina, I. (1985). Isolation and characterization of mannan-binding proteins from chicken liver. Arch Biochem Biophys *241*, 95-105.

Oohashi, T., Ueki, Y., Sugimoto, M., and Ninomiya, Y. (1995). Isolation and structure of the COL4A6 gene encoding the human alpha 6(IV) collagen chain and comparison with other type IV collagen genes. J Biol Chem *270*, 26863-7.

Orgeig, S., Barr, H. A., and Nicholas, T. E. (1995). Effect of hyperpnea on the cholesterol to disaturated phospholipid ratio in alveolar surfactant of rats. Exp Lung Res *21*, 157-74.

Overdier, D. G., Porcella, A., and Costa, R. H. (1994). The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. Mol Cell Biol *14*, 2755-66.

Paananen, R., Glumoff, V., and Hallman, M. (1999). Surfactant protein A and D expression in the porcine Eustachian tube. FEBS Lett. *452*, 141-4.

Paral, D., Sohns, B., Crass, T., Grove, M., Kohl, J., Klos, A., and Bautsch, W. (1998). Genomic organization of the human C3a receptor. Eur J Immunol *28*, 2417-23.

Pastinen, T., Liitsola, K., Niini, P., Salminen, M., and Syvanen, A. C. (1998). Contribution of the CCR5 and MBL genes to susceptibility to HIV type 1 infection in the Finnish population. AIDS Res Hum Retroviruses *14*, 695-8.

Pattanajitvilai, S., Kuroki, Y., Tsunezawa, W., McCormack, F. X., and Voelker, D. R. (1998). Mutational analysis of Arg197 of rat surfactant protein A. His197 creates specific lipid uptake defects. J Biol Chem *273*, 5702-7.

Patthy, L. (1987). Intron-dependent evolution: preferred types of exons and introns. FEBS Lett *214*, 1-7.

Patthy, L. (1991). Homology of the precursor of pulmonary surfactant-associated protein SP-B with prosaposin and sulfated glycoprotein 1. J Biol Chem *266*, 6035-7.

Perez Gil, J., Cruz, A., and Casals, C. (1993). Solubility of hydrophobic surfactant proteins in organic solvent/water mixtures. Structural studies on SP-B and SP-C in aqueous organic solvents and lipids. Biochim Biophys Acta *1168*, 261-70.

Persson, A., Chang, D., and Crouch, E. (1990). Surfactant protein D is a divalent cation-dependent carbohydrate-binding protein. J Biol Chem *265*, 5755-60.

Persson, A. V., Gibbons, B. J., Shoemaker, J. D., Moxley, M. A., and Longmore, W. J. (1992). The major glycolipid recognized by SP-D in surfactant is phosphatidylinositol. Biochemistry *31*, 12183-9.

Petry, F., McClive, P. J., Botto, M., Morley, B. J., Morahan, G., and Loos, M. (1996). The mouse C1q genes are clustered on chromosome 4 and show conservation of gene organization. Immunogenetics *43*, 370-6.

Phelps, D. S., and Floros, J. (1991). Localization of pulmonary surfactant proteins using immunohistochemistry and tissue in situ hybridization. Exp Lung Res *17*, 985-95.

Pikaar, J. C., Voorhout, W. F., van Golde, L. M., Verhoef, J., Van Strijp, J. A., and van Iwaarden, J. F. (1995). Opsonic activities of surfactant proteins A and D in phagocytosis of gram-negative bacteria by alveolar macrophages. J Infect Dis *172*, 481-9.

Pison, U., Wright, J. R., and Hawgood, S. (1992). Specific binding of surfactant apoprotein SP-A to rat alveolar macrophages. Am J Physiol *262*, L412-7.

Possmayer, F. (1988). A proposed nomenclature for pulmonary surfactant-associated proteins. Am Rev Respir Dis *138*, 990-8.

Possmayer, F., Yu, S. H., Weber, J. M., and Harding, P. G. (1984). Pulmonary surfactant. Can J Biochem Cell Biol *62*, 1121-33.

Postle, A. D., Mander, A., Reid, K. B., Wang, J. Y., Wright, S. M., Moustaki, M., and Warner, J. O. (1999). Deficient hydrophilic lung surfactant proteins A and D with normal surfactant phospholipid molecular species in cystic fibrosis. Am J Respir Cell Mol Biol *20*, 90-8.

Pryhuber, G. S., Khalak, R., and Zhao, Q. (1998). Regulation of surfactant proteins A and B by TNF-alpha and phorbol ester independent of NF-kappa B. Am J Physiol *274*, L289-95.

Qanbar, R., Cheng, S., Possmayer, F., and Schurch, S. (1996). Role of the palmitoylation of surfactant-associated protein C in surfactant film formation and stability. Am J Physiol *271*, L572-80.

Raftos, D., Green, P., Mahajan, D., Nair, S., Pearce, S., and Hutchinson, A. (1997). An opsonic collectin-like protein from tunicates. Dev Comp Immunol *21*, 89 (Abstr.).

Ravetch, J. V., and Clynes, R. A. (1998). Divergent roles for Fc receptors and complement in vivo. Annu Rev Immunol *16*, 421-32.

Reading, P. C., Allison, J., Crouch, E. C., and Anders, E. M. (1998). Increased susceptibility of diabetic mice to influenza virus infection: compromise of collectin-mediated host defense of the lung by glucose? J Virol *72*, 6884-7.

Reid, K. B. M., and Law, S. K. A. (1995). Complement, 2nd Edition, D. Male, ed. (Oxford: IRL Press).

Rettenberger, G., Klett, C., Zechner, U., Kunz, J., Vogel, W., and Hameister, H. (1995). Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. Genomics *26*, 372-8.

Rogde, S., Olaisen, B., Gedde Dahl, T., Jr., and Teisberg, P. (1986). The C8A and C8B loci are closely linked on chromosome 1. Ann Hum Genet *50*, 139-44.

Rogers, J. (1985). Exon shuffling and intron insertion in serine protease genes [news]. Nature *315*, 458-9.

Rooney, S. A., Young, S. L., and Mendelson, C. R. (1994). Molecular and cellular processing of lung surfactant. Faseb J *8*, 957-67.

Rosseau, S., Guenther, A., Seeger, W., and Lohmeyer, J. (1997). Phagocytosis of viable Candida albicans by alveolar macrophages: lack of opsonin function of surfactant protein A. J Infect Dis *175*, 421-8.

Ruano, M. L., Miguel, E., Perez Gil, J., and Casals, C. (1996). Comparison of lipid aggregation and self-aggregation activities of pulmonary surfactant-associated protein A. Biochem J *313*, 683-9.

Ruano, M. L., Perez Gil, J., and Casals, C. (1998). Effect of acidic pH on the structure and lipid binding properties of porcine surfactant protein A. Potential role of acidification along its exocytic pathway. J Biol Chem *273*, 15183-91.

Rubio, S., Lacaze Masmonteil, T., Chailley Heu, B., Kahn, A., Bourbon, J. R., and Ducroc, R. (1995). Pulmonary surfactant protein A (SP-A) is expressed by epithelial cells of small and large intestine. J Biol Chem *270*, 12162-9.

Rust, K., Bingle, L., Mariencheck, W., Persson, A., and Crouch, E. C. (1996). Characterization of the human surfactant protein D promoter: transcriptional regulation of SP-D gene expression by glucocorticoids. Am J Respir Cell Mol Biol *14*, 121-30.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science *239*, 487-491.

Saitoh, H., Okayama, H., Shimura, S., Fushimi, T., Masuda, T., and Shirato, K. (1998). Surfactant protein A2 gene expression by human airway submucosal gland cells. Am J Respir Cell Mol Biol *19*, 202-9.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: A Laboratory Manual, Second Edition (New York: Cold Spring Harbor Laboratory Press).

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J., and Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol *143*, 161-78.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A *74*, 5463-7.

Sano, H., Sohma, H., Muta, T., Nomura, S., Voelker, D. R., and Kuroki, Y. (1999). Pulmonary surfactant protein A modulates the cellular response to smooth and rough lipopolysaccharides by interaction with CD14. J Immunol *163*, 387-95.

Sastry, K., Herman, G. A., Day, L., Deignan, E., Bruns, G., Morton, C. C., and Ezekowitz, R. A. (1989). The human mannose-binding protein gene. Exon structure reveals its evolutionary relationship to a human pulmonary surfactant gene and localization to chromosome 10. J Exp Med *170*, 1175-89.

Sastry, R., Wang, J. S., Brown, D. C., Ezekowitz, R. A., Tauber, A. I., and Sastry, K. N. (1995). Characterization of murine mannose-binding protein genes Mbl1 and Mbl2 reveals features common to other collectin genes. Mamm Genome *6*, 103-10.

Sato, T., Endo, Y., Matsushita, M., and Fujita, T. (1994). Molecular characterization of a novel serine protease involved in activation of the complement system by mannose-binding protein. Int Immunol *6*, 665-9.

Scavo, L. M., Ertsey, R., and Gao, B. Q. (1998). Human surfactant proteins A1 and A2 are differentially regulated during development and by soluble factors. Am J Physiol *275*, L653-69.

Schelenz, S., Malhotra, R., Sim, R. B., Holmskov, U., and Bancroft, G. J. (1995). Binding of host collectins to the pathogenic yeast Cryptococcus neoformans: human surfactant protein D acts as an agglutinin for acapsular yeast cells. Infect Immun *63*, 3360-6.

Schots, A., Van Der Leede, B. J., De Jongh, E., and Egberts, E. (1988). A method for the determination of antibody affinity using a direct ELISA. Journal Of Immunological Methods *109*, 225-234.

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

Schurch, S. (1982). Surface tension at low lung volumes: dependence on time and alveolar size. Respir Physiol *48*, 339-55.

Schwaeble, W. J., and Reid, K. B. (1999). Does properdin crosslink the cellular and the humoral immune response? Immunol Today *20*, 17-21.

Sears, L. E., Moran, L. S., Kissinger, C., Creasey, T., Perry, O. K. H., Roskey, M., Sutherland, E., and Slatko, B. E. (1992). CircumVent thermal cycle sequencing and alternative manual and automated DNA sequencing protocols using the highly thermostable VentR (exo-) DNA polymerase. Biotechniques *13*, 626-33.

Sellar, G. C., Blake, D. J., and Reid, K. B. (1991). Characterization and organization of the genes encoding the A-, B- and C-chains of human complement subcomponent C1q. The complete derived amino acid sequence of human C1q. Biochem J *274*, 481-90.

Sheriff, S., Chang, C. Y., and Ezekowitz, R. A. (1994). Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil [published erratum appears in Nat Struct Biol 1996 Jan;3(1):103]. Nat Struct Biol *1*, 789-94.

Smit, A. F. (1993). Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res *21*, 1863-72.

Smith, S. L. (1999). Shark complement: an assessment. Immunol Rev *166*, 67-78.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol *98*, 503-17.

Spragg, R. G. (1998). Surfactant therapy in acute respiratory distress syndrome. Biol Neonate *74 Suppl 1*, 15-20.

Stamme, C., and Wright, J. R. (1999). Surfactant protein A enhances the binding and deacylation of E. coli LPS by alveolar macrophages. Am J Physiol *276*, L540-7.

Storgaard, P., Holm Nielsen, E., Skriver, E., Andersen, O., and Svehag, S. E. (1995). Mannan-binding protein forms complexes with alpha-2-macroglobulin. A protein model for the interaction. Scand J Immunol *42*, 373-80.

Stover, C. M., Schwaeble, W. J., Lynch, N. J., Thiel, S., and Speicher, M. R. (1999a). 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*, 148-9.

Stover, C. M., Thiel, S., Lynch, N. J., and Schwaeble, W. J. (1999b). The rat and mouse homologues of MASP-2 and MAp19, components of the lectin activation pathway of complement. J Immunol *163*, 6848-59.

Stover, C. M., Thiel, S., Thelen, M., Lynch, N. J., Vorup Jensen, T., Jensenius, J. C., and Schwaeble, W. J. (1999c). 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*, 3481-90.

Stubbs, L., Carver, E. A., Shannon, M. E., Kim, J., Geisler, J., Generoso, E. E., Stanford, B. G., Dunn, W. C., Mohrenweiser, H., Zimmermann, W., Watt, S. M., and Ashworth, L. K. (1996). Detailed comparative map of human chromosome 19q and related regions of the mouse genome. Genomics *35*, 499-508.

Suh, E., Chen, L., Taylor, J., and Traber, P. G. (1994). A homeodomain protein related to caudal regulates intestine-specific gene transcription. Mol Cell Biol *14*, 7340-51.

Sullivan, L. C., Daniels, C. B., Phillips , I. D., Orgeig, S., and Whitsett, J. A. (1998). Conservation of Surfactant Protein A: Evidence for a Single Origin for Vertebrate Pulmonary Surfactant. J Mol Evol. *46*, 131-8.

Sumiya, M., Super, M., Tabona, P., Levinsky, R. J., Arai, T., Turner, M. W., and Summerfield, J. A. (1991). Molecular basis of opsonic defect in immunodeficient children. Lancet *337*, 1569-70.

Summerfield, J. A., Ryder, S., Sumiya, M., Thursz, M., Gorchein, A., Monteil, M. A., and Turner, M. W. (1995). Mannose binding protein gene mutations associated with unusual and severe infections in adults [see comments]. Lancet *345*, 886-9.

Summerfield, J. A., Sumiya, M., Levin, M., and Turner, M. W. (1997). Association of mutations in mannose binding protein gene with childhood infection in consecutive hospital series [see comments]. Bmj *314*, 1229-32.

Sun, B. (1996). Use of surfactant in pulmonary disorders in full-term infants. Curr Opin Pediatr *8*, 113-7.

Sun, B., Curstedt, T., Lindgren, G., Franzen, B., Alaiya, A. A., Calkovska, A., and Robertson, B. (1997). Biophysical and physiological properties of a modified porcine surfactant enriched with surfactant protein A. Eur Respir J *10*, 1967-74.

Super, M., Thiel, S., Lu, J., Levinsky, R. J., and Turner, M. W. (1989). Association of low levels of mannan-binding protein with a common defect of opsonisation. Lancet *2*, 1236-9.

Suzuki, Y., Fujita, Y., and Kogishi, K. (1989). Reconstitution of tubular myelin from synthetic lipids and proteins associated with pig pulmonary surfactant. Am Rev Respir Dis *140*, 75-81.

Tabona, P., Mellor, A., and Summerfield, J. A. (1995). Mannose binding protein is involved in first-line host defence: evidence from transgenic mice. Immunology *85*, 153-9.

Takada, F., Seki, N., Matsuda, Y., Takayama, Y., and Kawakami, M. (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*, 757-9.

Takada, F., Takayama, Y., Hatsuse, H., and Kawakami, M. (1993). A new member of the C1s family of complement proteins found in a bactericidal factor, Ra-reactive factor, in human serum. Biochem Biophys Res Commun *196*, 1003-9.

Takahashi, A., Takayama, Y., Hatsuse, H., and Kawakami, M. (1993). Presence of a serine protease in the complement-activating component of the complement-dependent bactericidal factor, RaRF, in mouse serum. Biochem Biophys Res Commun *190*, 681-7.

Takahashi, M., Endo, Y., Fujita, T., and Matsushita, M. (1999). A truncated form of mannose-binding lectin-associated serine protease (MASP)-2 expressed by alternative polyadenylation is a component of the lectin complement pathway. Int Immunol *11*, 859-63.

Takahashi, N., and Ko, M. S. (1993). The short 3'-end region of complementary DNAs as PCR-based polymorphic markers for an expression map of the mouse genome. Genomics *16*, 161-8.

Takayama, Y., Takada, F., Takahashi, A., and Kawakami, M. (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*, 2308-16.

Taneva, S., Voelker, D. R., and Keough, K. M. (1997). Adsorption of pulmonary surfactant protein D to phospholipid monolayers at the air-water interface. Biochemistry *36*, 8173-9.

Tashiro, K., Nishizuka, K., Matsumoto, Y., Ohta, K., Suzuki, Y., and Kobayashi, T. (1999). Modified natural and synthetically reconstituted surfactant therapies for acute lung injury caused by endotoxin in rats. Acta Anaesthesiol Scand *43*, 821-8.

Tenner, A. J. (1999). Membrane receptors for soluble defense collagens. Curr Opin Immunol *11*, 34-41.

Tenner, A. J., Robinson, S. L., Borchelt, J., and Wright, J. R. (1989). Human pulmonary surfactant protein (SP-A), a protein structurally homologous to C1q, can enhance FcR- and CR1-mediated phagocytosis. J Biol Chem *264*, 13923-8.

Terai, I., Kobayashi, K., Matsushita, M., and Fujita, T. (1997). Human serum mannose-binding lectin (MBL)-associated serine protease-1 (MASP-1): determination of levels in body fluids and identification of two forms in serum. Clin Exp Immunol *110*, 317-23.

Terai, I., Kobayashi, K., Matsushita, M., Fujita, T., and Matsuno, K. (1995). alpha 2-Macroglobulin binds to and inhibits mannose-binding protein-associated serine protease. Int Immunol *7*, 1579-84.

Thiel, S., Holmskov, U., Hviid, L., Laursen, S. B., and Jensenius, J. C. (1992). The concentration of the C-type lectin, mannan-binding protein, in human plasma increases during an acute phase response. Clin Exp Immunol *90*, 31-5.

Thiel, S., Vorup Jensen, T., Stover, C. M., Schwaeble, W., Laursen, S. B., Poulsen, K., Willis, A. C., Eggleton, P., Hansen, S., Holmskov, U., Reid, K. B., and Jensenius, J. C. (1997). A second serine protease associated with mannan-binding lectin that activates complement. Nature *386*, 506-10.

Tino, M. J., and Wright, J. R. (1996). Surfactant protein A stimulates phagocytosis of specific pulmonary pathogens by alveolar macrophages. Am J Physiol *270*, L677-88.

Tino, M. J., and Wright, J. R. (1999a). Glycoprotein-340 binds surfactant protein-A (SP-A) and stimulates alveolar macrophage migration in an SP-A-independent manner. Am J Respir Cell Mol Biol *20*, 759-68.

Tino, M. J., and Wright, J. R. (1999b). Surfactant proteins A and D specifically stimulate directed actin-based responses in alveolar macrophages. Am J Physiol *276*, L164-74.

Toonen, R. F., Gowan, S., and Bingle, C. D. (1996). The lung enriched transcription factor TTF-1 and the ubiquitously expressed proteins Sp1 and Sp3 interact with elements located in the minimal promoter of the rat Clara cell secretory protein gene. Biochem J *316*, 467-73.

Tornetta, M. A., Foley, J. J., Sarau, H. M., and Ames, R. S. (1997). The mouse anaphylatoxin C3a receptor: molecular cloning, genomic organization, and functional expression. J Immunol *158*, 5277-82.

Tosi, M., Duponchel, C., Meo, T., and Couture Tosi, E. (1989). Complement genes C1r and C1s feature an intronless serine protease domain closely related to haptoglobin. J Mol Biol *208*, 709-14.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A *76*, 4350-4.

Turner, M. W. (1996). Mannose-binding lectin: the pluripotent molecule of the innate immune system. Immunol Today *17*, 532-40.

Van Eijk, M., Haagsman, H. P., Skinner, T., Archibald, A., Reid, K. B. M., and Lawson, P. R. (2000). Porcine lung surfactant protein D: Complementary DNA cloning, chromosomal localization and tissue distribution. J Immunol *164*, 1442-50.

van Emmerik, L. C., Kuijper, E. J., Fijen, C. A., Dankert, J., and Thiel, S. (1994). Binding of mannan-binding protein to various bacterial pathogens of meningitis. Clin Exp Immunol *97*, 411-6.

van Iwaarden, F., Welmers, B., Verhoef, J., Haagsman, H. P., and van Golde, L. M. (1990). Pulmonary surfactant protein A enhances the host-defense mechanism of rat alveolar macrophages. Am J Respir Cell Mol Biol *2*, 91-8.

Van Iwaarden, J. F., Pikaar, J. C., Storm, J., Brouwer, E., Verhoef, J., Oosting, R. S., van Golde, L. M., and van Strijp, J. A. (1994). Binding of surfactant protein A to the lipid A moiety of bacterial lipopolysaccharides. Biochem J *303*, 407-11.

Van Iwaarden, J. F., Shimizu, H., Van Golde, P. H., Voelker, D. R., and Van Golde, L. M. (1992). Rat surfactant protein D enhances the production of oxygen radicals by rat alveolar macrophages. Biochem J *286*, 5-8.

van Iwaarden, J. F., Teding van Berkhout, F., Whitsett, J. A., Oosting, R. S., and van Golde, L. M. (1995). A novel procedure for the rapid isolation of surfactant protein A with retention of its alveolar-macrophage-stimulating properties. Biochem J *309*, 551-5.

van Iwaarden, J. F., van Strijp, J. A., Ebskamp, M. J., Welmers, A. C., Verhoef, J., and van Golde, L. M. (1991). Surfactant protein A is opsonin in phagocytosis of herpes simplex virus type 1 by rat alveolar macrophages. Am J Physiol *261*, L204-9.

van Iwaarden, J. F., van Strijp, J. A., Visser, H., Haagsman, H. P., Verhoef, J., and van Golde, L. M. (1992). Binding of surfactant protein A (SP-A) to herpes simplex virus type 1-infected cells is mediated by the carbohydrate moiety of SP-A. J Biol Chem *267*, 25039-43.

Van Reeth, K., Nauwynck, H., and Pensaert, M. (1998). Bronchoalveolar interferon-alpha, tumor necrosis factor-alpha, interleukin-1, and inflammation during acute influenza in pigs: a possible model for humans? J Infect Dis *177*, 1076-9.

Vandenbussche, G., Clercx, A., Curstedt, T., Johansson, J., Jornvall, H., and Ruysschaert, J. M. (1992). Structure and orientation of the surfactant-associated protein C in a lipid bilayer. Eur J Biochem *203*, 201-9.

Veldhuizen, R., Nag, K., Orgeig, S., and Possmayer, F. (1998). The role of lipids in pulmonary surfactant. Biochim Biophys Acta *1408*, 90-108.

Vorup Jensen, T., Davis, S. J., Poulsen, K., Schwaeble, W., Stover, C. M., Sim, R., Reid, K. B. M., Peterson, S. V., Thiel, S., and Jensenius, J. C. (1998). Studies of recombinant MBL and MASP-2: Is MASP-2 self-activating? Mol Immunol *35*, 400 (Abstract).

Voss, T., Melchers, K., Scheirle, G., and Schafer, K. P. (1991). Structural comparison of recombinant pulmonary surfactant protein SP-A derived from two human coding sequences: implications for the chain composition of natural human SP-A. Am J Respir Cell Mol Biol *4*, 88-94.

Wada, M., Ozaki, K., Itoh, N., Yamashina, I., and Kawasaki, T. (1990). Two forms of messenger RNA encoding rat liver mannan-binding protein are generated by differential utilization of polyadenylation sites of one transcript. J Biochem Tokyo *108*, 914-7.

Walenkamp, A. M., Verheul, A. F., Scharringa, J., and Hoepelman, I. M. (1999). Pulmonary surfactant protein A binds to Cryptococcus neoformans without promoting phagocytosis. Eur J Clin Invest *29*, 83-92.

Wallis, R., and Drickamer, K. (1999). Molecular determinants of oligomer formation and complement fixation in mannose-binding proteins. J Biol Chem *274*, 3580-9.

Wang, J. Y., Kishore, U., Lim, B. L., Strong, P., and Reid, K. B. (1996a). Interaction of human lung surfactant proteins A and D with mite (Dermatophagoides pteronyssinus) allergens. Clin Exp Immunol *106*, 367-73.

Wang, J. Y., Shieh, C. C., You, P. F., Lei, H. Y., and Reid, K. B. (1998). Inhibitory effect of pulmonary surfactant proteins A and D on allergen-induced lymphocyte proliferation and histamine release in children with asthma. Am J Respir Crit Care Med *158*, 510-8.

Wang, Z., Hall, S. B., and Notter, R. H. (1996b). Roles of different hydrophobic constituents in the adsorption of pulmonary surfactant. J Lipid Res *37*, 790-8.

Warr, R. G., Hawgood, S., Buckley, D. I., Crisp, T. M., Schilling, J., Benson, B. J., Ballard, P. L., Clements, J. A., and White, R. T. (1987). Low molecular weight human pulmonary surfactant protein (SP5): isolation, characterization, and cDNA and amino acid sequences. Proc Natl Acad Sci U S A *84*, 7915-9.

Weaver, T. E. (1991). Surfactant proteins and SP-D. Am J Respir Cell Mol Biol *5*, 4-5.

Weaver, T. E., Lin, S., Bogucki, B., and Dey, C. (1992). Processing of surfactant protein B proprotein by a cathepsin D-like protease. Am J Physiol *263*, L95-103.

Weaver, T. E., and Whitsett, J. A. (1989). Processing of hydrophobic pulmonary surfactant protein B in rat type II cells. Am J Physiol *257*, L100-8.

Weikert, L. F., Edwards, K., Chroneos, Z. C., Hager, C., Hoffman, L., and Shepherd, V. L. (1997). SP-A enhances uptake of bacillus Calmette-Guerin by macrophages through a specific SP-A receptor. Am J Physiol *272*, L989-95.

Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem *55*, 631-61.

Weis, W. I., and Drickamer, K. (1994). Trimeric structure of a C-type mannose-binding protein. Structure *2*, 1227-40.

Weis, W. I., Drickamer, K., and Hendrickson, W. A. (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. Nature *360*, 127-34.

Weissbach, S., Neuendank, A., Pettersson, M., Schaberg, T., and Pison, U. (1994). Surfactant protein A modulates release of reactive oxygen species from alveolar macrophages. Am J Physiol *267*, L660-6.

Whitsett, J. A., and Glasser, S. W. (1998). Regulation of surfactant protein gene transcription. Biochim Biophys Acta *1408*, 303-11.

Williams, M. C., Hawgood, S., and Hamilton, R. L. (1991). Changes in lipid structure produced by surfactant proteins SP-A, SP-B, and SP-C. Am J Respir Cell Mol Biol *5*, 41-50.

Williams, M. D., Wright, J. R., March, K. L., and Martin, W. J. n. (1996). Human surfactant protein A enhances attachment of Pneumocystis carinii to rat alveolar macrophages. Am J Respir Cell Mol Biol *14*, 232-8.

Wintergerst, E., Manz Keinke, H., Plattner, H., and Schlepper Schafer, J. (1989). The interaction of a lung surfactant protein (SP-A) with macrophages is mannose dependent. Eur J Cell Biol *50*, 291-8.

Wong, C. J., Akiyama, J., Allen, L., and Hawgood, S. (1996). Localization and developmental expression of surfactant proteins D and A in the respiratory tract of the mouse. Pediatr Res *39*, 930-7.

Wong, N. K. H., Dobo, J., and Sim, R. B. (1998). Interaction of MBL-associated serine proteases (MASPs) with sunthetic and natural substrates and inhibitors. Mol Immunol *35*, 375 (Abstract).

Wright, J. R. (1997). Immunomodulatory functions of surfactant. Physiol Rev. *77*, 931-962.

Wright, J. R., and Youmans, D. C. (1993). Pulmonary surfactant protein A stimulates chemotaxis of alveolar macrophage. Am J Physiol *264*, L338-44.

Yan, C., Sever, Z., and Whitsett, J. A. (1995). Upstream enhancer activity in the human surfactant protein B gene is mediated by thyroid transcription factor 1. J Biol Chem *270*, 24852-7.

Yerle, M., Echard, G., Robic, A., Mairal, A., Dubut Fontana, C., Riquet, J., Pinton, P., Milan, D., Lahbib Mansais, Y., and Gellin, J. (1996). A somatic cell hybrid panel for pig regional gene mapping characterized by molecular cytogenetics. Cytogenet Cell Genet *73*, 194-202.

Yu, S. H., and Possmayer, F. (1996). Effect of pulmonary surfactant protein A and neutral lipid on accretion and organization of dipalmitoylphosphatidylcholine in surface films. J Lipid Res *37*, 1278-88.

Yu, S. H., and Possmayer, F. (1998). Interaction of pulmonary surfactant protein A with dipalmitoylphosphatidylcholine and cholesterol at the air/water interface. J Lipid Res *39*, 555-68.

Zhang, M., Damodarasamy, M., Elhalwagi, B. M., and McCormack, F. X. (1998). The longer isoform and Cys-1 disulfide bridge of rat surfactant protein A are not essential for phospholipid and type II cell interactions. Biochemistry *37*, 16481-8.

Zimmerman, P. E., Voelker, D. R., McCormack, F. X., Paulsrud, J. R., and Martin, W. J. d. (1992). 120-kD surface glycoprotein of Pneumocystis carinii is a ligand for surfactant protein A. J Clin Invest *89*, 143-9.

Appendix 1: Primers Used to Characterise the Murine SP-D Gene.

Primer Name	Sequence
CA1f	CAGAGTGTAGGTGGAGTAAG
CA1r	GGCTATTTATGCCTGCCCTG
CA2f	TACACAGTAACTAATATTATCATG
CA2r	AACATCTGATATCAACCTCTAG
mAf	GGAATTCTAGATCAAAATTCACAAATGGCCAG
mAnif	ACAGTAGGTCGCTTTCTGGGC
mAnlr	GGAATTCTAGATCAAAATTCACAAATGGCCAG
mAr	GGAATTCATATGGGGCTTCCAGGGTTTCCA
mD1f	GCTGCCCTTTCTCTCCAT
mD1r	CACCCTTCTCACCCCGT
mD2f	GGCTCTCAGGGTTGCAG
mD2r	CACGTTCTCCCTTTGGTC
mD4r	CAGCATTCCCTGGGGCT
mD5f	GTCCACAGGGAGCTCCA
mD5r	CGCTTTCACCTTTGATTCC
mD6f	CTGAGGCAGCAGATGGAG
mD6r	GTGGGAGAAGGCAACCTC
mD7f	GATGGCCGAAGTGTTGGAG
MD7r	CAGAACTCACAGATAACAAG
MDcon-f	CCGGAATTCCTGGAAGCAGAAATGAAGA
MDcon-r	CCGGAATTCAAGATCTCCACACAGTCCTC
mDn1-f	GGAATTCAGAACTCACAGATAACAAG
mDn1-r	GGAATTCGGATCCAAAGGTGAAAGCGGGCTT
mMBL1f	GCCATTGAGGAGAAGCTGGCAAATATG
mMBL1r	CCGGAATTCGAATATGGAGGCACTCGTTTCCTC
pAf	AATGATAAGGCTTGTGGAGAG
pAr	GGATGTTGGCAACATACAGG
XSC	GACTCGAGTCGACATCGAT
$XSCdG_{17}$	$GACTCGAGTCGACATCGATG_{17}H$[1]

[1]H represents the single IUB code that indicates the nucleotides; A, T and C. This extra nucleotide in $XSCdG_{17}$ acts to anchor the oligonucleotide at the junction of a polyC tract and normal sequence.

Appendix 2: The Sequence of the 5' clone for the Murine SP-D Gene.

```
LOCUS       MMSPD1       6090 bp    DNA             ROD           20-MAY-1999
DEFINITION  Mus musculus surfactant protein D precursor (Sftp4) gene, exons 1
            and 2.
ACCESSION   AF047741
NID         g4877553
VERSION     AF047741.1  GI:4877553
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 6090)
  AUTHORS   Lawson,P.R., Perkins,V.C., Holmskov,U. and Reid,K.B.M.
  TITLE     Genomic organisation of the mouse gene for surfactant protein D
            (sftp4)
  JOURNAL   Am. J. Respir. Cell Mol. Biol. 20 (5), 953-963 (1999)
REFERENCE   2  (bases 1 to 6090)
  AUTHORS   Lawson,P.R., Perkins,V.C., Holmskov,U. and Reid,K.B.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1998) Department of Biochemistry, MRC
            Immunochemistry Unit, South Parks Road, Oxford University, Oxford,
            Oxfordshire OX1 3QU, England
FEATURES             Location/Qualifiers
     source          1..6090
                     /organism="Mus musculus"
                     /strain="129/SV"
                     /sub_strain="PAS"
                     /db_xref="taxon:10090"
                     /chromosome="14"
                     /map="between D14Mit212 and D14Mit141"
                     /cell_line="D3"
                     /cell_type="embryonal stem"
                     /dev_stage="embryo"
  promoter           1..3671
  repeat_region      1..44
                     /note="SINE"
                     /rpt_family="B2 (B3A)"
                     /rpt_type=dispersed
  repeat_region      172..247
                     /note="SINE"
                     /rpt_family="ID (ID3)"
                     /rpt_type=dispersed
  repeat_region      252..320
                     /note="SINE"
                     /rpt_family="B1"
                     /rpt_type=dispersed
  repeat_region      487..668
                     /note="LINE"
                     /rpt_family="L1 (L1ME1)"
                     /rpt_type=dispersed
  repeat_region      1071..1538
                     /note="LINE"
                     /rpt_family="L1 (L1M4)"
                     /rpt_type=dispersed
  repeat_region      1679..1863
                     /note="LINE"
                     /rpt_family="L1 (L1MC4)"
                     /rpt_type=dispersed
  repeat_region      3551..3592
                     /note="polymorphic microsatellite used to genetically map
                     SP-D"
                     /rpt_type=tandem
                     /rpt_unit=ca
  TATA_signal        3639..3646
  mRNA               join(3672..3710,5702..5900,AF047742.1:312..428,
                     AF047742.1:669..785,AF047742.1:1112..1228,
                     AF047742.1:1977..2093,AF047742.1:3162..3245,
```

```
                     AF047742.1:5010..5507)
                     /gene="Sftp4"
                     /product="surfactant protein D precursor"
     5'UTR           join(3672..3710,5702..5704)
                     /gene="Sftp4"
                     /evidence=experimental
     exon            3672..3710
                     /gene="Sftp4"
                     /number=1
     gene            order(3672..6090,AF047742.1:1..5507)
                     /gene="Sftp4"
     exon            5702..5900
                     /gene="Sftp4"
                     /number=2
     sig_peptide     5705..5761
                     /gene="Sftp4"
     CDS             join(5705..5900,AF047742.1:312..428,AF047742.1:669..785,
                     AF047742.1:1112..1228,AF047742.1:1977..2093,
                     AF047742.1:3162..3245,AF047742.1:5010..5386)
                     /gene="Sftp4"
                     /codon_start=1
                     /product="surfactant protein D precursor"
                     /protein_id="AAD31380.1"
                     /db_xref="PID:g4877556"
                     /db_xref="GI:4877556"
                     /translation="MLPFLSMLVLLVQPLGNLGAEMKSLSQRSVPNTCTLVMCSPTEN
                     GLPGRDGRDGREGPRGEKGDPGLPGPMGLSGLQGPTGPVGPKGENGSAGEPGPKGERG
                     LSGPPGLPGIPGPAGKEGPSGKQGNIGPQGKPGPKGEAGPKGEVGAPGMQGSTGAKGS
                     TGPKGERGAPGVQGAPGNAGAAGPAGPAGPQGAPGSRGPPGLKGDRGVPGDRGIKGES
                     GLPDSAALRQQMEALKGKLQRLEVAFSHYQKAALFPDGRSVGDKIFRTADSEKPFEDA
                     QEMCKQAGGQLASPRSATENAAIQQLITAHNKAAFLSMTDVGTEGKFTYPTGEPLVYS
                     NWAPGEPNNNGGAENCVEIFTNGQWNDKACGEQRLVICEF"
BASE COUNT     1672 a   1249 c   1283 g   1886 t    ORIGIN
        1 gatcaacaag tttgtcttaa ccactgagct atctctccag accctaactg cacctttag
       61 acaaggggtg tggggtgctg agcttgtaag atgagctttg aaacgggccc cttgggtaat
      121 acacaatgat tattacgaaa taaagtcata aaaatcaaat ctatagttac attttattta
      181 tttttgtaat gttaagaacc caagactcag tgatgctagg caagcactct atcgatgagt
      241 atatctcagc attttgagac agtcttgcta tgtagcccag gctagcctaa accttatgat
      301 cttcctgcct tctccttctg taaccatcat aaccttggca atccctcttc agtaatttcc
      361 tccagccact gaggcccaga cccactgctc tgctttctgt tagaacacag ggcttctcat
      421 ttccagaact tcaagtaatt atgcagaata tcctgttact ttttgttatt tttatttaat
      481 taaaaaataa tgggattcat ttggcagcat gttttttgtt tgtttgtttg ttttactctg
      541 cctgtttttg agattattca tgctgtgcct gtcaatggtt tatccctctt gtcactgctg
      601 agcaacattt cattgttcct atatatgaca atttatccat tcatggtttg tggcttgagt
      661 tgtttccaga atgggcaatt aatacagcta aaaaaaaaca ttgtataatc catctcaagt
      721 taatgtttgt ggggtaaata aggttgaaat gatttttttt gcagccattt cccatatggt
      781 gttccagtac caatcactaa agtatcttcc caccttgag taagatcaag tgaatgatac
      841 acatttgagg tctgttcagg aaagaaaaag aaatacatac atacaaaaaa atcagtagta
      901 catatcttct aaattatgtt ctagatagtc tatattttat actaaattat atattatata
      961 caatactact tatattatgt aatataatat cttatatatt atactattta atatattaat
     1021 catgatgtta taaatatgca tactatataa tattatatgt attatacatt ttatttttata
     1081 ttatatttat tatagattgt ctgtaaaatt catattttaa tatcaatatt agagtcaagt
     1141 tagtttctat ataatgacta ctaaagtatt gataggtact gaattgaata tgtagatcaa
     1201 ttcagagaaa tttgaattaa caatatcaag tcctttatta cataaacata cctacttaac
     1261 atttaaaatt tgcatggatg ttttatagga ttccagtgtg aatatcttga ctatatttta
     1321 aattacttag aagtagtttt catttggctg ctttgtgcca ttgccaaatt tcttgttatt
     1381 aaatgtgtgt atcatgtgga aatgcaattg ctttttgtat gttgatattg cttcctgtaa
     1441 cctcagaaaa ctgctagctg aagctgctct atgcatgtcc ttcaggactt cctactctta
     1501 cgtgccagtg atgggtaaag acagtttaat ttctttctaa tatgacatat tttcctatct
     1561 tcatgtctta agtagtggtc ttcagtagag tggcaaggga cactgataac cagtctcagg
     1621 gaaggaagca ttctttcttc aacatgtggt gctagctcaa gggtttctat aatgtcctac
     1681 agccacagta cgatgaccta gttcaggaaa ttaacctggg ctcagtgcta tttgggaatc
     1741 caaggcctta cttaaattct gcaagtttcc actctcatgc tgtttctggt ctagggttct
     1801 gtccatatct cacattcaat ttagttgtta tgtttcctta gcatgctttt acttggtaca
     1861 gctcatgcct tggacctgaa tgatgcactg ttttgcattg ttcctcaaca tgaaattttg
     1921 ttgtaattat actgagttta tacgactctg gcaaaagaat cacacacgcg cgtgcgtgca
     1981 cacacacata cacacataaa attacttttt gctcctcttg atacattctg tcaagggctc
     2041 agatgtcaat ttgtcaactt ttatttttct tgacttaaat tttaattttc tgacatggaa
     2101 ccattgtgca cccatagata atatctactg ctgggataca taagtgagtg tgctatacaa
```

2161 cacatatgcc aacaaactgg agaaaagaac aaattagtta acaggtaaaa aaaatgaaac
2221 ctctgaccag ggtctcagag aaaatacaag cccagaccaa gacgccatct cataagggga
2281 gggcaacacc attctctgaa agtgttcaat agcttcctgg gcctgagtct ctgcacttat
2341 atatgaaggg aaagacatat tggtttgtta cagtgtgcac acattgcata gaatcacctc
2401 cagatcaata tggagtgtag ggccatcgac cttgtctgga ataactgtgg gaaaggttgg
2461 gttagttggt tcacacagcc atccatttct atgattaggc acttcagagg actaatgaca
2521 gccttagtgc tatgaatatg gtcatcattg tcttccctgt caactgtatt gactcatctt
2581 ccctcccaga gaggctggta tgtctgtccc tgctggccct atgagtatgt gtgttatgtt
2641 ccctttagct atttatgttg taatctcttg ttttcactta gagaaataag tatctaaatt
2701 gtctttcatc agaatgaaac ttgcttcgga agggaaatat tgtcttaatt ctaggaaaag
2761 gaggtaaagt caccttgacc ttgagatata tggagcttga atgggtgagc ataaacacaa
2821 tgggtggggg gcaatttggg aaggaaaaaa aaacctcaag tcacaggaat ctagaaacaa
2881 agattgtaac aatggtggaa agatagacaa cttcccaata aactctggaa attaagagct
2941 gcctgatggg ttttgatcca tcaagagttc aaggtgtctc tttaatgacc agggcaagat
3001 acaagtcaaa ttattgctag gaagatggaa gaacaagggc caacatgcca catgctaagg
3061 gaagtccatt ccatggataa taagtgatga gctctttgta ttttagagca gtattagtac
3121 tacaatcctc acaggttgct gcctttgtag cctccttatc tgccccaccc ctccttgcta
3181 tgccaagcag ctcccagctc tttctcaata tctctgcaat gcttttcaga tgaaaaagga
3241 agtggacaaa gcaatggatt ctacctttgg agatatgcaa atgtgagggt tttcttttc
3301 ttttttttt ttccttcaag cgttgtaggc tactgtgcaa aatgatgggc tctggacctc
3361 agattcttgg acatctcaaa tactcgggca ttgggaatgc ttcctcactc ctggagctgt
3421 agaaaggttg tagagctcac tgcctgacca tgcatagcta agggagcagg tgcacatagt
3481 gtcaggtata ttgaaggcag ttttccagag tgtaggtgga gtaagagtag gtgagtcaac
3541 aggaaaaaat tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgacttttct
3601 cagaagacca atgctgggta gctccactta cagggcaggc ataaatagcc gagcctgaca
3661 aacagaggtg catttacctg gacctcctgg ggccaacaag gaagcaatct gtgagtgcct
3721 ctgcttctca ttttatctct ctacctcccg tggagatggc tttgtccatc tgggctgttt
3781 atcccactga ttttgagaaa ggggaatagg attctggttc taggttctgg atcggggaca
3841 agggatttga ggggtggagg tgagggacac agtgagggaa aagaatccag tccctgtgtc
3901 accttccttg actcagagaa agatttggtt tctgagatgg gagtcgtgtt gatgcatgtt
3961 atgtgatgat tctatagacc tacacagatt gtatggagaa ttcaggtgaa ttgtggttga
4021 atctcaattt cataagaaaa ataatagggg aagaactccc tagccttccc taggctgggg
4081 ttttaatctc aggctagtgc cgaaagcata gaaacaagga caaaatgagt tctctcttag
4141 cctaaactct ttattaaaca ataaacttaa cagctcacgg cctctgattt gcgagccaac
4201 attacaaaca gcttctgttc tcaatgaact tttactctga aagcacggag tgccacacac
4261 actgtgggct gtatctctct gcacactcca tccactgccc cagctaatta gtctgtctca
4321 gatctggaag cctgtagatc tgatctctgc aggcttcatt aatgctaagg tgaacagatg
4381 gtagatcaga aggccttttt tcttgattct tccacagaat gaggcatggg gccccaaagt
4441 taaggtggct actggctaag gaacagtggc cttgctctga cttgggcctg ccctttagta
4501 ctgacgacag aacctctgct aaaacaatgg taaagaagta gcttgtgagc tcactaaggt
4561 atttctagga agttccaaga cacatgaggc aattgtaaat aagaagcaca aagctttaac
4621 ccagaatcac caggtggaga aacagcaggt caggaaacct gacaaacatg caatccaccc
4681 ccagctgtgc tgaaactctt ccctcgttat agaagcctct tttgctttgt cctcagccac
4741 cgtgctgcct atgggctctt tccttccttg agaggaaaga tgttataata attcaggtct
4801 tggccccctg tatcccatat cccagacgtg tatagttatc agccttgtgt cactattgga
4861 aaatgctggt catgagaagt taagaccttg ataccctgtg tagggagtta ccaaagtctg
4921 acatcctgca cgaggtctca gccggactat gtccctctca gacagtgact ggatggatac
4981 tctgctaaat gttttctctt ggttcctcct caaaccaacc tccaatctac ataagcccat
5041 ctggaatgtt cccgctcccg caacaaaact tttcccttgt gccctctcag attgtagact
5101 tctaggtatc tctgactctc ttcaactgcc aacttaatat ctctcaagct ctttatttgc
5161 attctctcta gtactttccg tatgaactct gcctggtcta ggaatgccaa acagccgggc
5221 aagcatgtgg ccgtactcca gcctgaactc agtacaagca tctagaagtg ctcagagcca
5281 ggttcatctc tggatcagta gtattcaatg tcggtgctaa tcccagaagc cagggagaga
5341 atgcagactc agcctcaaat gagttaagcc tgctatccta gaacatagaa ggaggaccag
5401 ggatgagaca gtcctcagtg atggacacta gactcagcca ctgttctcat ggggcaccag
5461 agaggaacac tacagtcaac ccgtgttaga gtgctagtca ggctactcac tccctcaaag
5521 gccagaccat gtacctacct ccataaactg cttttggagt cttactgtct tattgctctg
5581 ggtgttcatc atggggaacg atgtgggggа tgaaggtttc tcttctctgg acttagtagc
5641 cactacctag aacctcatcc tagaaccatc attgcctctt ttctctttct caccctgca
5701 ggacatgctg ccctttctct ccatgcttgt cttgcttgta cagcccctgg gaaatctggg
5761 agcagaaatg aagagcctct cgcagagatc agtacccaac acctgcaccc tagtcatgtg
5821 tagcccaaca gagaatggcc tgcctggtcg tgatggacgg gatgggagag aaggtccacg
5881 gggtgagaag ggtgatccag gtaagctggg aatagggggt tggaatgtga actaaactga
5941 tcgatccccc aggtgattag ggatcaactt ggatatgagc ttctttctcc tagggattga
6001 gtttcccccc ggtgtccatg tttcctgcca aacacatttt ttttttgttt tttggggttt
6061 ttttgtgggg gggggggggg agcgtgtttt

Appendix 3: The sequence of the 3' clone for the murine SP-D gene.

```
LOCUS       MMSPD2       5593 bp    DNA             ROD 20-MAY-1999
DEFINITION  Mus musculus surfactant protein D precursor (Sftp4) gene, exons 3-8
            and complete cds.
ACCESSION   AF047742
NID         g4877554
VERSION     AF047742.1  GI:4877554
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 5593)
  AUTHORS   Lawson,P.R., Perkins,V.C., Holmskov,U. and Reid,K.B.M.
  TITLE     Genomic organisation of the mouse gene for surfactant protein D
            (sftp4)
  JOURNAL   Am. J. Respir. Cell Mol. Biol. 20 (5), 953-963 (1999)
REFERENCE   2  (bases 1 to 5593)
  AUTHORS   Lawson,P.R., Perkins,V.C., Holmskov,U. and Reid,K.B.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1998) Department of Biochemistry, MRC
            Immunochemistry Unit, South Parks Road, Oxford University, Oxford,
            Oxfordshire OX1 3QU, England
FEATURES             Location/Qualifiers
     source          1..5593
                     /organism="Mus musculus"
                     /strain="129/SV"
                     /sub_strain="PAS"
                     /db_xref="taxon:10090"
                     /chromosome="14"
                     /map="between D14Mit212 and D14Mit141"
                     /cell_line="D3"
                     /cell_type="embryonal stem"
                     /dev_stage="embryo"
     exon            312..428
                     /gene="Sftp4"
                     /number=3
     exon            669..785
                     /gene="Sftp4"
                     /number=4
     exon            1112..1228
                     /gene="Sftp4"
                     /number=5
     repeat_region   1301..1668
                     /note="MaLR"
                     /rpt_type=dispersed
     repeat_region   1368..1408
                     /note="minisatellite"
                     /rpt_type=tandem
                     /rpt_unit=AAAC
     exon            1977..2093
                     /gene="Sftp4"
                     /number=6
     repeat_region   2538..2648
                     /note="SINE"
                     /rpt_family="B1 (PD1B10)"
                     /rpt_type=dispersed
     repeat_region   2654..2851
                     /note="SINE"
                     /rpt_family="B4 (B4A)"
                     /rpt_type=dispersed
     exon            3162..3245
                     /gene="Sftp4"
                     /number=7
     repeat_region   4242..4285
                     /note="microsatellite; polymorphic"
                     /rpt_type=tandem
                     /rpt_unit=ca
```

```
repeat_region   4266..4341
                /note="SINE"
                /rpt_family="B4 (B4A)"
                /rpt_type=dispersed
repeat_region   4389..4539
                /note="SINE"
                /rpt_family="B4 (RSINE1)"
                /rpt_type=dispersed
exon            5010..5507
                /gene="Sftp4"
                /number=8
repeat_region   5399..5420
                /note="microsatellite; polymorphic"
                /rpt_type=tandem
3'UTR           5484..5507
                /gene="Sftp4"
polyA_signal    5489..5494
                /gene="Sftp4"
mRNA            join(AF047741.1:3672..3710,AF047741.1:5702..5900,312..428,
                669..785,1112..1228,1977..2093,3162..3245,5010..5507)
                /gene="Sftp4"
                /product="surfactant protein D precursor"
gene            order(AF047741.1:3672..6090,1..5507)
                /gene="Sftp4"
mat_peptide     join(AF047741.1:5762..5900,312..428,669..785,1112..1228,
                1977..2093,3162..3245,5010..5383)
                /gene="Sftp4"
                /product="surfactant protein D"
CDS             join(AF047741.1:5705..5900,312..428,669..785,1112..1228,
                1977..2093,3162..3245,5010..5386)
                /gene="Sftp4"
                /codon_start=1
                /product="surfactant protein D precursor"
                /protein_id="AAD31380.1"
                /db_xref="PID:g4877556"
                /db_xref="GI:4877556"
                /translation="MLPFLSMLVLLVQPLGNLGAEMKSLSQRSVPNTCTLVMCSPTEN
                GLPGRDGRDGREGPRGEKGDPGLPGPMGLSGLQGPTGPVGPKGENGSAGEPGPKGERG
                LSGPPGLPGIPGPAGKEGPSGKQGNIGPQGKPGPKGEAGPKGEVGAPGMQGSTGAKGS
                TGPKGERGAPGVQGAPGNAGAAGPAGPAGPQGAPGSRGPPGLKGDRGVPGDRGIKGES
                GLPDSAALRQQMEALKGKLQRLEVAFSHYQKAALFPDGRSVGDKIFRTADSEKPFEDA
                QEMCKQAGGQLASPRSATENAAIQQLITAHNKAAFLSMTDVGTEGKFTYPTGEPLVYS
                NWAPGEPNNNGGAENCVEIFTNGQWNDKACGEQRLVICEF"
BASE COUNT     1550 a   1244 c   1333 g   1466 t          ORIGIN
        1 atgcacataa atctacactc ttgctactga tcaatcaatg ctggcagatt gccttctgct
       61 gatacagagg ccttggggca ggggaaatga aaactacatg aaaagttgtg aagctcttga
      121 cttgaattca tctccaagcc aagaatgtct attttgtcta agtggagaca ctaattttta
      181 tttagtgggt taccacagga ccaagccatg ctcagcagaa ccttaaaggg atgtcaggga
      241 gctcatcttg tagacatcag tctctcctga gcacgggact aactttgggt ccatttcttt
      301 tgtttctata ggtttgccag gacctatggg gctctcaggg ttgcagggcc ctacaggtcc
      361 agttggaccc aaaggagaga atggctctgc tggcgaacct ggaccaaagg gagaacgtgg
      421 actaagtggt gagcaatggg tcatgggatt cagatggtgc ctccagctgt tgcttctatc
      481 atgaggatgt gtaagtggca ggcacagaac ctgtcccagg aagaggaaac atggatgatc
      541 tgaagaaggc catgggttgt ttcagggtgg actgttgtct ctttgacctc tgctctgaca
      601 tgcatcttct aggctcccac ttccaaagga atttactgtc attgactggg ctgatccttt
      661 tctgccagga cctccaggac ttccaggtat tcctggtcca gctgggaaag aaggtccctc
      721 tgggaagcag gggaacatag gacctcaagg caaaccaggt cctaaaggag aggctgggcc
      781 caaaggtggg tgagggctgt ggtttgtctg ggagggagaa aagaggaatg acagtagaag
      841 ccattagtgt gatatgaggc atggtcagac ggacaggaag aatccagagt aataacatta
      901 tgaagacctt ggcatgcaca cctctctcac atctctccac accaacagga tttctccatc
      961 accctccaca gaggaaccct aggcagcaag aaacagtatt gccaactccc tatctaaccc
     1021 tgtagttctc tgtgacagac ctactgctga agaaatgttg cacactgaac acactcttag
     1081 cagtatttct ttctttttgt tcctttctca ggagaagtag gtgctcctgg catgcaagga
     1141 tctacagggg caaaaggctc cacaggcccc aagggagaaa gaggtgcccc tggtgtgcaa
     1201 ggagccccag ggaatgctgg agcagcaggt gagtgatgac tagatagagg gctggatgga
     1261 gttctattgt gccccctttg agctgacaag gcctagatcc tgtcttagtt agggttttac
     1321 tgctgtgaac agacaccatg atcaaggcaa gtcttattta tatgtataaa aacaaaacaa
     1381 aacaaaacaa aacaaaacaa aacaaaacca cacatttaat tgggcctggc ttacaggttt
     1441 aggtgggagc agggtagcat ccaggcagga atggtgcagg aggagctgag agttctatgt
```

1501 cttcatccaa aggctgctag tgaaaaactg acttccagga aactagggtg aggatcttaa
1561 gcccacatcc acagtgacac acctactcca accaggtcac ccaattccaa caaggccaca
1621 cctccaaatg gtgccactcc ctggtccaag aatatacaaa ccatcacaga tcccttccaa
1681 aatttaccct ttaagtaaag cctgtgctgt ggaggcaatc tgtttctaag gaggttgagt
1741 gtagagccaa tgggactaat cacacatagg gagaggaatg ctttacgcca tagggaatcg
1801 ttttttagga ttcagataaa gttgattgcc agatgctaag gacagccaaa cacatagtgg
1861 gcatgtgtct gggttctgcg aagagtctgt ttgtgggata aacataaact atggacagtg
1921 aggtgaaaat ccccagaaat aaacatacaa ccaagccaga tcttcttgaa tttcaggacc
1981 tgccggacct gccggtccac agggagctcc aggttccagg gggcccccag gactcaaggg
2041 ggacagaggt gttcctggag acagaggaat caaaggtgaa agcgggcttc caggtaagcc
2101 agaagcttca ggtcctggga cctgcactgg gatgtagccc tcttcatcga ggcaggagca
2161 tctcattagt atgggtggtc aatagaactt ggggtcagct gataagtttg aaagaaagcc
2221 aattatattc tggcttagct atctatcgaa gtcatatggg aaggcaaatt gagataattt
2281 catcatctta ccatgggaaa taggtgacaa cagctggctt cctggtttat aggctagcct
2341 tgtgaatcta acaatccagt ggatgagaat tatgtaagta ctatgtgttt ctgttctttt
2401 cccaagactg tcaagttgtc attatatttc ctgaacctca actcagaata ccttcttgtt
2461 tgatctgcca tctgcttcag gctcagttgg gggttcaggg gcttatttag ggatcataag
2521 aaatccctac acagaatctg ggtgtcatga atcatacttg tcatcctagt actagggagg
2581 ttatagcagg atcacagagt gagatgctag cctggcctac tttagatcat gtcttgaaaa
2641 caagcaaaca acagactagg gagatggatg ggtgggtgaa atgcttgcca aacagatatg
2701 aacccagagt tcagagttcc aaagtacaaa tccccaaatg gtgagtggac atgacagcct
2761 ccccagcaat cccagctcta tagagtccga ggtgaggtcc ctggaacaaa ctggctagct
2821 agactaggca aatcagcaag ctctgggttc cagagagatt ctacctcaat atataaggtg
2881 atagaagaag atgaccaaca gaattgttgg gccacatgtg cataaacatg taacaagtac
2941 aacacacaca ctcactcccc ccaacaaaca aaggattttt gagaaaagaa agaaaagcat
3001 tcatagcagg gacctgctgt gcctacagat gtaaagtttg ttatgagtaa caagctcccc
3061 accttaacct ttacccttt tctggggcac ggggacaatt tatccttagg cttgtatcat
3121 gatctccaca tggtcatttc ccttatttac ctcctatata gacagtgctg ctctgaggca
3181 gcagatggag gccttaaaag gaaaactaca gcgtctagag gttgccttct cccactatca
3241 gaaaggtgag tttctggata caaataaaat tgagctaggg ttgctcattt ggttgagctt
3301 tgctctaggt ctaggatgag aatgagcaag aactaagtct aaggctgaac ttggggcttg
3361 aacaacacag ttggcatcta tcaacataca tgtgtgtatg tgtgtcatag acacacatat
3421 acatcaagat tgggtcctcc tatgtacttt ggataaacga gtgggcatat tcatgtagga
3481 cacacaagtc caccaaggat caaccctggg acagcacatg actcttacat ggccctgatc
3541 agggctgctc ttggtaccat actgcaccta tggagactct tccccaggat ccttttaaag
3601 atactcagag gaaaagaaca gtgaaccagg gtgcagcttg gattctgacc agcttcagtg
3661 taggattgga acacagactt tgccagtatc tagtttcttt tgctgacttt catgcatggg
3721 cacaaagcca cagagatagg tgcacactga aggaaagatg tatggcacgc tatacagatg
3781 atatttttg tgatgtcatt tgttacttca aggtcaagga agtggaaagg aagtgacttc
3841 gtcttgccct cttagatagt cattgccaga gtcttaataa ccaacttgat tcctttctgg
3901 gttcacaatt aaacccatcc cctatttccc tgaatacgtg gacctctcaa atgtacttca
3961 aacccctgac cagcactttc aaggctagga ggtccaggac agcctgggac ttgacccaac
4021 cctcagatgc ccactatttc cttagctgtt ccttaactgt agcacatagt gcaaaccctg
4081 gctgacaacc tatgtttcta ccaccaggtt gtctatgggc tcactgcact gtgtcttctg
4141 tacactcaaa agtaatttcc tacaaaaagt tgttcatttc tacacagtaa ctaatattat
4201 catgaaaagt cagagttttg ttttaggggt agctccgtgt atgtgtgtgt gtgtgtgtgt
4261 gtgtgtgtgt gtgtgtgtgt gtgtgttcaa gctagaggtt gatatcagat gttcttcttc
4321 accactttat actttatttt taaattttta attatgtata tgtatatgca tgggctatgt
4381 ggccaactgt accacagtgc tcatgtggaa gtcagagaac aaccttatgg attcagttct
4441 ctccttttac ctttatgtag cttccaaaga ttgaactcag gtcacaaagc tgtgcagcaa
4501 aaggcaaaca cctttacctg tgagccatct ctctagctct ccatcgtgtc ttttttgaga
4561 caagatctct cactaaaact ttctgtttta gtgatattgt ctggttagtg gggcctttct
4621 ccctatcctg gggtccacct gtctccatac cctctgtccc catggacagc ttgtatatga
4681 tgcagtgtgt cagctcctca gacttgcaaa gctagcattt gatccattga gctctggtat
4741 cagagtgagt aaggctcccc acaatcttac caacttgggg ccttgtactg aacttctttg
4801 agcttacggg ctcttgcttg tgaatgaggc aagatggcat tctctcatag agtttactgt
4861 ggtgctagga ccacagtctt tgcttcatca tttctggttc cccaataccc ttcattagaa
4921 ggttgagaga aggctccaac tccagctcaa atttctcaca ggcatctaag ggctgccaca
4981 actagagccc accttctatc ttctttcagc tgcattgttc cctgatggcc gaagtgttgg
5041 agacaagatc ttcaggacag cagactctga aaagcctttt gaggatgccc aggagatgtg
5101 caaacaggct ggaggacagc tggcctcccc acgttctgct actgagaatg ctgccataca
5161 gcaactcatc acagcccaca acaaggctgc tttcctgagt atgacagatg tgggcacaga
5221 gggcaagttc acttacccca caggagagcc cctggtctat tctaattggg ctccagggga
5281 gcccaacaac aatggtggag cagagaactg tgtggagatc ttcaccaatg ggcaatggaa
5341 tgataaggct tgtggagagc agcgccttgt tatctgtgag ttctgagcca ctggaggtaa
5401 aaaaaaaaaa aaaaaaaaaa caggacaata tttggccagg agtttaacca gtgtactatg
5461 gctcagacct gtatgttgcc aacatcctaa taaaaggtgt atagcttgag ttctctgtat
5521 agcttgggtt ctctacagaa agcagcccaa aggcagggct attgcagaat acatttctca
5581 gagtatcatg cac

Appendix 4: The Sequence of the Murine SP-D cDNA.

```
LOCUS       NM_009160    1288 bp    mRNA            ROD    04-JAN-2000
DEFINITION  Mus musculus surfactant associated protein D (Sftpd), mRNA.
ACCESSION   NM_009160
NID         g6677920
VERSION     NM_009160.1  GI:6677920
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1288)
  AUTHORS   Lawson,P.R., Perkins,V.C., Holmskov,U. and Reid,K.B.M.
  TITLE     Genomic organisation of the mouse gene for surfactant protein D
            (sftp4)
  JOURNAL   Am. J. Respir. Cell Mol. Biol. 20 (5), 953-963 (1999)
COMMENT     REFSEQ: This reference sequence was derived from AF047742.1.
            PROVISIONAL RefSeq: This is a provisional reference sequence record
            that has not yet been subject to human review. The final curated
            reference sequence record may be somewhat different from this one.
FEATURES             Location/Qualifiers
     source          1..1288
                     /organism="Mus musculus"
                     /strain="129/Sv"
                     /sub_strain="PAS"
                     /db_xref="taxon:10090"
                     /chromosome="14"
                     /map="14 14.0 cM"
                     /cell_line="D3"
                     /cell_type="embryonal stem"
                     /dev_stage="embryo"
     gene            1..1288
                     /gene="Sftpd"
                     /db_xref="LocusID:20390"
                     /db_xref="MGD:MGI:109515"
     CDS             43..1167
                     /gene="Sftpd"
                     /codon_start=1
                     /db_xref="LocusID:20390"
                     /db_xref="MGD:MGI:109515"
                     /product="surfactant associated protein D"
                     /protein_id="NP_033186.1"
                     /db_xref="PID:g6677921"
                     /db_xref="GI:6677921"
                     /translation="MLPFLSMLVLLVQPLGNLGAEMKSLSQRSVPNTCTLVMCSPTEN
                     GLPGRDGRDGREGPRGEKGDPGLPGPMGLSGLQGPTGPVGPKGENGSAGEPGPKGERG
                     LSGPPGLPGIPGPAGKEGPSGKQGNIGPQGKPGPKGEAGPKGEVGAPGMQGSTGAKGS
                     TGPKGERGAPGVQGAPGNAGAAGPAGPAGPQGAPGSRGPPGLKGDRGVPGDRGIKGES
                     GLPDSAALRQQMEALKGKLQRLEVAFSHYQKAALFPDGRSVGDKIFRTADSEKPFEDA
                     QEMCKQAGGQLASPRSATENAAIQQLITAHNKAAFLSMTDVGTEGKFTYPTGEPLVYS
                     NWAPGEPNNNGGAENCVEIFTNGQWNDKACGEQRLVICEF"
     sig_peptide     43..99
     mat_peptide     100..1164
                     /product="surfactant associated protein D"
BASE COUNT      351 a    310 c    385 g    242 t                ORIGIN
        1 atttacctgg acctcctggg gccaacaagg aagcaatctg acatgctgcc ctttctctcc
       61 atgcttgtct tgcttgtaca gcccctggga aatctgggag cagaaatgaa gagcctctcg
      121 cagagatcag tacccaacac ctgcacccta gtcatgtgta gcccaacaga gaatggcctg
      181 cctggtcgtg atggacggga tgggagagaa ggtccacggg gtgagaaggg tgatccaggt
      241 ttgccaggac ctatggggct ctcagggttg cagggcccta caggtccagt tggacccaaa
      301 ggagagaatg gctctgctgg cgaacctgga ccaaagggag aacgtggact aagtggacct
      361 ccaggacttc caggtattcc tggtccagct gggaaagaag gtccctctgg gaagcagggg
      421 aacataggac ctcaaggcaa accaggtcct aaaggagagg ctgggcccaa aggagaagta
      481 ggtgctcctg gcatgcaagg atctacaggg gcaaaaggct ccacaggccc caagggagaa
      541 agaggtgccc ctggtgtgca aggagcccca gggaatgctg gagcagcagg acctgccgga
      601 cctgccggtc cacagggagc tccaggttcc agggggcccc caggactcaa gggggacaga
      661 ggtgttcctg gagacagagg aatcaaaggt gaaagcgggc ttccagacag tgctgctctg
      721 aggcagcaga tggaggcctt aaaaggaaaa ctacagcgtc tagaggttgc cttctcccac
```

781 tatcagaaag ctgcattgtt ccctgatggc cgaagtgttg gagacaagat cttcaggaca
841 gcagactctg aaaagccttt tgaggatgcc caggagatgt gcaaacaggc tggaggacag
901 ctggcctccc cacgttctgc tactgagaat gctgccatac agcaactcat cacagcccac
961 aacaaggctg ctttcctgag tatgacagat gtgggcacag agggcaagtt cacttacccc
1021 acaggagagc ccctggtcta ttctaattgg gctccagggg agcccaacaa caatggtgga
1081 gcagagaact gtgtggagat cttcaccaat gggcaatgga atgataaggc ttgtggagag
1141 cagcgccttg ttatctgtga gttctgagcc actggaggta aaaaaaaaaa aaaaaaaaaa
1201 acaggacaat atttggccag gagtttaacc agtgtactat ggctcagacc tgtatgttgc
1261 caacatccta ataaaaggtg tatagctt

Appendix 5: Primers used for the characterisation of porcine SP-D

Primer name	Sequence
PD-10F	CTGGGCATGCAGGGCTCTAC
PD-11F	GGACAGAGGTCCTCCTGGAG
PD-12F	CAGAGAACTGTGTGGAGATC
PD-1F	CTGATCCTGCTCACACAGC
PD-1R	CCTGGATCCCCTTTCTCGC
PD-2F	TGCCTGGTCGTGATGGTCG
PD-2R	AGGTCCTATGTTCCCCTGC
PD-3F	CAGTATAAGAAAGTGGAGCAC
PD-3R	GATGTCAGTCATGCTCAGG
PD-4F	CGGAGGGCAATTTCACCTAC
PD-4R	TGGCCAGCAGAAGGTCAC
PD-5'PE	GGCTACACATGACCAGGGCGCAGGCGTTGGC
PD-5F	CCAAAGGAGAAGTGGGTGC
PD-5R	CAGTGATGCCTGGGAGTC
PD-7F	ATGCGCCCTTTGAGGCCGAG
PD-8F	ATGGAGAATGGCCTGCCTGG
PD-9F	CATTGGACCATGCGGGCCTC
PD-CRD-F	CTTCCCCAATGGCCGAGGTG
PD-CRD-R	TCAGAACTCGCAGATCACGAG
PD-CYS-F	GTTTGCCAGGAGCTGTAGGGCGAGC
PD-CYS-R	CTCTCCTTTGGGGCCTGGTGTGCC
PSPA-F	TCCTCAGTTTCCAGGAGTTCATGC
PSPA-R	AAGGCAGGACCCTATCCATCCTGTG
XSC	GACTCGAGTCGACATCGAT
$XSCdG_{17}$	$GACTCGAGTCGACATCGATG_{17}H$[1]

[1]H represents the single IUB code that indicates the nucleotides; A, T and C. This extra nucleotide in $XSCdG_{17}$ acts to anchor the oligonucleotide at the junction of a polyC tract and normal sequence.

Appendix 6: The Sequence of the Chimeric Porcine cDNA Clone, pPIG17d

```
   1 AATTCCGGCA TGAAGGTCAC TGTAAACTTG CAGACTTTGG GATGTGCAAG
  51 GAGGGGATCT GTAACGGGGT CACCACGGCA ACYTTCTGTG GCACGCCCGA
 101 CTACATTGYT CCAGAGATCC TCCAGGAAAT GCTGTACGGG CCTGCAGTAG
 151 ACTGGTGGGC GATGGGTGTG CTGCTGTACG AGATGCTCTG CGGCCATGCG
 201 CCCTTTGAGG CCGAGAATGA AGACAGGAGC TCCTGCTGTG GTATGGGATC
 251 AGCGGCGCGT CTGGACATGC TTCTCCTCCC TCTCTCCGTG CTGATCCTGC
 301 TCACACAGCC CCCGAGGTCA CTGGGAGCAG AAATGAAGAC CTATTCCCAG
 351 AGAGCAGTGG CCAACGCCTG CGCCCTGGTC ATGTGTAGCC CCATGGAGAA
 401 TGGCCTGCCT GGTCGTGATG GTCGGGATGG GAGAGAGGGC CCTCGGGGCG
 451 AGAAGGGGGA TCCAGGTTTG CCAGGAGCTG TAGGGCGAGC GGGGATGCCT
 501 GGACTGGCTG GCCCAGTTGG GCCCAAAGGG GACAACGGCT CTACTGGAGA
 551 ACCCGGAGCA AAGGGAGACA TTGGACCATG CGGGCCTCCA GGACCTCCAG
 601 GTATACCTGG TCCAGCCGGA AAAGAAGGTC CCTCAGGGCA GCAGGGGAAC
 651 ATAGGACCTC CAGGCACACC AGGCCCCAAA GGAGAGACTG GGCCCAAAGG
 701 AGAAGTGGGT GCCCTGGGCA TGCAGGGCTC TACAGGGGCA AGAGGCCCTG
 751 CAGGTCTTAA AGGAGAGAGA GGTGCCCCCG GTGAGCGTGG AGCCCCTGGA
 801 AGTGCTGGGG CAGCAGGGCC TGCTGGAGCC ACGGGCCCTC AGGGCCCTTC
 851 AGGTGCCAGG GGCCCCCCAG GACTGAAAGG GGACAGAGGT CCTCCTGGAG
 901 AAAGAGGAGC CAAGGGAGAG AGTGGACTCC CAGGCATCAC TGCTCTGAGG
 951 CAACAGGTGG AGACCTTACA GGGGCAGGTA CAACGCCTCC AGAAGGCCTT
1001 CTCTCAGTAT AAGAAAGTGG AGCTCTTCCC CAATGGCCGA GGTGTCGGGG
1051 AGAAGATCTT CAAGACGGGA GGCTTTGAAA AGACTTTTCA GGATGCTCAG
1101 CAGGTATGCA CACAGGCCGG GGG
```

nucleotides	
1 - 1123	the porcine cDNA clone pPIG17d
1 - 263	procine protein kinase C.
264 - 1123	porcine surfactant protein D.

Appendix 7: The Sequence of the Porcine SP-D cDNA

```
LOCUS       AF132496     1385 bp    mRNA            MAM   01-JAN-2000
DEFINITION  Sus scrofa lung surfactant protein D precursor (SFTPD) mRNA,
            complete cds.
ACCESSION   AF132496
NID         g6651047
VERSION     AF132496.1  GI:6651047
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1385)
  AUTHORS   van Eijk,M., Haagsman,H.P., Skinner,T., Archibold,A., Reid,K.B.M.
            and Lawson,P.R.
  TITLE     Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
            localisation and tissue distribution
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1385)
  AUTHORS   van Eijk,M., Haagsman,H.P., Reid,K.B.M. and Lawson,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAR-1999) Biochemistry, Oxford University, South
            Parks Road, Oxford, Oxon OX1 3QU, England
FEATURES             Location/Qualifiers
     source          1..1385
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /chromosome="10"
     5'UTR           1..43
                     /gene="SFTPD"
     gene            1..1385
                     /gene="SFTPD"
     sig_peptide     44..103
                     /gene="SFTPD"
     CDS             44..1180
                     /gene="SFTPD"
                     /function="Involved in innate immune system and lipid
                     homeostasis in the lung"
                     /note="collectin; SPD; SP-D; CP4"
                     /codon_start=1
                     /product="lung surfactant protein D precursor"
                     /protein_id="AAF22145.1"
                     /db_xref="PID:g6651048"
                     /db_xref="GI:6651048"
                     /translation="MLLLPLSVLILLTQPPRSLGAEMKTYSQRAVANACALVMCSPME
                     NGLPGRDGRDGREGPRGEKGDPGLPGAVGRAGMPGLAGPVGPKGDNGSTGEPGAKGDI
                     GPCGPPGPPGIPGPAGKEGPSGQQGNIGPPGTPGPKGETGPKGEVGALGMQGSTGARG
                     PAGLKGERGAPGERGAPGSAGAAGPAGATGPQGPSGARGPPGLKGDRGPPGERGAKGE
                     SGLPGITALRQQVETLQGQVQRLQKAFSQYKKVELFPNGRGVGEKIFKTGGFEKTFQD
                     AQQVCTQAGGQMASPRSETENEALSQLVTAQNKAAFLSMTDIKTEGNFTYPTGEPLVY
                     ANWAPGEPNNNGGSSGAENCVEIFPNGKWNDKACGELRLVICEF"
     mat_peptide     104..1177
                     /gene="SFTPD"
                     /product="lung surfactant protein D"
     3'UTR           1181..1347
                     /gene="SFTPD"
     polyA_signal    1232..1237
                     /gene="SFTPD"
     polyA_signal    1324..1329
                     /gene="SFTPD"
BASE COUNT      351 a    364 c    453 g    217 t                ORIGIN
        1 cgagtttgcc tggagattct gagctctaga ggacgcaact gacatgcttc tcctccctct
       61 ctccgtgctg atcctgctca cacagccccc gaggtcactg ggagcagaaa tgaagaccta
      121 ttcccagaga gcagtggcca acgcctgcgc cctggtcatg tgtagcccca tggagaatgg
      181 cctgcctggt cgtgatggtc gggatgggag agagggccct cggggcgaga agggggatcc
      241 aggtttgcca ggagctgtag ggcgagcggg gatgcctgga ctggctggcc cagttgggcc
      301 caaaggggac aacggctcta ctggagaacc cggagcaaag ggagacattg gaccatgcgg
      361 gcctccagga cctccaggta tacctggtcc agccggaaaa gaaggtccct cagggcagca
```

```
 421 ggggaacata ggacctccag gcacaccagg ccccaaagga gagactgggc ccaaaggaga
 481 agtgggtgcc ctgggcatgc agggctctac aggggcaaga ggccctgcag gtcttaaagg
 541 agagagaggt gcccccggtg agcgtggagc ccctggaagt gctggggcag cagggcctgc
 601 tggagccacg ggccctcagg gcccttcagg tgccaggggc cccccaggac tgaaagggga
 661 cagaggtcct cctggagaaa gaggagccaa gggagagagt ggactcccag gcatcactgc
 721 tctgaggcaa caggtggaga ccttacaggg gcaggtacaa cgcctccaga aggccttctc
 781 tcagtataag aaagtggagc tcttccccaa tggccgaggt gtcggggaga agatcttcaa
 841 gacgggaggc tttgaaaaga cttttcagga tgctcagcag gtatgcacac aggccggggg
 901 acagatggcc tccccacgct ctgagactga gaacgaggcc ttgagccagc tggtcacagc
 961 tcagaataag gctgctttcc tgagcatgac tgacatcaag acggagggca atttcaccta
1021 ccccacgggg gagcccctgg tctatgccaa ctgggccccc ggggagccca acaacaatgg
1081 tggcagcagc ggagcagaga actgtgtgga gatctttccc aatggcaagt ggaatgacaa
1141 ggcctgcgga gaactgcgcc tcgtgatctg cgagttctga gcccctgggg agggaggggc
1201 ggtgtccaga gctgtgtgct accaacgtcc caataaatag gtgaccttct gctggccagg
1261 gcttctccac agagccgtgg gacgaggcca gaaggtaggg agcctatgga acgcctccct
1321 cagaataaag tacgaaactg gcctcacaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
1381 aaaaa
```

Appendix 8: The Partial cDNA Sequence of Porcine SP-A

```
LOCUS       AF133668      353 bp    DNA             MAM       31-DEC-1999
DEFINITION  Sus scrofa lung surfactant protein A (SFTPA) gene, partial cds.
ACCESSION   AF133668
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   van Eijk,M., Haagsman,H.P., Skinner,T., Archibald,A., Reid,K.B.M.
            and Lawson,P.R.
  TITLE     Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
            localisation and tissue distribution
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 353)
  AUTHORS   van Eijk,M., Haagsman,H.P., Reid,K.B.M. and Lawson,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-1999) Biochemistry, Oxford University, South
            Parks Road, Oxford, Oxon OX1 3QU, England
FEATURES             Location/Qualifiers
     source          1..353
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
     gene            <1..353
                     /gene="SFTPA"
     CDS             <1..353
                     /gene="SFTPA"
                     /function="involved in the innate immune system and lipid
                     homeostasis within the lung"
                     /note="collectin; SPA; SP-A"
                     /codon_start=3
                     /product="lung surfactant protein A"
                     /translation="AVGEKVFSTNGQSVAFDVIRELCARAGGRIAAPRSPEENEAIAS
                     IVKKHNTYAYLGLVEGPTAGDFFYLDGTPVNYTNWYPGEPRGRGKEKCVEMYTDGQWN
                     DRNCQQYRLAICEF"
BASE COUNT       90 a     79 c    109 g     75 t
ORIGIN
        1 tggcagtggg agagaaggtc ttctccacca atgggcagtc agtcgctttt gatgtcatta
       61 gagagttgtg tgccagagca ggtggacgca tcgctgcccc aaggagtcca gaggagaatg
      121 aggccattgc aagcattgtg aagaaacaca acacttatgc ttacctcggc ctggttgagg
      181 gccccactgc tggagacttc ttctacttgg atggaacccc tgtgaattac accaactggt
      241 acccagggga acccaggggt cggggcaaag agaagtgtgt ggagatgtac acagatggcc
      301 agtggaatga caggaactgc cagcagtacc gactggccat atgtgagttt tga
```

Appendix 9: Primers used for chromosomal localisation of complement genes

Primer	Primer sequence
C1r-2f	CAAGATGCTGACCAGAGCTGGCAGC
C1r-2r	CTTCATTCTTCCAAATGCCTTGGGC
C1s-1f	CTCCACGATGACATGAGGAATTGTG
C1s-1r	TGCAGCTGAATCTGGTATTCACACC
C3aR-1f	TTTTCTCACTGAGGCATCTATTCAG
C3aR-1r	TAGACAAGTGAGACCAAGAATGACC
Crarf-1f	CTGGGGGAAGCAGTTCTTACAGAGG
Crarf-1r	ATAGGCCTCCTGGCAGGTGTCAGAG
Masp2-1f	AACAGCCGCTCATGCTGTATATGAG
Masp2-2r	CCCCACTGTCACCTCTGCAGCTGTC
Masp2-3f	TGATCCAACGGCGCCACCTAATGGC
Masp2-3r	AAGATCCATCTTTCTGACAGACAGC

PUBLICATION LIST

A. Original Communications in Reviewed Journals

Holmskov, U., Lawson, P., Teisner, B., Tornoe, I., Willis, A. C., Morgan, C., Koch, C., and Reid, K. B. (1997). Isolation and characterization of a new member of the scavenger receptor superfamily, glycoprotein-340 (gp-340), as a lung surfactant protein-D binding molecule. J Biol Chem *272*, 13743-9.

Lawson, P. R., Perkins, V. C., Holmskov, U., and Reid, K. B. M. (1999). Genomic organisation of the mouse gene for lung surfactant protein D (Sftpd). Am J Respir Cell Mol Biol *20*, 953-63.

Van Eijk, M., Haagsman, H. P., Skinner, T., Archibald, A., Reid, K. B. M., and Lawson, P. R. (2000). Porcine lung surfactant protein D: Complementary DNA cloning, chromosomal localization and tissue distribution. J Immunol *164*, 1442-50.

Lawson P. R. and Reid K. B. M. (2000). A Novel PCR-based Technique Using Expressed Sequence Tags and Gene Homology for Murine Genetic Mapping: Localisation of the Complement Genes. Int. Immunol *12*, 231-40.

B. Chapters in Books and Review Articles

Lawson P. R. and Reid KBM. (1999). Conglutinin. In: The Complement Factsbook, ed. Walport M and Morley B. Academic Press Ltd., London.

Lawson P. R. and Reid K. B. M. (1999). Mannose-Binding Lectin. In: The Complement Factsbook, ed. Walport M and Morley B. Academic Press Ltd., London.

Shah P. L., Lawson P. R., Reid K. B. M. and Morgan C. (2000). Current Concepts on the Pathogenesis and Treatment of Pulmonary Alveolar Proteinosis. Thorax *55*, 67-77.

Lawson P. R. and Reid K. B. M. (2000). The Roles of Surfactant Proteins A and D in Innate Immunity. Immunol Rev, *173*, 66-78.

C. Abstracts

Reid K. B. M., Lawson P., Hoppe H. J., Kishore U., Wang J. Y., Lu J. L., Lim B. and Holmskov U. (1993). Human and Bovine Lung Surfactant Protein D: Cloning, Expression and Functional Studies Involving the Binding to Allergens and Receptors. Applied Cardiopulmonary Pathophysiology *5*:(supp. 3) 96-98.

Holmskov U., Lawson P., Teisner B., Tornøe I., Willis A. C., Morgan C., Koch C and Reid K. B. M. (1996). Isolation and Tissue Localisation Of Glycoprotein-340 (Gp-340) and Characterisation of its Binding to the Collectins. Mol. Immunol. *33*, 85.

Lawson P. R and Reid K. B. M. (1998). Genetic Mapping of the Murine Anaphylatoxin Receptor for C3a and Three Activating Serine Proteases of the Complement System (Masp-2, C1s, C1r). Mol. Immunol., *35*, 376.