UNIVERSITY OF TECHNOLOGY, SYDNEY

Proteomic Investigation of the Genome-Reduced Pathogen, *Mycoplasma hyopneumoniae*

Jessica Leigh Tacchi

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

The ithree institute

University of Technology, Sydney

2015

Declaration

I, Jessica L. Tacchi, certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.
I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.
This thesis includes chapters that have been published, or are in preparation for publication in refereed journals. I declare that I was primarily responsible for data collection, analysis and interpretation and responding to editing suggestions from co-authors. The contributions from other authors are described at the beginning of each chapter.

Jessica Leigh Tacchi

2015

Acknowledgements

Right from the outset, completing my PhD has been an amazing experience, and I owe it to so many people.

Professor Steve Djordjevic and Dr. Matt Padula. I would never have come this far if not for their enthusiasm, dedication and support. I could not have asked for better role models, each inspirational in their own, often completely different ways. I could not be prouder of my academic roots and hope that I can in turn pass on the values and the passion for science that they have instilled in me.

I extend my sincerest gratitude to those who have helped contributed to this thesis, whether directly or through providing technical assistance or discussions and ideas: Cheryl Jenkins, Linda Falconer, Lauren Woolley, Chris Minion, Paul Haynes, Mark Raftery, Joyce To, Joe Moxon and Jens Coorssen. I am also fortunate to have followed in the footsteps of Lisa Seymour, Ania Deutscher and Daniel Bogema and I thank them for being so open to answering my questions and letting me build on their foundations.

To the researchers at UTS and the ithree institute with whom I shared scientific discussions, ideas and endeavours, I must thank for their encouragement and collaboration. I am also grateful to have had the support of the most wonderful technical staff and lab managers over the years; they were always happy to help.
I am also grateful to UTS for generously providing a Scholarship to fund my PhD.

Thanks also goes to the AB Sciex representatives and technical support staff for keeping the instruments running and always being willing to let me watch and learn.

To my family; Mum, Wayne and Gabbie, who had no idea what I was talking about half the time but smiled and nodded anyway.
To Luke, for everything, but mostly for bringing me coffee. Soon it'll be my turn.

Finally, I have to thank the proteomics lab family. Ben, Jerran, Isa, Mike, Krish, Kate, Jacqui, Ronnie, Iain, Marz, Samira, "George", Joel and the countless students and visitors. I could not have asked for better labmates or friends.

Table of Contents

List of Figures

List of Tables

Abbreviations

↓	Denotes site of cleavage
1D	One-dimensional
2D	Two-dimensional
6xHis-tagged	Hexahistidine tagged
ACN	Acetonitrile
amu	atomic mass units
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BALF	Bronchoalveolar lavage fluid
BLAST	Basic Local Alignment Search Tool
BM	Basement membrane
BN	Blue native
BSA	Bovine Serum Albumin
C-terminus	Carboxyl-terminus
C7BzO	3-(4-Heptyl)phenyl-3-hydroxypropyl)dimethylammoniopropanesulfonate
CDS	Coding sequences
CN	Clear Native
CO_2	Carbon Dioxide
ddH2O	Double-distilled water
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediamine tetraacetic acid
EfTu	Elongation factor Tu
ELISA	Enzyme-linked immunosorbent assay
EMAI	Elizabeth Macarthur Agricultural Institute
ESI	Electrospray Ionisation
ExPASy	Expert Protein Analysis System
Fn	Fibronectin
FnBP	Fibronectin binding protein
FT	Fourier transform
FT ICR	Fourier transform ion cyclotron resonance
G + C	Guanine and cytosine
GAG	Glycosaminoglycan
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GeLC-MS/MS	Gel-electrophoresis-coupled liquid chromatography tandem mass spectrometry
GO	Gene Ontology
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLB	Hydrophillic-lipophillic balanced
HRP	Horseradish peroxidase

HSP	Heat shock protein
ID	Inside/Inner Diameter
IDA	Intelligent Data Acquisition
IEF	Isoelectric focusing
Ig	Immunoglobulin
IMAC	Immobilised metal affinity chromatography
IPG	Immobilised pH gradient
IPTG	Isopropyl-β-D-thiogalactopyranoside
KCl	Potassium Chloride
kDa	Kilodalton
kVh	KiloVolt Hours
LB	Luria-Bertani
LC	Liquid chromatography
LPS	Lipopolysaccharide
LTQ	Linear Trap Quadrupole
MØ	Macrophages
MAb	Monoclonal antibody
MALDI	Matrix-assisted laser desorption/ionization
MMP	Matrix metalloproteinases
mRNA	Messenger Ribonucleic Acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MudPIT	Multidimensional protein identification technology
N-terminus	Amino-terminus
NH_4HCO_3	Ammonium bicarbonate
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAI-1	Plasminogen activation inhibitor-1
PAMP	Pathogen-associated molecular pattern
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline-Tween 20
PDB	Protein Database
PEP	Porcine enzootic pneumonia
p*I*	Isoelectric point
PICS	Proteomic identification of protease cleavage site
PK15 cell	Porcine kidney epithelial-like cell
Plg	Plasminogen
PONDR	Predictor of Naturally Disordered Regions
PPLO	Pleuropneumonia-like organism
ppm	Parts per million
PRRSV	Porcine reproductive and respiratory syndrome virus
PVDF	Polyvinylidene fluroride
Q-TOF	Quadrupole time-of-flight

R1	Repeat region 1
R2	Repeat region 2
rpm	Revolutions per minute
SCX	Strong cation exchange
SDS	Sodium Dodecyl Sulphate
SEM	Scanning electron microscopy
SPE	Solid phase extraction
Sulfo-NHS-LC-biotin	Sulfosuccinimidobiotin long chain biotin
TAILS	Terminal amine isotopic labelling of substrates
TBP	Tributylphosphine
TCA	Trichloroacetic acid
TCA cycle	Tricarboxylic acid cycle
TEM	Transmission electron microscopy
TFA	Trifluoroaceteic Acid
TOF	Time of flight
tPA	Tissue-type plasminogen activator
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Tuf_{Sp}	Elongation factor Tu (Tuf) of *Streptococcus pneumoniae*
Tween-20	Polyoxyethylenesorbitan monolaurate
TX-100	Triton X-100
TX-114	Triton X-114, Octylphenoxy polyethoxyethanol
uPA	Urokinase-type plasminogen activator
uPAR	Urokinase-type plasminogen activator receptor
v/v	Volume per volume
w/v	Weight per volume
WHO	World Health Organisation

Publications

"Sequence TTKF ↓ QE defines the site of proteolytic cleavage in Mhp683 protein, a novel glycosaminoglycan and cilium adhesin of *Mycoplasma hyopneumoniae*."
Bogema DR, Scott NE, Padula MP, Tacchi JL, Raymond BB, Jenkins C, Cordwell SJ, Minion FC, Walker MJ, Djordjevic SP. *J Biol Chem.* 2011.

"Mhp182 (P102) binds fibronectin and contributes to the recruitment of plasmin(ogen) to the *Mycoplasma hyopneumoniae* cell surface."
Seymour LM, Jenkins C, Deutscher AT, Raymond BB, Padula MP, Tacchi JL, Bogema DR, Eamens GJ, Woolley LK, Dixon NE, Walker MJ, Djordjevic SP. *Cell Microbiol.* 2012.

"*Mycoplasma hyopneumoniae* Surface proteins Mhp385 and Mhp384 bind host cilia and glycosaminoglycans and are endoproteolytically processed by proteases that recognize different cleavage motifs."
Deutscher AT, Tacchi JL, Minion FC, Padula MP, Crossett B, Bogema DR, Jenkins C, Kuit TA, Walker MJ, Djordjevic SP. *J Proteome Res.* 2012.

"Characterization of cleavage events in the multifunctional cilium adhesin Mhp684 (P146) reveals a mechanism by which *Mycoplasma hyopneumoniae* regulates surface topography."
Bogema DR, Deutscher AT, Woolley LK, Seymour LM, Raymond BB, Tacchi JL, Padula MP, Dixon NE, Minion FC, Jenkins C, Walker MJ, Djordjevic SP. *MBio*. 2012.

"MHJ_0125 is an M42 glutamyl aminopeptidase that moonlights as a multifunctional adhesin on the surface of *Mycoplasma hyopneumoniae."*
Robinson MW, Buchtmann KA, Jenkins C, Tacchi JL, Raymond BBA, To J, Roy Chowdhury P, Woolley LK, Labbate M, Turnbull L, Whitchurch CB, Padula MP, Djordjevic SP. *Open Biol.* 2013.

"Formation of assemblies on cell membranes by secreted proteins: molecular studies of free λ light chain aggregates found on the surface of myeloma cells."

Hutchinson AT, Malik A, Berkahn MB, Agostino M, To J, Tacchi JL, Djordjevic SP, Turnbull L, Whitchurch CB, Edmundson AB, Jones DR, Raison RL, Ramsland PA. *Biochem J.* 2013.

"P159 from *Mycoplasma hyopneumoniae* binds porcine cilia and heparin and is cleaved in a manner akin to ectodomain shedding."
Raymond BBA[¥], Tacchi JL[¥], Jarocki VM, Minion FC, Padula MP, Djordjevic SP. *J Proteome Res.* 2013. ([¥]Co-first authors)

"Proteogenomic mapping of *Mycoplasma hyopneumoniae* virulent strain 232."
Pendarvis K, Padula MP, Tacchi JL, Petersen AC, Djordjevic SP, Burgess SC, Minion FC. *BMC Genomics.* 2014.

"Proteome analysis of multidrug-resistant, breast cancer-derived microparticles."
Pokharel D, Padula MP, Lu JF, Tacchi JL, Luk F, Djordjevic SP, Bebawy M. *Journal of extracellular vesicles.* 2014.

"Proteolytic processing of the cilium adhesin MHJ_0194 (P123) in *Mycoplasma hyopneumoniae* generates a functionally diverse array of cleavage fragments that bind multiple host molecules."
Raymond BB, Jenkins C, Seymour LM, Tacchi JL, Widjaja M, Jarocki VM, Deutscher AT, Turnbull L, Whitchurch CB, Padula MP, Djordjevic SP. *Cell Microbiol*. 2014.

"Non-proteolytic functions of microbial proteases increases pathological complexity."
Jarocki VM, Tacchi JL, Djordjevic SP. *Proteomics*. 2014.

"MHJ_0461 is a multifunctional leucine aminopeptidase on the surface of *Mycoplasma hyopneumoniae*."
Jarocki VM, Santos J, Tacchi,JL, Raymond BB, Deutscher AT, Jenkins C, Padula MP, and Djordjevic SP. *Open biology* 2015.

Conference Presentations – Presenting author

- 20th Lorne Proteomics Symposium, 2015 – Poster presentation: "Post-translational processing, multifunctionality and moonlighting."
- 19th Lorne Proteomics Symposium, 2014 – Poster presentation: "Proteolytic processing of adhesin proteins on the surface of *Mycoplasma hyopneumoniae"*
- BacPath 12: The Molecular Biology of Bacterial Pathogens, Tangalooma Resort QLD, 2013 – Poster presentation: "Protein Processing and Promiscuity in Pathogenesis"
- 18th Lorne Proteomics Symposium, 2013 – Oral presentation: "It's all about cleavage: generating diversity in a genome-reduced organism"
- Proteomics and Beyond Symposium, Macquarie University, 2012 – Poster Presentation: "Mining 2D gels for a wealth of biological information"
- 17th Lorne Proteomics Symposium, 2012 – Poster presentation: "The Surface Proteome of *Mycoplasma hyopneumoniae*"
- 19th Congress of the International Organisation for Mycoplasmology, Toulouse, France 2012 – Poster presentation: "Endoproteolytic Adhesin Fragments and Geographical Moonlighting Proteins Dominate the Surfaceome of *Mycoplasma hyopneumoniae*"
- 16th Lorne Proteomics Symposium, 2011 – Poster presentation: "Surface Proteins of a Reduced-Genome Pathogen: *Mycoplasma hyopneumoniae*"
- Royal North Shore Hospital Annual Scientific Research Meeting, 2010 – Poster Presentation: "Protein Complexes of *Mycoplasma hyopneumoniae* Revealed by High-Resolution Clear Native Electrophoresis"
- HUPO 9th Annual World Congress, Sydney 2010 – Poster presentation: "Protein Complexes of *Mycoplasma hyopneumoniae* Revealed by High-Resolution Clear Native Electrophoresis"
- Australian Society for Microbiology Annual Scientific Meeting & Exhibition, Sydney 2010 – Poster presentation: "A Comprehensive Proteomics Analysis of *Mycoplasma hyopneumoniae* Strain J"
- 18th Congress of the International Organisation for Mycoplasmology, Chianciano Terme, Italy 2010 – Oral presentation: "A Comprehensive Proteomics Analysis of *Mycoplasma hyopneumoniae* Strain J"

Awards and Scholarships

- Poster Presentation Award – 20th Lorne Proteomics Symposium, 2015
- Student Oral Presentation Award – 18th Lorne Proteomics Symposium, 2013
- Best poster – 17th Lorne Proteomics Symposium, 2012
- IOM Student Travel Award – 19th Congress of the International Organisation for Mycoplasmology, Toulouse, France, 2012
- Student Travel Award to attend the Human Proteome Organization Congress (HUPO) Congress in Sydney, 2010
- Vice-Chancellor's Postgraduate Research Students Conference Funding, 2010
- Faculty of Science Post Graduate Conference Funding, 2010
- IOM Student Travel Award – 18th Congress of the International Organisation for Mycoplasmology, Chianchiano Terme, Italy, 2012
- UTS Doctoral Scholarship, 2010

Abstract

Mycoplasma hyopneumoniae is a genome-reduced bacterium and an economically significant pathogen that chronically infects the respiratory tract of swine. This infection often leads to pneumonia and secondary infections, costing agricultural industries significantly in the use of antibiotics and vaccines, which are currently largely ineffective. An improved understanding of the molecular mechanisms behind the infection process is essential to our ability to rationally design better vaccine and therapeutic interventions. With fewer than 700 predicted protein coding sequences, *M. hyopneumoniae* possesses one of the smallest genomes of any free-living organism. As such, it lends itself well to thorough proteomic interrogation.

In this thesis, a range of proteomic techniques have been used to investigate the *M. hyopneumoniae* global and surface proteome at the protein and peptide level, including surface shaving and labelling techniques, ligand and immuno-blotting and affinity chromatography, as well as N-terminal dimethyl labelling to determine true N-termini of mature proteins. This conceptually unbiased, function-oriented approach has revealed an unexpected level of complexity in the use of proteolytic processing, multifunctional proteins and moonlighting to compensate for reduced coding capacity at the genome level. While microarray and transcriptome studies suggest that under normal culture conditions, the majority of genes are transcribed; our analyses identified less than 400 detectable expressed protein products under similar conditions. A significant number of the expressed proteins were discovered to be multifunctional, post-translationally modified by proteolysis.

Surface proteome analyses identified a range of proteins to be surface exposed, despite lacking known signal peptides. Even though many of these proteins had well-characterised functions in the cytoplasm, they were also identified to have secondary functions at the cell surface, a phenomenon known as moonlighting. Many of the proteins present at the cell surface were identified to be subjected to proteolytic cleavage events. These were predominantly cell surface adhesins, many of which have already been described in the literature, however a large number of cytoplasmic

"housekeeping" proteins are also found to be post-translationally cleaved, multifunctional proteins or moonlighting proteins.

These findings can be applied to improve the rational design and development of vaccines and therapeutics for the prevention and treatment of *Mycoplasma hyopneumoniae*, as well as having wider implications for the field of biology as a whole, if similar levels of post-translational regulation can be found in other bacterial pathogens.

Chapter 1. Introduction

Pork is currently the most consumed meat (per ton) worldwide, with the growing population constantly increasing the demand for animal-based protein. This increasing demand also calls for improvement in the efficiency of raising meat-production animals, with consideration of the associated environmental and agronomic impact. In a US study, the environmental burden of pork production was estimated to be approximately 1 order of magnitude less than that of beef (Eshel *et al*., 2014). This could be further improved with the control of prevalent economically-significant respiratory diseases which currently pose a challenge to pig farmers. Prevention or improved management of porcine respiratory diseases has the potential to improve efficiency and reduce the environmental impact of meat production, increasing the long-term sustainability of pig-farming. Up to 96% of swine herds worldwide experience outbreaks of mycoplasmal pneumonia (Escobar *et al*., 2002), a chronic, high morbidity, low mortality disease which ultimately leads to reduced feed conversion efficiency. The result is reduced output of product to market and massive economic losses to the agricultural industry estimated in the billions of dollars per annum (Clark *et al*., 1991). The primary etiological agent of porcine mycoplasmal pneumonia is the genome-reduced bacterium, *Mycoplasma hyopneumoniae*.

1.1. The Mollicutes

M. hyopneumoniae belongs to a distinct class of cell wall-less micro-organisms known as the mollicutes (Latin: *mollis* meaning 'soft', *cutis* meaning 'skin'), which encompasses mycoplasmas, phytoplasmas and spiroplasmas, and are often interchangeably referred to simply as mycoplasmas (Razin and Hayflick, 2010). The mollicutes are widespread in nature, and colonise plants and animals, including insects, reptiles, fish and humans. Most occur as commensals, organisms that colonize a host without causing pathogenesis, and although many cause disease, particularly the mycoplasma species, little is known about the mechanisms of pathogenicity. The mollicutes classically lack a cell wall (and any genes required for cell wall synthesis) and have very small genomes, displaying an apparent lack of redundancy and

accordingly are often referred to as ‘minimal genomes’ or ‘genome reduced organisms’. As a result, the mollicutes have stimulated the basis of research into the development and recent successful creation of the first “synthetic genome” by the J. Craig Venter Institute (Gibson *et al*., 2008; Gibson *et al*., 2010). The genomes of the mollicutes often lack genes involved in amino acid biosynthesis and may be totally dependent on exogenous sources of amino acids. Mycoplasmas also lack a functional tricarboxylic acid (TCA) cycle, cannot synthesise many essential lipids, including cholesterol, and rely on importation of nucleotides for the synthesis of DNA (Fraser *et al*., 1995; Guimaraes *et al*., 2011; Jaffe *et al*., 2004b). Consequently, these organisms possess inefficient energy-yielding pathways and are extremely fastidious in their nutritional requirements (Dybvig and Voelker, 1996). This phenomenon is thought to be the result of degenerative evolution, presumably arising from a lifestyle that is predominantly parasitic in nature (Razin and Hayflick, 2010; Razin *et al*., 1998). Individual species thus tend to be highly host- and site-specific. For example, *Mycoplasma pneumoniae* infects the respiratory tract of human hosts, resulting in primary atypical pneumonia (walking pneumonia) and *Mycoplasma genitalium* infects the urogenital tract of humans, despite large similarities in their genomes. *Mycoplasma hyopneumoniae* infects the respiratory tract of swine exclusively (Razin *et al.*, 1998; Zielinski and Ross, 1993).

The mollicutes are amongst the smallest free-living organisms (Figure 1.1), many less than 1 μm in diameter, and for many years were thought to be viruses due to their ability to pass though bacterial filters and the difficulty to culture *in vitro* because of complex nutritional requirements (Razin and Hayflick, 2010; Razin *et al.*, 1998). It is only through recent developments in the areas of genomics and ribosomal DNA sequencing that the species classed as mollicutes were defined as a group of eubacteria, phylogenetically related to gram-positive bacteria, having evolved through a process of reductive evolution from the firmicutes. Mollicute genomes have characteristically low G + C content in the range of 24 – 33 mol%, and unlike other bacteria, the TGA codon does not encode a stop codon, but a tryptophan, except in members of the genera *Phytoplasma* and *Acholeplasma* (Inamine *et al*., 1990; Osawa *et al*., 1992; Razin *et al.*, 1998). This difference in the genetic code creates problems when genetically manipulating the organism, particularly when expressing cloned genes in heterologous systems such as *E. coli* resulting in prematurely truncated proteins. Accordingly, cloned

mycoplasmal genes were in the past required to be expressed in UGA suppressor strains to study protein function (Dybvig and Voelker, 1996). To alleviate this problem, TGA codons can be readily converted to tryptophan-encoding TGG codons.

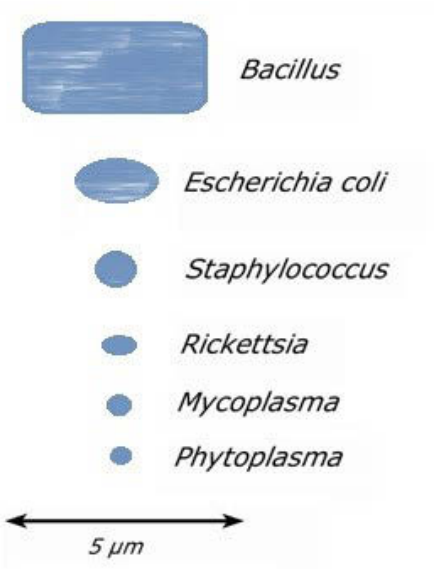


Figure 1.1. Diagrammatic representation of size comparisons of bacteria and mollicutes.

The mollicutes are bounded only by a single plasma membrane, lacking a cell wall and intracytoplasmic membranes. Consequently, all proteins that are not cytosolic are either membrane bound or secreted. Mollicutes possess eubacterial signal peptide sequences that direct proteins into a general secretory pathway for transport across the plasma membrane (Dybvig and Voelker, 1996). However, these signal peptides for the mollicutes have been found to be significantly different from *E. coli* and Gram positive bacteria, differing in their N-terminal charge, peptide length and periodicity of side chain hydrophobicity (Edman *et al*., 1999). Signal peptides for lipoproteins were also found to be longer than for any other bacteria, which may be attributed to a thicker membrane rich in host-derived cholesterol (Edman *et al.*, 1999). The plasma membrane contains almost all mycoplasmal lipid, and like other biological membranes is made up of phospholipids, glycolipids and neutral lipids, although of the total mass of membrane, more than two thirds is protein and lipoprotein (Razin and Hayflick, 2010). The high number of lipoproteins possessed by members of the mollicutes is unusual compared to other bacteria and may be attributed to the lack of a periplasmic space. Lipoproteins are attached to the membrane by a lipid moiety, with the protein fraction

of the molecule on the outside surface. The function of most lipoproteins in the mollicutes is not yet known, however, several of these proteins demonstrate antigenic and/or size variation and may mediate interactions between the mycoplasma and its environment (e.g. ligand binding) (Dybvig and Voelker, 1996). Due to the lack of a cell wall, the membrane is directly in contact with the environment and the host, and membrane proteins are among the most dominant antigens of mycoplasma (Kim *et al*., 1990). Thus, the plasma membrane is of great interest in the study of pathogenesis of mycoplasmas as host-pathogen interactions are mediated primarily through cell-cell contact.

1.2. *Mycoplasma hyopneumoniae* features and pathogenesis

Mycoplasma hyopneumoniae is a respiratory pathogen of swine and is the primary causative agent of porcine enzootic pneumonia (PEP); an economically significant, chronic respiratory disease prevalent in commercial pig farms worldwide (Kolodziejczyk and Pejsak, 2004). *M. hyopneumoniae* adheres directly to cilia and colonizes the respiratory tract, where it causes ciliostasis and epithelial cell death, causing characteristic lesions associated with PEP. Colonization also disrupts the mucocilliary escalator, resulting in a chronic infection, which reduces food conversion efficiency and increases the animal's susceptibility to secondary infections which can result in further pathogenesis (DeBey and Ross, 1994; Kolodziejczyk and Pejsak, 2004) (Figure 1.2). Coughing is a characteristic symptom of infection and the bacteria can be aerosolized in mucous droplets, leading to rapid dissemination throughout pig herds, particularly in intensive rearing environments, but the droplets are also capable of being carried distances of up to 4.7 km to neighbouring farms (Otake *et al*., 2010). The occurrence of PEP is currently controlled in commercial pig farms through the use of vaccines and antibiotics to prevent or treat secondary infections, however such measures are often ineffective and may promote the evolution of antibiotic resistant bacteria (Haesebrouck *et al*., 2004). In Australian pig farms, infections attributed to *Lawsonia*, *Mycoplasma* and *Escherichia coli* motivated the most antimicrobial usage (Jordan *et al*., 2009). There is a strong international consensus that antimicrobials in agriculture should be used judiciously to minimise the risk of the development of antibiotic resistant bacteria (Jordan *et al.*, 2009). Current commercial vaccines developed to combat PEP are comprised of whole cell bacterin formulations based on killed or attenuated whole

bacteria and although they alleviate some of the losses associated with reduced growth rate, they are costly to produce and induce the production of serum antibodies to provide limited protection against the symptoms of infection, however they do not prevent colonization (Haesebrouck *et al.*, 2004; Thacker *et al*., 1998a and b; Woolley *et al*., 2014). In addition to this, as mycoplasmas characteristically lack a cell wall, they also lack peptidoglycan and lipopolysaccharides (LPS) which are highly antigenic and are often targets for antimicrobial therapies. Hence *M. hyopneumoniae* can evade elimination by antibacterial agents that target cell wall synthesis (Razin *et al.*, 1998). Currently quinolones, which inhibit the activity of DNA gyrase, which is essential for DNA replication in bacteria, are the most effective antibiotics against mycoplasmas (Razin *et al.*, 1998). 31 tonnes of antimicrobials were sold for therapeutic use in swine in 2001-2002 according to the last reported figures in 2005 by the Australian Pesticides & Veterinary Medicines Authority (APVMA, 2005). Fortunately, Australian pig farmers tend to rely most on antimicrobials that are of low importance to human medicine, such as sulphonamides, tetracyclines and penicillins (Jordan *et al.*, 2009). The sustained use of high volumes of antibiotics in food production results in the release of partially-metabolised antibiotics into the environment via animal waste. These environmental pollutants may subsequently contaminate waterways or pastures where they drive the capture of antibiotic resistance genes from complex microbial communities in soil and aquatic environments onto mobile elements that circulate between food animal and human populations (Gillings *et al*., 2014; Jechalke *et al*., 2014). This use of antibiotics has been shown to be a driver of antibiotic resistance in pig-associated bacteria such as *Campylobacter spp*, enterococci and *E. coli* which may also directly or indirectly affect human health (Barton, 2014; Faldynova *et al*., 2013; Vicca *et al*., 2007). The development of novel, effective vaccines and therapeutics to combat *M. hyopneumoniae* infection could reduce not only the cost of raising food producing animals, but also decrease the use of antibiotics. An improved understanding of pathogenesis and the pathogen itself is necessary to enable rational drug design and/or vaccine development with greater efficacy and fewer undesirable side effects.

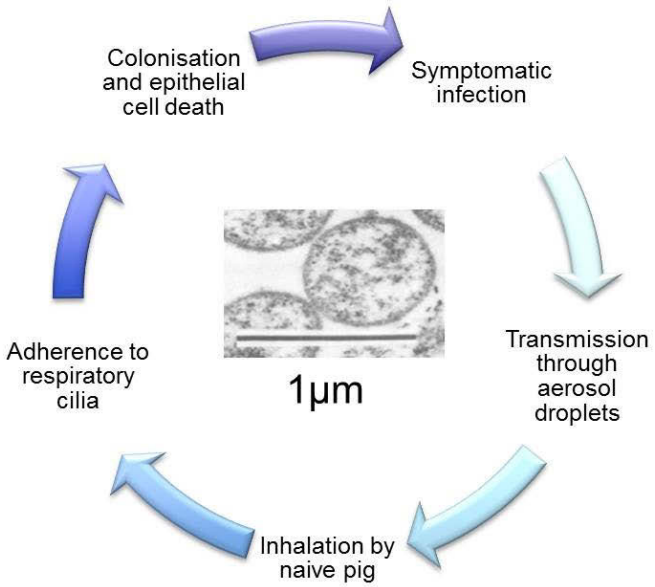


Figure 1.2. Cycle of pathogenicity of *M. hyopneumoniae.*

Our understanding of many organisms has been bolstered by the ability to sequence the genome, allowing for prediction of genes and potentially gene functions. The genomes of four, geographically-distinct strains of *Mycoplasma hyopneumoniae* have been completely sequenced; including the virulent strains 168 from China, 232 from the United States and 7448 from Brazil, and the avirulent strain J (ATCC 25934) (Liu *et al*., 2011; Minion *et al*., 2004; Vasconcelos *et al*., 2005). These genomes are approximately 900 kbp and contain less than 700 predicted coding sequences (Minion *et al.*, 2004). Of the predicted open reading frames (ORFs), around 60% show homologies to known proteins leaving approximately 40% of the predicted proteome comprised of putative uncharacterised proteins. Genomic comparisons of the virulent versus avirulent strains have revealed variations between orthologous proteins in the number of amino acid repeats they contain (de Castro *et al*., 2006). These insertions/deletions result from variations in the number of tandem nucleotide repeats within coding regions, which is indicative of a molecular mechanism generating functional and/or antigenic variants. Such variation in surface proteins is likely to be a key determinant of different pathogenic properties of each *M. hyopneumoniae* strain (Vasconcelos *et al.*, 2005). Phase variation has been previously demonstrated in other members of the Mollicutes and is particularly well documented to occur through stochastic expression patterns of diverse lipoprotein genes in the *Mycoplasma mycoides* cluster of mycoplasmas (Wise *et*

al., 2006). As of yet, however, there is not sufficient evidence to suggest phase variation occurs in *M. hyopneumoniae*.

1.2.1. Adhesin proteins of M. hyopneumoniae

Mycoplasma hyopneumoniae is poorly suited to survival outside of the host, as it relies on the scavenging of host components that it is unable to produce itself, such as cholesterol which is required for the stability of the cell membrane. In order to survive as a species, it depends on its ability to establish an infection through the direct binding of respiratory cilia, colonising the respiratory tract and thus persisting in the host. Porcine cilia beat at a frequency of ~11-15 Hz (Joki and Saano, 1994), and are designed to clear bacteria and debris from the respiratory tract. In addition, unlike other adherent bacteria, *M. hyopneumoniae* is pleomorphic in structure and does not possess projections or appendages such as flagella or pili which classically aid in motility and adhesion (Layh-Schmitt *et al*., 2000; Minion *et al*., 2000). Given all this, the ability to directly bind to and colonize the host respiratory cilia is a testament to the evolution of the highly-specialized proteins that are responsible for initial binding events (Figure 1.3).

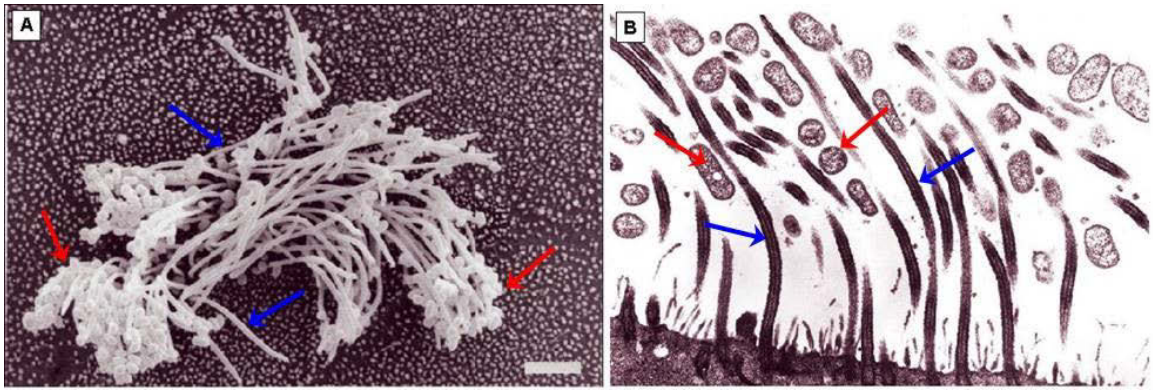


Figure 1.3. Microscopy of *Mycoplasma hyopneumoniae* binding cilia.
A. Scanning electron micrographs showing adherence to cilia *in vitro* of large numbers of pathogenic *M. hyopneumoniae* (bar = 2 μm) (adapted from Young *et al*. (2000)). **B.** *Mycoplasma hyopneumoniae* attaching to ciliated epithelial cells in the respiratory tract of infected swine. Image by F. Chris Minion (Bentley *et al*., 2005). Red arrows indicate mycoplasma cells, blue arrows indicate cilia.

It was established in 1990 that the nature of the bacterial adhesins involved in host-pathogen interaction were likely to be protein in nature, due to the decrease in binding to epithelial cell monolayers and cilia seen when intact mycoplasma cells were pre-treated with the proteolytic enzyme, trypsin (Zielinski *et al*., 1990). In 1995 the first adhesin protein was identified through antibody affinity chromatography and inhibition studies using monoclonal antibodies (MAbs) raised against various *M. hyopneumoniae* antigens. A 97 kDa protein, called P97 (Mhp183) reacted with monoclonal antibodies F2G5 and F1B6 and inhibited adherence of mycoplasma cells to porcine cilia by up to 67% (Zhang *et al*., 1995). Further, it was identified to be surface-exposed by immunogold labelling and scanning electron microscopy (Figure 1.4). It is now understood that P97 is a key mediator of cilium adherence in *M. hyopneumoniae*. The P97 protein contains two tandem repeat regions; R1, comprised of tandem pentapeptide repeats with the sequence A-A-K-P-V-(E); and R2 comprised of tandem decapeptide repeats with the sequence G-T-P-N-Q-G-K-K-A-E. A study by Hsu *et al*. concluded that the R1 region was critical to cilium adhesion while the R2 region did not influence ciliary binding (Hsu and Minion, 1998a). Further, through the creation of fusion proteins with a series of R1 repeats containing different numbers of repeating units, it was shown that a minimum of eight consecutive R1 repeated units are required for cilium binding whilst only three repeating units are required for antibody recognition (Minion *et al.*, 2000). However, more recent studies have found that both R1 and R2 are necessary to bind to the glycosaminoglycan, heparin (the significance of this is discussed later), and it has been suggested that R2 is required for conformational stability and not its amino acid motif (Jenkins *et al*., 2006). A comparison of the P97 R1 regions in strains of *M. hyopneumoniae* with varying degrees of virulence failed to identify any differences that could account for the reduced adherence of the avirulent strains; all strains had more than the minimum of eight repeats required for cilium binding, and as such it must be concluded that P97 is intricately, but not exclusively, involved in the process of adherence (Hsu and Minion, 1998a; Jenkins *et al.*, 2006).

The *p97* gene occurs in a two-gene operon with partner *p102*, which encodes a 102 kDa protein (P102). Due to the close genetic linkage with P97, P102 was originally predicted to play some role in adherence either directly by interacting with host surface structures or indirectly through a supportive function for P97 activity (Adams *et al*., 2005; Hsu

and Minion, 1998b). To test this, P102 expression was examined *in vivo* by Adams *et al*. using immunogold labelling and electron microscopy, performed on respiratory tissues from infected pigs. From this, it was concluded that P102 is expressed *in vivo* and is secreted, as gold particles could be seen within mycoplasmas and attached to swine cilia, often in aggregates in high concentrations (Figure 1.4) (Adams *et al.*, 2005). In addition, it did not appear that P102 interacts directly with the R1 cilium-binding region of P97 found exclusively along the mycoplasma membrane, suggesting an indirect mechanism for affecting virulence (Adams *et al.*, 2005; Djordjevic *et al*., 2004; Hsu *et al*., 1997).

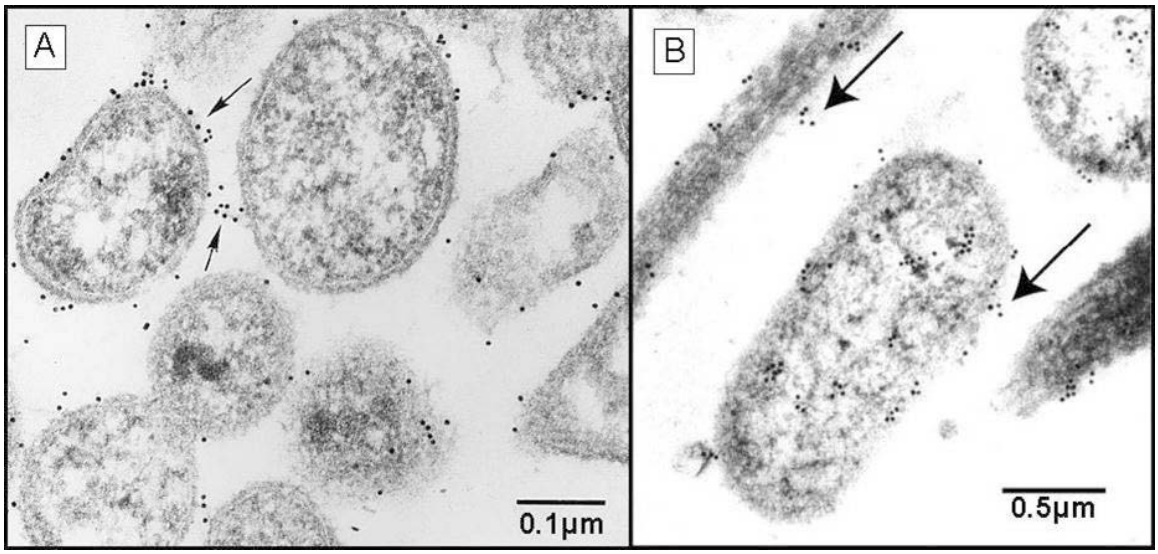


Figure 1.4. Immunogold labelling and electron microscopy, showing the localization of adhesins, indicated by black arrows.

A. *Mycoplasma hyopneumoniae* reacted with monoclonal antibody against P97 (F2G5) and gold conjugates (adapted from Zhang *et al.* (1995)); **B.** Infected trachea from pigs challenged with strain 232. Regions containing ciliated epithelium were sectioned and stained with gold-labelled mouse anti-P102 antibodies (adapted from Adams *et al.* (2005)).

Genome analysis of *Mycoplasma hyopneumoniae* strain 232 in 2004 suggested that adhesion could indeed be attributed to interactions other than P97, through the discovery of the P97 and P102 paralogous gene families of adhesins. A BLAST-P search discovered that there are six paralogs of P97 and six paralogs of P102 in the sequenced genomes, with >30% amino acid identity over 70% of the length and many of these occur as two-gene operons containing a *p97* and *p102* paralog (Figure 1.5) (Adams *et al.*, 2005; Minion *et al.*, 2004). At the time the prevailing hypothesis was that P97 was the archetype cilium adhesin because it was the only gene in the genome of *M. hyopneumoniae* that expressed an adhesin protein with the requisite number of pentapeptide repeats in the R1 cilium-binding domain. The gene products are collectively referred to as adhesins and are highly conserved in the genomes of the four strains of *M. hyopneumoniae* so far sequenced (Table 1.1). Many members of these families are prominently expressed during broth culture and mRNA transcripts encoding most of these adhesins have been detected in *M. hyopneumoniae* recovered from infected pigs (Adams *et al.*, 2005), demonstrating expression *in vivo*. As *M. hyopneumoniae* is incapable of surviving for long periods outside the host, the initial adherence and colonisation steps are critical for survival, thus from an evolutionary standpoint, redundancy in adhesive capacity is likely essential to survival of the species.

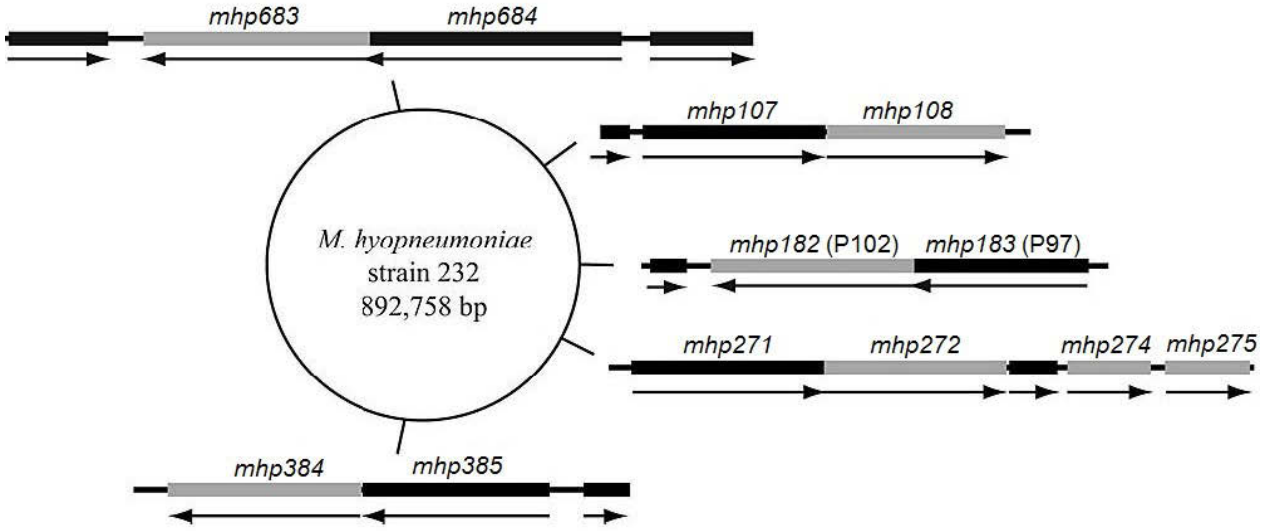


Figure 1.5. Paralog gene structure and genome location in *M. hyopneumoniae* strain 232.

Grey bars indicate P102 gene paralogs, and black bars indicate P97 gene paralogs. *mhp182* is the P102 gene while *mhp183* is the gene for P97. The arrows indicate direction of transcription. Not shown are *mhp280* (encoding P95) and *mhp493* (encoding P216) which do not occur as two-gene operons. Adapted from Adams *et al.* (2005).

Members of the adhesin families have predicted molecular weights ranging from ~100 to 216 kDa; however they have subsequently been shown through the use of immunoblotting and mass spectrometry to be extensively post-translationally cleaved, sometimes in a strain-specific manner (discussed below). Cleavage products of all the adhesins have been found to be enriched in hydrophilic amino acids, soluble in non-ionic detergents and lack any known membrane anchorage motifs. However it has also been demonstrated that cleavage products are prominently displayed on the cell surface and remain associated with the surface despite rigorous washing cycles to remove media derived contaminants (Burnett *et al*., 2006; Djordjevic *et al.*, 2004; Seymour *et al*., 2010; Wilton *et al*., 2009; Zhang *et al.*, 1995).

Table 1.1: Members of the P97/P102 paralog family of adhesins and their equivalent genes in strains J and 232.

P97 family		**P102 family**	
Strain 232	**Strain J**	**Strain 232**	**Strain J**
Mhp183 (P97)	*MHJ_0194*	*Mhp182* (P102)	*MHJ_0195*
Mhp107	*MHJ_0264*	*Mhp108*	*MHJ_0263*
Mhp271	*MHJ_0105*	*Mhp272*	*MHJ_0104*
Mhp684 (P146)	*MHJ_0663*	*Mhp683*	*MHJ_0662*
Mhp385	*MHJ_0369*	*Mhp384*	*MHJ_0368*
Mhp493 (P216)	*MHJ_0493*	*Mhp275*	No gene
Mhp280	*MHJ_0096*	*Mhp274*	No gene

1.2.2. Proteolytic cleavage of adhesins

In the initial investigation into P97, immunoblots were performed using the monoclonal antibody F2G5 against *M. hyopneumoniae* whole cell lysates (Figure 1.6C). The appearance of multiple reactive bands in the 1D immunoblot was hypothesised to be unlikely due to degradation of P97 during sample handling as protease inhibitors were used, but rather due to the presence of the epitope in multiple antigens (Zhang *et al.*, 1995). Almost ten years later, it was confirmed that the appearance of multiple bands was a legitimate finding, with the discovery that the adhesin undergoes proteolytic processing and is displayed on the cell surface (Djordjevic *et al.*, 2004).

Genomic analyses revealed the gene encoding the P97 cilium adhesin actually encodes a ~120 kDa pre-protein (Hsu and Minion, 1998b; Minion *et al.*, 2004; Vasconcelos *et al.*, 2005). Through one-dimensional gel electrophoresis in combination with immunoblotting and Edman N-terminal sequencing to identify cleavage sites, it was demonstrated that this pre-protein is post-translationally cleaved into dominant fragments of 22, 28, 70 and 97 kDa (Figure 1.6). Endoproteolytic cleavage at amino acid position 195 removes an amino-terminal fragment of approximately 22 kDa ($P22_{P97}$) and generates the mature cilium adhesin known as P97. Amino-terminal

sequence analysis of P97 from strains J and 232 detected the sequence[1] ^{195}ADEKTSS^{201} and further studies identified that an identical sequence is present in the cilium adhesin paralog Mhp271 (Deutscher *et al*., 2010; Djordjevic *et al.*, 2004). Consistent with this, P22 has been detected in cell lysates of *M. hyopneumoniae* strains J and 232 (Djordjevic *et al.*, 2004). The only putative transmembrane domain in P97 occurs between amino acids 8 and 22 (TMpred score 2202) and this region is likely to be critical for targeting the P124 pre-protein to the general secretion machinery (Figure 1.7). Edman sequencing of the amino-terminal 22 kDa cleavage fragment (P22) detected the sequence ^{2}SKKSKTF^{8} (Djordjevic *et al.*, 2004). Although *M. hyopneumoniae* contains a gene which has a signal peptidase I signature motif (Minion *et al.*, 2004), existing biochemical data from amino-terminal sequence analysis of cleavage products indicates that this species lacks SPase I activity, at least in so far that cleavage at the A-X-A site immediately following the signal sequence does not occur. We cannot exclude the possibility that SPaseI plays a role in the processing of proteins with a secretion signal by cleaving at sites with different recognition motifs. Clearly, SPase I activity is not required to cleave the signal peptide during processing of P124/P97, but other processing events take place, presumably during or immediately after secretion. A carboxyl-terminal fragment of the cilium adhesin of approximately 28 kDa ($P28_{P97}$) is also removed by an endoproteolytic cleavage event. Edman analysis detected the amino-terminal sequence ^{863}NTNTGFS^{869}. The lack of sequence similarity spanning the two endoproteolytic cleavage sites suggests that these cleavage events are executed by different proteases, a hypothesis that has since been further supported by investigations of processing patterns in other adhesins (Bogema *et al*., 2012; Bogema *et al*., 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al*., 2012; Djordjevic *et al.*, 2004; Seymour *et al*., 2011; Seymour *et al*., 2012; Wilton *et al.*, 2009). Cleavage fragments derived from P97 have been located on the surface of *M. hyopneumoniae* cells by immunogold labelling (Adams *et al.*, 2005; Djordjevic *et al.*, 2004; Hsu *et al.*, 1997; Hsu and Minion, 1998a; Minion *et al.*, 2000; Zhang *et al.*, 1995). Cleavage fragments are also detected in cell-free culture supernatants (See Chapter 4).

[1] Superscript numbers refer to the amino acid position in the complete translated pre-protein and are used throughout this thesis.

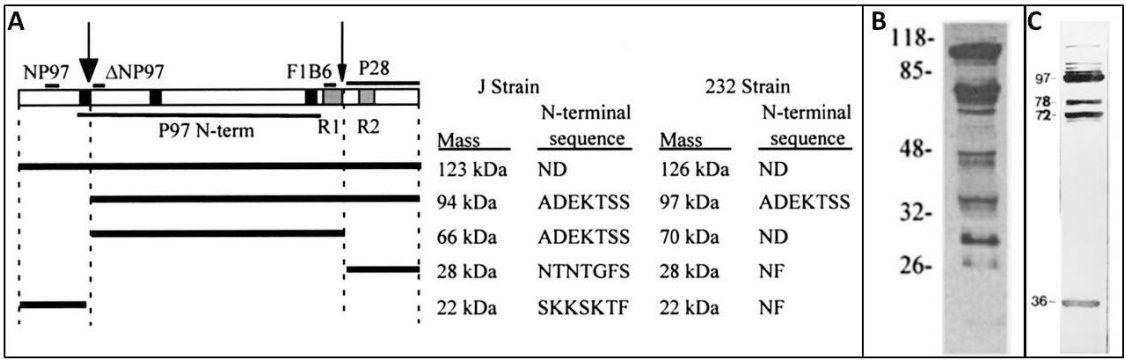


Figure 1.6. Map of cleavage of the P97 cilium adhesin and immunoblot analyses using MAb F1B6 and F2G5.

(A) The map shows antibody epitope locations, repeat regions, and selected cleavage sites identified by peptide mass fingerprint analysis, fragment masses, and N-terminal sequences. The major cleavage event at amino acid 195 (large arrow) and another event at amino acid 891 (small arrow) are shown. The locations of the R1 and R2 repeat regions are represented by gray-shaded boxes. The locations of the epitope for MAb F1B6, ΔNP97 peptide, and NP97, P97 N-terminal (P97 N-term), and P28 antisera are shown as bars above or below the map. The coiled-coil regions are represented by the black boxes. The cleavage products of strains J and 232 are shown as gray-shaded bars below the map. Their molecular masses and N-terminal sequences are shown to the right of the map. NF: not found; ND: not determined. (B) 1D immunoblot using MAb F1B6 against *M. hyopneumoniae* whole cell lysates performed by Djordjevic *et al.* (C) 1D immunoblot using MAb F2G5 against *M. hyopneumoniae* whole cell lysates performed by Zhang *et al*. Adapted from (Djordjevic *et al.*, 2004; Zhang *et al.*, 1995).

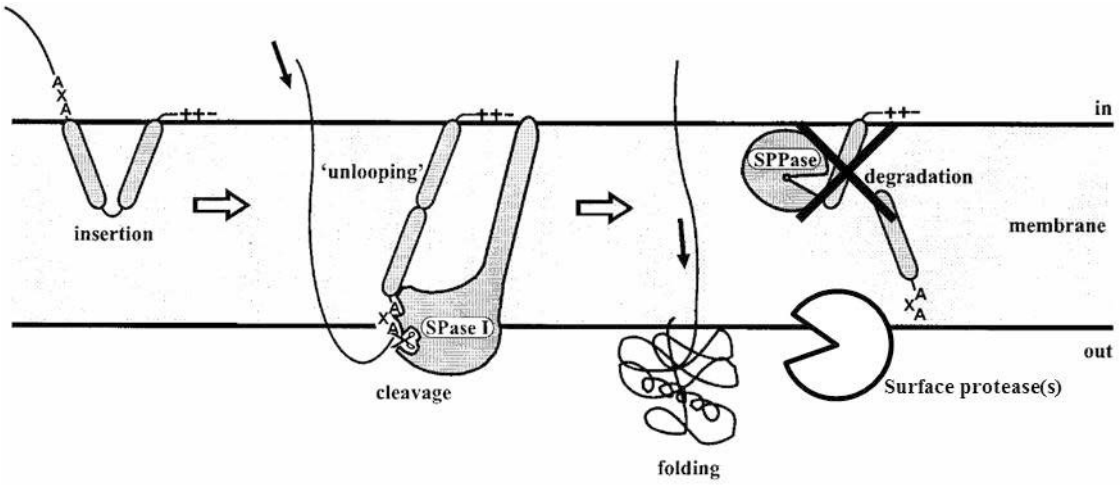
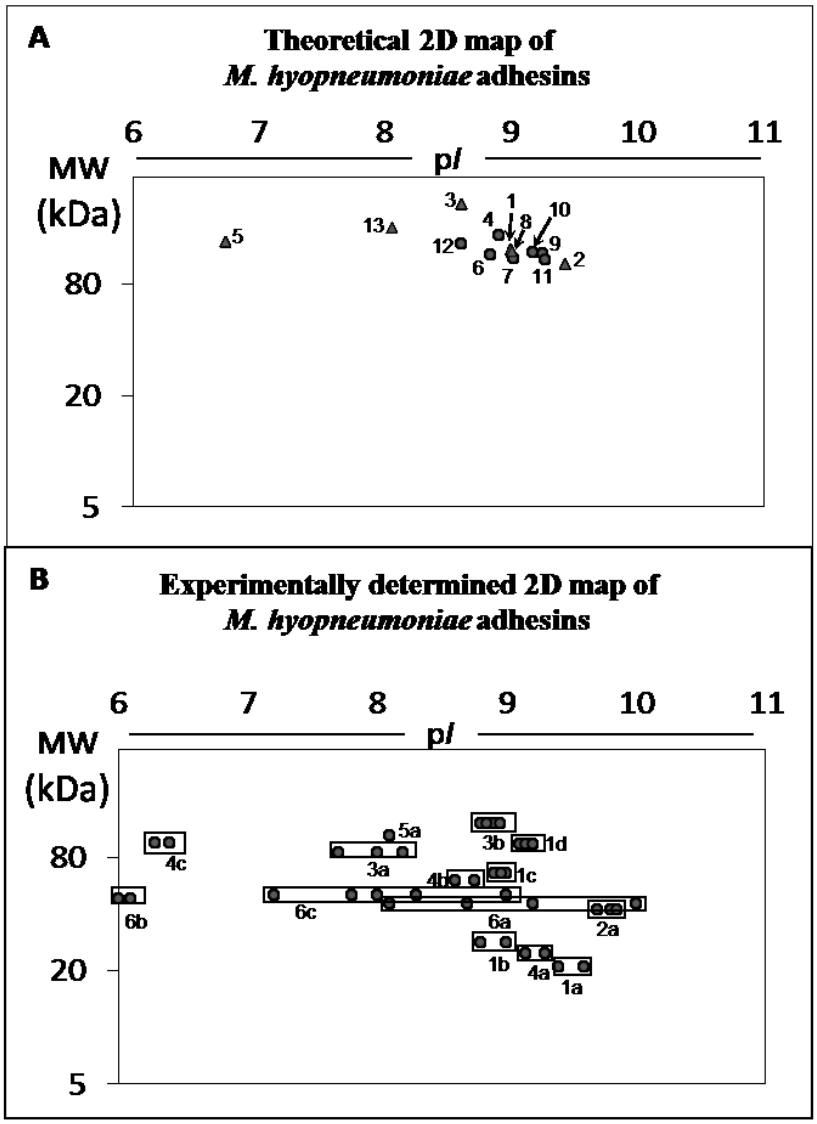


Figure 1.7. Model for signal peptide insertion into the cytoplasmic membrane and cleavage by SPase I.

This is the general secretion mechanism, similar to the proposed mechanism for export of adhesins of *M. hyopneumoniae.* First, the positively charged N-domain of the signal peptide interacts with negatively charged phospholipids in the membrane, after which the H-domain integrates loopwise into the membrane. Next, the H-domain unloops, whereby the first part of the mature protein is pulled through the membrane. During or shortly after translocation by a translocation machinery (not shown), the signal peptide is cleaved by SPase I and subsequently degraded by signal peptide peptidases (SPPases). After its translocation across the membrane, the mature protein folds into its native conformation. Adapted from Tjalsma *et al*. (2000). There is no evidence that degradation by SPPases occurs in *M. hyopneumoniae*, but rather, the adhesin is then proteolytically processed into functional domains, in a manner akin to ectodomain shedding in eukaryotes.

Djordjevic *et al.* also demonstrated the cleavage of P102 into 72 and 42 kDa fragments; showing that complex processing is protein specific. The 42 kDa protein represented the C-terminal cleavage fragment of P102 (Djordjevic *et al.*, 2004). It has been assumed that proteolytic activity on the mycoplasma membrane surface is selective, and although the mechanisms for this remain unclear, the discovery of the specific cleavage sites and the identification of other post-translationally cleaved surface proteins and proteases may aid in the elucidation of this process.

It has now been shown that many members of the P97 and P102 paralogous families of adhesins are post-translationally cleaved into functional domains such that they are retained on the surface of the cells harvested from broth culture, but the mechanism enabling this to occur is not understood (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Raymond *et al*., 2014; Raymond *et al*., 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al*., 2014; Wilton *et al.*, 2009). Cleavage fragments display multifunctional binding abilities for extracellular matrix components and other key host circulatory molecules, including various glycosaminoglycans (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Jenkins *et al.*, 2006; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Wilton *et al.*, 2009; Zhang *et al.*, 1995), fibronectin (Deutscher *et al.*, 2010; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012) and plasminogen (Bogema *et al.*, 2012; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012). Figure 1.8 depicts how proteolytic processing has a dramatic effect on how adhesin proteins migrate in 2D gels compared to their predicted migration patterns based on sequence data. In addition to proteolytic processing, other PTMs such as phosphorylation or deamidation can cause a shift in p*I*, and where this is not present in all proteins or fragments, this may result in spot trains across p*I* units. Phosphorylation has been shown to occur in P216 previously (Wilton *et al.*, 2009).



Legend: (Panel A)

1. P97 (P94)
2. P102
3. P216
4. P146
5. Mhp683
6. Mhp385
7. Mhp384
8. Mhp107
9. Mhp108
10. Mhp271
11. Mhp272
12. P95
13. P159

Legend: (Panel B)

1. P97 (P94) (Djordjevic 2004)
   - 1a. $P22_{P97}$
   - 1b. $P28_{P97}$
   - 1c. $P66_{P97}$
   - 1d. $P94_{P97}$
2. P102 (Djordjevic 2004)
   - 2a. $P42_{P102}$
3. P216 (Wilton 2009)
   - 3a. $P85_{P216}$
   - 3b. $P120_{P216}$
4. P159 (Burnett 2006)
   - 4a. $P27_{P159}$
   - 4b. $P52_{P159}$
   - 4c. $P110_{P159}$
5. Mhp107 (Seymour 2011)
   - 5a. $P104_{Mhp107}$
6. Mhp683 (Bogema 2011)
   - 6a. $P45_{683}$
   - 6b. $P48_{683}$
   - 6c. $P50_{683}$

Figure 1.8. Diagrammatical illustration of the effect of post-translational modifications of adhesin molecules of *M. hyopneumoniae*.

Panel A shows a computationally composed theoretical 2D gel map of *M. hyopneumoniae* adhesins based on *in silico* predicted ORFs. Panel B shows a computationally composed 2D gel map of *M. hyopneumoniae* adhesins based on experimentally determined isoelectric points and molecular mass of adhesins in panel A denoted by triangular markers. From Djordjevic and Tacchi (2014).

Considerable progress has been made in deciphering proteolytic cleavage motifs that reside within the P97 and P102 paralog families. In a landmark study, Bogema *et al.* (2011) identified an amino acid sequence motif TTKF↓QE in the P102 paralog Mhp683 that resided in two regions of the molecule where cleavage was determined to occur. A combination of Edman sequencing and LC-MS/MS was used to decipher the cleavage site at the phenylalanine residue in both these motifs that produced the three cleavage fragments $P45_{683}$, $P48_{683}$ and $P50_{683}$ (Bogema *et al.*, 2011). The cleavage site was refined to S/T-X-F↓X-D/E by examining cleavage data (Edman sequence and peptide identifications by LC-MS/MS) for all members of the P97 and P102 paralogs and searching for atypical tryptic peptides (semitryptic peptides) that reside within regions of these molecules that span known cleavage sites (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Seymour *et al.*, 2012; Wilton *et al.*, 2009). It has since been shown that efficient cleavage occurs at S/T-X-F↓X-D/E motifs found in most of the P97 and P102 adhesin families and in P159, and has been designated as the "dominant" adhesin cleavage motif (Bogema *et al.*, 2011; Deutscher *et al.*, 2012). Peptides spanning these motifs are never identified and proteins are very rarely identified at their intact ORF mass, hence cleavage is considered efficient. It has been demonstrated that these motifs typically reside within disordered regions of more than 40 amino acids (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2012; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Tacchi *et al.*, 2014), suggesting that disorder can influence the accessibility of cleavage motifs in the pre-protein to proteolytic cleavage. Regions of intrinsic protein disorder are regions which do not have defined or rigid 3D structures under physiological conditions *in vitro* and are thus hypothesised to allow for increased flexibility in the protein and thereby improved accessibility of the cleavage site (Dunker *et al*., 2005; Uversky and Dunker, 2010). This discovery has been instrumental in discovering and accurately mapping potential cleavage sites in the adhesins and other proteins (discussed further in Chapters 3 and 5), as protein disorder may be predicted using bioinformatics algorithms such as PONDR (Linding *et al*., 2003; Xue *et al*., 2010) (Table 1.2).

Other efficient cleavage sites have been mapped in P97 at T-N-T↓N-T-N (Djordjevic *et al.*, 2004), in P159 at L-K-V↓G-A-A (Raymond *et al.*, 2013) and in P97 paralog Mhp385 at L-N-V↓A-V-S (Deutscher *et al.*, 2012). Trypsin-like cleavage events in

members of the P97 and P102 adhesin families and in P159 have also been characterised (Moitinho-Silva *et al*., 2013; Raymond *et al.*, 2013). A dominant cleavage event has also been shown to occur within a putative transmembrane domain in the N-terminus of P216 (See Chapter 5) with sequence ^{7}T-L-L↓L↓A↓T↓A↓A-I-I-G-S-T-V-F-G-T-V-V-G-L-A-S^{30}. Consecutive cleavage events at positions L10, A11, T12, A13 and A14 are indicative of aminopeptidase activity. Aminopeptidase activity was observed at a number of putative endoproteolytic cleavage sites in P216 suggesting that *M. hyopneumoniae* expresses several aminopeptidases on its cell surface (See Chapter 5).

It should be noted that host-derived proteases are not responsible for these cleavage events as they are consistently described in cultured bacteria. The Friis media used in culture is unlikely to possess protease activity as media incubated with recombinant P97 adhesin displayed no cleavage activity (Djordjevic *et al*, 2004). In addition, inefficient cleavage sites also exist (mostly as sub-fragments) where peptides spanning the proposed cleavage site are sometimes identified and peptides mapping to the protein are identified at both intact (or dominant fragment) and fragment (or sub-fragment) masses.

Table 1.2. Adhesin proteins of *M. hyopneumoniae* strain J and respective cleavage fragments so far identified.

Protein	J	Intact MW /pI	Fragments	Fragment location	Features	References
P97 Paralogs						
P97	MHJ_0194 Q4AAD6	124903.59 9.50	P22	N-term (cleaved)	Leader sequence	(Djordjevic *et al.*, 2004; Jenkins *et al.*, 2006; Okamba *et al.*, 2007; Raymond *et al.*, 2014)
			P94	Full length	Intact (- leader sequence)	
			P66/P70	N-term	Binds heparin	
			P28	C-term	TNT//NTNT	
P216*	MHJ_0493 Q4A9J2	215649.06 9.15	P120	N-term	Binds heparin	(Tacchi *et al.*, 2014; Wilton *et al.*, 2009)
			P85	C-term	Phosphorylated serine Binds heparin	
			Extensive minor cleavage and aminopeptidase activity*			
P146/ Mhp684	MHJ_0663 Q4A925	148202.35 9.32	P50	N-term	Immunogenic, binds cilia and heparin TYF//AE putative site, 1 TMD	(Bogema *et al.*, 2012)
			P40	Mid	PQ repeat region	
			P85	C-term	TEF//QQ semitryptic ID PS and S repeat regions, 3 coiled-coil domains Binds cilia, plasminogen and heparin	
			P25N (P50N)	P50 N-term	TKSF//QT inefficient site	
			P25C (P50C)	P50 C-term		

			P50minor (P85N)	P85 N-term	PS repeat region 1 coiled coil domain	
			P35 (P85C)		2 coiled coil domains, S repeat region	
Mhp385	MHJ_0369 Q4A9W4	114786.69 9.38	P115	Full length preprotein	Novel motif identified – LNV//AVS Reduced efficiency of cleavage Binds heparin	(Deutscher *et al.*, 2012)
			P88	N-term		
			P27	C-term		
Mhp107	MHJ_0264 Q4AA66	119854.64 9.55	23.4 kDa	N-term	TNF//SDQ efficient cleavage Binds heparin	(Deutscher *et al.*, 2012; Seymour *et al.*, 2011)
			96.4 kDa	C-term		
Mhp271	MHJ_0105 Q4AAM4	118835.63 9.72	Intact	Full length	Intact Binds heparin, fibronectin and porcine cilia	(Deutscher *et al.*, 2010)
			P96.6	C-term	Cleavage like P97	
P95	MHJ_0096 Q4AAN1	131553.36 9.23	-	-	-	ND
P102 Paralogs						
P102	MHJ_0195 Q4AAD5	102343.99 10.03	P60	N-term		(Djordjevic *et al.*, 2004; Seymour *et al.*, 2012)
			P42	C-term		
P76/ P159**	MHJ_0494 Q4A9J1	158676.10 9.07	P27	N-term	1 TMD (Probable motif: TEF//QQQ)	(Burnett *et al.*, 2006)
			P110	Mid	(Probable motif: TTF//QDD)	

					Binds PK15 monolayers and heparin	
			P52	C-term	Binds PK15 monolayers and heparin	
			P76 (3)	Extensive minor cleavage fragments and aminopeptidase activity		(Raymond *et al.*, 2013)
			P35 (7)			
			P50 (8)			
			P70 (9)			
P135/ Mhp683	MHJ_0662 Q4A926	134633.60 7.89	P45	N-term	Motif identified – TTK//FQE Binds heparin	(Bogema *et al.*, 2011)
			P48	Mid		
			P50	C-term		
Mhp384	MHJ_0368 Q4A9W5	109224.14 9.53	P60	N-term	Similar motif to Bogema – ILF//NEE Bind heparin	(Deutscher *et al.*, 2012)
			P50	C-term		
P116/ Mhp108	MHJ_0263 Q4AA67	116173.20 9.80	P116	Full length	First to be shown to bind fibronectin and plasminogen	(Seymour *et al.*, 2010)
			17 kDa			
			47 kDa	C-term	Binds fibronectin	
			70 kDa	C-term	Binds fibronectin	
			55 kDa	N-term		
			39 kDa			
Mhp272	MHJ_0104 Q4AAM5	107795.25 9.84	-	-	-	ND

*P216 is described in full detail in Chapter 5.

**P159 shall be included as an adhesin in this thesis due to its close similarity to other recognized adhesin proteins.

It is remarkable that *M. hyopneumoniae* can secrete and process paralogs of the P97 and P102 families without genetic evidence for the presence of genes with significant homology to *spaseI*, *secB*, *secD*, *secE*, *secF* and *secG* or the entire *groEL-groES* operon. Nonetheless, endoproteolytic cleavage fragments of the majority of the P97 and P102 paralog families are prominent components comprising the surface topography of this species. Collectively, these observations suggest that *M. hyopneumoniae* maintains an "adhesion-competent state" during broth culture and that expression of members of these two paralog families are crucial during the infection process.

1.2.3. Processing in other bacteria

In other pathogenic bacteria, large extracellular proteins, particularly those that function as adhesins, are often processed by limited and timely endoproteolysis (Coutte *et al*., 2003; Coutte *et al*., 2001). Processing is considered essential for the maturation of large mass proteins to avoid aggregation on the external membrane surface (Coutte *et al.*, 2001). Like *M. hyopneumoniae,* various *Bordetella* species target the respiratory cilia as the preferred site of colonisation. Adherence to cilia by *Bordetella brochiseptica* is mediated by multiple adhesins to ensure maximal adherence and is a critical early step in pathogenesis (Edwards *et al*., 2005). Mutants of the respiratory pathogen *Bordetella pertussis* deficient in the enzyme responsible for processing the major adhesin, filamentous haemagglutinin, are severely affected in their ability to colonise the murine respiratory tract. Processing involves chaperones and a protease(s) that has access to the proteolytic cleavage site(s), at least for a limited period of time before the molecule refolds on the extracellular side of the cell membrane. Secretion is likely to occur immediately following or concomitantly with translation. Once cleaved, the products either remain attached non-covalently to the external surface of the membrane or are released into the extracellular milieu. It has been suggested that adhesin processing plays a key role in respiratory tract colonisation by assisting in the dispersal of *B. pertussis* from microcolonies (Coutte *et al.*, 2003). It is highly likely that *Mycoplasma hyopneumoniae* follows a similar pattern with regards to adhesin processing and involvement in pathogenesis.

1.2.4. Proteases of M. hyopneumoniae and identification of proteolytic processing

Proteases are enzymes which hydrolyse peptide bonds between amino acids within a protein, constituting an irreversible post-translational modification (PTM). Proteases are classified based on the catalytic site residue (serine, cysteine, threonine, aspartic or glutamic acid), if the mode of action is mediated through coordination with a metal ion (metalloproteases) and the location of the cleavage site within the polypeptide chain. Proteases that cleave at the N-terminus (usually one or two amino acids at a time) are referred to as aminopeptidases, while carboxypeptideases cleave at the C-terminus and endoproteases cleave within the polypeptide chain (Turk, 2006).

The *M. hyopneumoniae* (strain J) genome possesses 18 ORFs that have been annotated in the UniProt database to have putative protease activity. Three of these have poorly predicted functions and five are predicted to have aminopeptidase activity. MHJ_0125 has been characterised as a glutamyl aminopeptidase (Robinson *et al*., 2013) and another, MHJ_0461, has been described as a leucine aminopeptidase (Jarocki *et al*., 2015). Both of these aminopeptidases of *M. hyopneumoniae* have also been shown to carry out moonlighting functions on the cell surface (described later). The remaining ORFs are predicted to have endoprotease activity based on sequence similarity to characterised endoproteases, and these are the likely candidates for carrying out processing at the dominant internal cleavage motifs in the adhesin proteins. Although considerable progress has been made in the identification of cleavage motifs and mapping of cleavage sites in the P97 and P102 adhesins, the endoproteases responsible remain unknown. In addition, aminopeptidase activity has also been described to modify the N-terminal regions of the adhesins (See Chapter 5), and may contribute to modification of new N-termini generated from endoproteolysis occurring at dominant cleavage motifs.

While the function of proteases was initially thought to be restricted to protein degradation, many microbial proteases have been implicated in prominent pathological roles and are recognised as important virulence factors, which is unsurprising given that host immune effector and signalling molecules are often proteins or peptides (Jarocki *et al*., 2014). Through peptide bond cleavage, many pathogens are capable of modulating host innate immune responses, including disruption of cascade systems, such as

complement, coagulation and fibrinolysis (Travis and Potempa, 2000). Some proteases may affect virulence directly by acting as toxins (Tonello *et al*., 1996) or indirectly through the processing of other virulence factors, for example the P97 and P102 paralogous families of adhesins (Coutte *et al.*, 2001). However, as proteases often function within intricate proteolytic networks, their exact targets and biological actions are often difficult to ascertain or even predict. The majority of early discoveries of proteolytic processing have been serendipitous, through the appearance of bands at multiple molecular masses on Western blots, or through untargeted identification of proteins in bands or gel spots following separation by PAGE in protein-centric analyses (auf dem Keller and Schilling, 2010; Davis and Wise, 2002). In this respect, high-throughput peptide-centric proteomics approaches would fail to detect this kind of post-translational modification and many of these fragments would go undiscovered.

The ability to identify post-translational modifications of proteins is a relatively new acquisition to the proteomics "toolbox". The importance of PTMs has increasingly become recognised in relation to biological systems, with apparently minor changes to proteins such as phosphorylation or methylation playing substantial roles in activation or deactivation of protein functions and cascade initiations (auf dem Keller and Schilling, 2010; Lange and Overall, 2013). In particular, the importance of post-translational processing has increasingly become recognised due to the ability to create enormous diversity (Lange and Overall, 2013; Shahinian *et al*., 2013) and contribute to virulence in pathogenic organisms as discussed above. Techniques to identify these modifications began initially with serendipitous findings, which then required follow-up by biochemical analyses performed on purified proteins, however with the progression of high-throughput proteomics, strategies were also developed for enrichment and high-throughput identification of PTMs (auf dem Keller and Schilling, 2010; Lange and Overall, 2013; Shahinian *et al.*, 2013). Unfortunately, the nature of high-throughput proteomics is that it largely utilises "bottom-up", or peptide-centric approaches whereby all proteins in a sample are digested into smaller, more homogenous peptides which provides increased sensitivity and specificity of mass spectrometry analyses. These techniques result in a loss of protein mass context which made identification of genuine proteolytic cleavage events *in vivo* difficult or even impossible to determine, and this in turn led to the under-representation of proteolytic cleavage as a PTM (Lange and

Overall, 2013). Protein-centric methods to detect proteolytic cleavage are predominantly gel-based, with 2D gels providing good resolution of individual proteoforms (Smith *et al*., 2013) which allows identification of cleavage fragments, and one-dimensional gels retaining mass context, however these techniques are limited by technical challenges (Gorg *et al*., 2004) and are not high-throughput. Edman degradation is a technique used to determine the native N-terminal sequence of mature proteins, which allows the identification of post-translational cleavage, however this technique is slow (at the rate of one amino acid per hour) and require proteins to be purified to homogeneity, preventing any high-throughput analysis of the proteome (Brune *et al*., 2006; Doucet and Overall, 2011).

Novel solutions to the identification of proteolytic processing, in a high-throughput manner have only been developed within the last decade, with the expansion of the field known as "degradomics" – the study of all elements involved in proteolysis, including proteases, inhibitors and the processed substrates (Doucet *et al*., 2008; Doucet and Overall, 2008; Impens *et al*., 2010a; Impens *et al*., 2010b; Lopez-Otin and Overall, 2002). Numerous experimental strategies have been developed to characterise the active site specificity of proteolytic enzymes, such as: substrate phage and bacterial display, peptide microarrays, mixture-based oriented peptide libraries, positional scanning peptide libraries, and proteome-derived, database-searchable peptide libraries; for example, proteomic identification of cleavage sites or "PICS" (Schilling and Overall, 2007, 2008). Many of these techniques are reviewed in detail by auf dem Keller and Schilling (2010). The identification of active site specificity alone is rarely sufficient to predict *in vivo* cleavage events, as cleavage of native, folded proteins is not only guided by active site but also by structural constraints of both enzyme and substrate (Overall, 2002; Schilling and Overall, 2008).

Terminal Amine Isotopic Labelling of Substrates (TAILS) utilises differential labelling of control and protease-treated samples in order to identify substrates of a specific protease (Kleifeld *et al*., 2011). As novel N-termini are identified in this method, it can also be used to infer proteolytic cleavage motifs of the protease. This method can also be adapted to screen for true N-termini at a global-proteome level, which makes it a high-throughput, non-targeted approach to identifying post-translationally cleaved

products. N-terminomics techniques have been applied to model mammalian systems yielding promising results (Guryca *et al*., 2012; Kleifeld *et al.*, 2011). These techniques have been applied to *M. hyopneumoniae* (Berry, Djordjevic, Unpublished data) with limited success, mainly due to variations in the lysine and arginine content of different organisms. TAILS and TAILS-like labelling techniques rely on dimethyl-labelling through primary amines on accessible lysine residues which effectively block those residues for cleavage using trypsin. When digestion for capture of neo-N-termini and mass spectrometric identification is required, only arginine residues are accessible for cleavage by trypsin and this creates very large peptides which are often incompatible with ionisation and mass spectrometry (Kleifeld *et al.*, 2011). In addition, knowledge of protein N-termini does not fully characterise cleavage events or provide comprehensive information about cleavage efficiency, particularly in the case of multiple cleavage events within a single pre-protein, generating overlapping fragments (See Chapter 5 for more detail). Thus, even with these promising new additions to the proteomics "toolbox", the most comprehensive information of post-translational proteolytic processing events *in vivo* should be obtained through complementary analyses, including analysis of intact proteins (i.e. a protein-centric approach).

1.3. Proteomic analysis of *M. hyopneumoniae*

Despite the apparent simplicity of the *Mycoplasma hyopneumoniae* genome, the proteome of these organisms is paradoxically complex, as a large number of important proteins, including the adhesin paralogs, are post-translationally modified and potentially multi-functional. Patterns of protein expression and cleavage are likely to influence pathogenesis (Djordjevic *et al.*, 2004). Microarray and transcriptome studies suggest that under normal culture conditions, the majority of genes (627 of the 691 genes of strain 232) are transcribed; while under various stressors all genes are transcribed (Madsen *et al*., 2006a; Madsen *et al*., 2006b; Madsen *et al*., 2007; Madsen *et al*., 2008; Oneal *et al*., 2008; Schafer *et al*., 2007). The molecular analyses so far carried out on *M. hyopneumoniae* can be supplemented by proteomic approaches that combine high-throughput analysis with further detailed analysis of surface-exposed proteins and post-translational cleavage patterns. This is necessary because protein cleavage and post-translational modification cannot be fully predicted by interpretation of the genome or transcriptome.

1.3.1. Global analyses and Mycoplasmas as model organisms

The minimal genomes of Mycoplasmas make them well-suited to large-scale, global analysis of the genome, transcriptome, proteome and metabolome, collectively referred to as systems biology. Mycoplasmas have long been considered to represent what is close to the "minimal requirements for life" and as such present as interesting model organisms for such studies (Balish, 2014; Karr *et al*., 2013; Kuhner *et al*., 2009; Wodke *et al*., 2013). In addition, this has led to recent interest in these organisms for use in synthetic biology (Gibson *et al.*, 2010).

Traditionally the exploration of gene function in a bacterium has been investigated through the generation of random insertional mutants by transposon mutagenesis (Maglennon 2013). Typically, this involves a transposon and transposase enzyme being delivered into the organism by transformation, resulting in random insertions throughout the chromosome. Insertions disrupt gene expression, and therefore insertions in essential genes will result in lethality, however insertions in non-essential genes can provide clues to gene function. Non-essential genes may be involved in pathogenicity and virulence, playing roles in adherence, invasion or immune evasion, and random mutants can be screened *in vivo* to identify attenuated mutants (Maglennon 2013, Hutchinson 1999). The genetic intractability of *Mycoplasma hyopneumoniae* has hindered the development of transposon mutagenesis strategies until very recently, Maglennon *et al*. (2013) developed a technique using simple artificial self-replicating plasmids containing the oriC of *M. hyopneumoniae* and an antimicrobial resistance gene. This is the first time that transposon mutagenesis has been demonstrated in this minimal genome pathogen.

The difficulty in performing genetic manipulation of Mycoplasmas has also seen an increase in global proteomic studies, in order to investigate gene function, without requiring genetic manipulation. Several members of the mollicutes have been profiled at the global proteome level including *M. genitalium* (Parraga-Nino *et al*., 2012), *M. pneumoniae* (Jaffe *et al*., 2004a; Ueberle *et al*., 2002), *Acholeplasma laidlawii* (Lazarev *et al*., 2011), *M. mobile* (Fisunov *et al*., 2011; Jaffe *et al.*, 2004b), *M. synoviae* (Menegatti *et al*., 2010), *M. gallisepticum* (Demina *et al*., 2009), *M. penetrans* (Ferrer-Navarro *et al*., 2006), *M. suis* (Yuan *et al*., 2010) and *M. hyopneumoniae* strains 7448

(Pinto *et al*., 2007) and 232 (Pendarvis *et al*., 2014). The work presented in this thesis will further contribute the comprehensive global proteome of strain J. In previously published work, coverage of the predicted proteomes varied from 4% presented in the partial proteome of *M. synoviae* using only 2D gels (Menegatti *et al.*, 2010) to 85.3% in *M. genitalium*, using gel-based and gel-free methods (Parraga-Nino *et al.*, 2012), and 88% of ORFs were validated in *M. mobile* by proteogenomic mapping (Jaffe *et al.*, 2004b). Some studies have also investigated sub-proteomes such as the membrane proteome (Cacciotto *et al*., 2010; Parraga-Nino *et al.*, 2012) and immunoproteome (Jores *et al*., 2009; Parraga-Nino *et al.*, 2012; Pinto *et al.*, 2007), illustrating the amenability of these organisms to comprehensive analyses.

Kuhner *et al.* (2009) investigated the proteome organisation of *M. pneumoniae* using a tandem affinity purification-mass spectrometry approach that enabled detection of protein interactions and putative complexes. This analysis revealed a number of homo- and heteromultimeric complexes, some of which showed higher level organisation. A number of protein complexes appeared to form larger, multiprotein complex entities, which suggested sequential steps in biological processes were taken. Proteins were also identified as belonging to more than one protein complex, or interacting amongst several known complexes, implying that these proteins are multifunctional. Overall, the study revealed an unexpected complexity in the proteome for an apparently minimal organism that could not be directly inferred from its genome composition and organization or from transcriptional analysis, emphasizing the importance of an integrated systems-wide approach to the study of organisms and their interactions (Balish, 2014; Flintoft, 2009; Kuhner *et al.*, 2009).

1.3.2. Surface proteome investigation

It is apparent that cell surface components of *Mycoplasma hyopneumoniae* are critically involved in infection and pathogenesis and as such it is desirable to be able to characterise the surface proteome of the bacteria to determine which components may be interacting with the host during infection. Despite the apparent suitability of mycoplasmas as model organisms for membrane proteomic studies (as they are bounded by only a single plasma membrane, with no organellar membranes to contaminate the sample) (Razin and Hayflick, 2010); with the exception of the detailed analysis of the

adhesins, very little cell surface proteome analysis had been performed for mycoplasmas until recently. This lack of detailed analysis is addressed and performed in Chapter 4.

Proteomic methods to examine the surface of cells, including but not limited to mycoplasmas, have been hindered by issues of insolubility, largely attributed to the hydrophobic nature of proteins with transmembrane domains (Cordwell, 2006; Cordwell and Thingholm). Several techniques have emerged in an attempt to circumvent this issue, including membrane enrichment with detergents such as Triton X-114 and gel-free techniques such as enzymatic cell surface shaving and cell surface labelling methods using reagents that are unable to permeate the cell membrane.

1.3.2.1. Membrane enrichment

Octylphenoxy polyethoxyethanol, or Triton X-114 (TX-114) is a non-ionic detergent that enables the separation and enrichment of hydrophilic and hydrophobic proteins, including integral membrane proteins with an amphiphilic nature (Bordier, 1981). When in solution above its critical micelle concentration and warmed from 0 to 20°C, TX-114 increases its micelle weight and in the process decreases its critical micelle concentration. This induces intermicellar interactions, which leads to turbidity known as the "cloud point" and phase separation of the detergent at 20°C. Hydrophilic proteins are recovered in the aqueous phase and hydrophobic proteins including the membrane component can be recovered in the detergent phase (Bordier, 1981; Rubin and Tzagoloff, 1973). This presents as a unique and highly effective pre-fractionation step prior to protein analysis of a complex sample, and enriches for membrane proteins in the detergent phase.

The use of TX-114 phase partitioning has historically been useful in the investigation of membrane proteins of *Mycoplasma hyopneumoniae*, including in the identification of immunogenic membrane proteins and lipoproteins. A study by Kim *et al*. identified immunogenic proteins of 70, 65, 50 and 44 kDa in the detergent phase following TX-114 extraction (Kim *et al.*, 1990; Wise and Kim, 1987a, b). Immunogenic *M. hyopneumoniae* proteins of apparent molecular masses 41, 66 and 82 kDa have also been identified following TX-114 enrichment, however they were found to partition to

the aqueous phase, and the 41 kDa protein was shown through antibody binding to whole cells to have a surface exposed epitope (Wise and Kim, 1987b). TX-114 phase separation may be useful for enrichment of membrane proteins, however it does not discriminate between outer, inner and integral membrane proteins and a subset of *M. hyopneumoniae* proteins, have been demonstrated to partition to both the detergent and aqueous phases (Djordjevic *et al.*, 2004). Whilst it may be suitable as a general enrichment technique for more hydrophobic proteins, more specific techniques are required to obtain an accurate representation of the surface topography.

1.3.2.2. Shaving

Proteolytic cell surface 'shaving' has been described extensively in the literature as an ideal method for identifying surface-exposed peptide epitopes and investigating the cell surface proteome of bacteria and eukaryotic cells (Bledi *et al*., 2003; Cordwell, 2006; Cordwell and Thingholm, 2010; Solis *et al*., 2010; Rodriguez-Ortega *et al*., 2006). Commonly, enzymes such as trypsin are used to liberate proteins from the cell surface without lysing the cells and these liberated proteins are further digested into peptides and analysed by mass spectrometry (Figure 1.9).

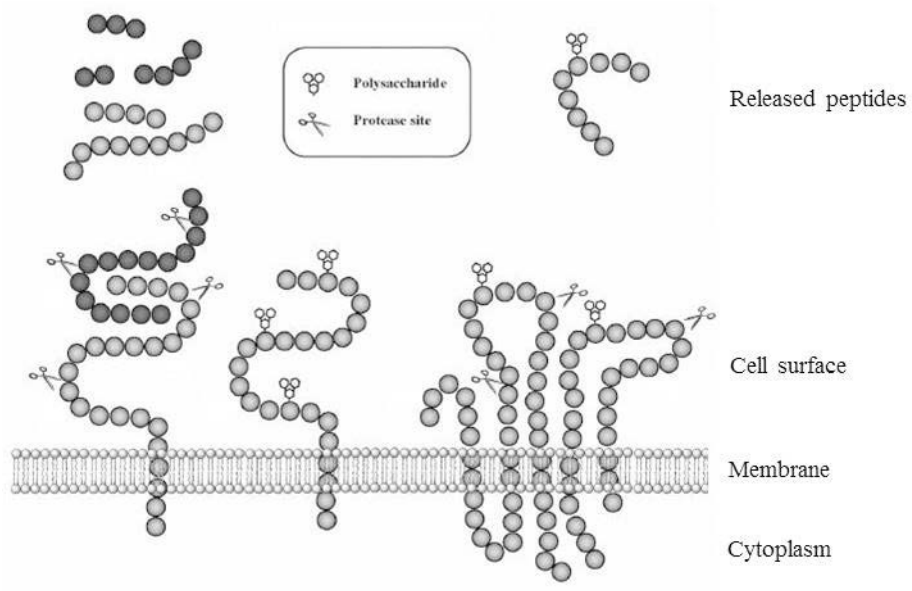


Figure 1.9. Diagrammatic representation of enzymatic cell surface shaving, releasing extracellular exposed domains.
Adapted from Bledi *et al.* (2003).

Literature covering bacterial and mammalian cell surface shaving experiments reveals that a major problem associated with this technique is the large number of proteins with known cytosolic functions that are (falsely or otherwise) identified as cell surface proteins. This is often attributed to contamination caused by cell lysis during enzymatic treatment. A major issue remains regarding those proteins that are truly surface-exposed and those that are theoretically cytoplasmic, but routinely identified in surface extracts (Solis *et al.*, 2010; Tjalsma *et al*., 2008). Some models of surface associated cytoplasmic proteins are based on non-specific secretion and then re-association with the cell surface. Such proteins are referred to as 'anchorless' membrane proteins and are said to be "non-classically secreted". These proteins are possibly exported by currently unknown secretion mechanisms, diffuse through porins, or re-associate to the cell membrane following autolysis particularly in bacterial communities, such as biofilms, which some *Mycoplasma* species have been shown to form (McAuliffe *et al*., 2006). Proteins that are weakly associated with the surface by charge interactions (or other mechanisms) are referred to as 'shed' proteins by Tjalsma *et al*. (Tjalsma *et al.*, 2008). This highlights the need for verification of the biological functions of different classes of cell surface-associated molecules by multiple methods and methods that do not require removal of proteins from the surface prior to analysis.

1.3.2.3. Cell surface labelling

A supporting method in cell surface topography investigation is analysis using cell surface labelling methods such as isotope labelling or biotinylation (Cole *et al*., 1987; Elia, 2008; Han *et al*., 2001). The strong biotin-avidin affinity ($K_d \sim 10^{-5}$ M) can be greatly exploited using such methods, and as such there are many commercially available forms of biotin, including water soluble and membrane impermeable variations such as N' Hydroxysuccinimido-biotin. The use of biotin in conjunction with avidin or streptavidin yields the highest binding affinity and allows the use of strong detergents during the purification of relatively insoluble proteins (Elia, 2008). Biotin labelling can be carried out under a variety of conditions to suit the purpose of labelling – minimising chances of cell lysis and active export or secretion of proteins and internalisation of biotin, while maximising surface protein labelling (Cole *et al.*, 1987; Elia, 2008). Once cell surface proteins are labelled, strong surfactants such as SDS can be used to disrupt the cells, maximising the solubility, and therefore recovery, of

hydrophobic membrane proteins. Biotin labelled surface proteins can be identified through gel electrophoresis followed by Western blotting and probed with streptavidin-conjugated detection reagents, or alternatively through affinity purification using for example monomeric avidin columns, which, if performed under mild enough conditions enables co-purification of binding partners and thus surface protein interactions may be examined under near-physiological conditions (Elia, 2008).

Recently, the surface proteome of *M. hyopneumoniae* strain 7448 was catalogued using biotin labelling of intact cells, identifying 59 protein species to be present at the cell surface (Reolon *et al*., 2014) (Discussed in Chapter 4).

1.4. Functional cell surface analyses

Although adhesin molecules have been well studied in mycoplasma, few studies have described other proteins that may be involved in pathogenesis of *M. hyopneumoniae.* Taking an unbiased approach to the study of the mycoplasma cell surface, where proteins are not excluded on the basis of predicted or annotated subcellular localisation; in conjunction with global studies of interactions with host molecules will provide valuable insight into host-pathogen interactions. This is further discussed in Chapter 4.

1.4.1. Interactions of M. hyopneumoniae *with extracellular matrix components*

The initial interaction between cell surface adhesins and host receptors is a crucial step in the infection process for any pathogen, however it is especially crucial in host-specific pathogens such as Mycoplasmas that have limited capacity to survive outside the host (Candela *et al*., 2010). This early colonization is also essential for pathogens to establish molecular processes to subvert the host cell machinery and allow bacterial replication, propagation and immune evasion, ultimately leading to chronic infection (Jacques *et al*., 1992; Nobbs *et al*., 2009). A key factor in understanding pathogenesis and enabling the intelligent design of vaccines or disease treatments is the understanding of host-pathogen interactions. In *M. hyopneumoniae*, the major interactions occur between cell-surface proteins of the bacteria and host extracellular matrix components. A number of mycoplasma proteins including the adhesins have

already been found to be capable of binding host extracellular matrix components such as heparan sulphate and fibronectin and other host molecules such as plasminogen.

1.4.1.1. Glycosaminoglycans

Host extracellular matrix is composed of a variety of proteoglycans and glycosaminoglycans (GAGs) which form a network in and around tissues. While the composition of the matrix varies from tissue to tissue, common components include fibronectin, collagens type I to XV, laminin and proteoglycans such as heparan sulphate and chondroitin sulphate (Hynes, 2009; Kim *et al*., 2011). Glycosaminoglycans are found on the surface of a majority of eukaryote cells and provide an attractive target for micro-organisms with the ability to bind them (Erlinger, 1995). Erlinger (1995) found that cilia and microvilli of the porcine respiratory tract display glycosaminoglycans which were sensitive to heparinase, an enzyme capable of degrading heparan sulphate. Jenkins *et al.* (2006) demonstrated that P97 is able to bind heparin, a highly sulphated glycosaminoglycan related to heparan sulphate, and that both R1 and R2 regions are required for this. P159 and P216 also have the ability to bind heparin and when bound, prevent binding to cilia, suggesting that *M. hyopneumoniae* binds to the GAG component of cilia (Jenkins *et al.*, 2006). Zhang *et al*. (1994) found that dextran sulfate, heparin, chondroitin sulfate, laminin, mucin, and fucoidan significantly inhibited the binding of the mycoplasmas, while laminin also blocked the receptors in cilia (Table 1.3). The six inhibitors also disrupted the adherence of the mycoplasmas to intact ciliated cells, indicating these host molecules interact with the adhesive molecules on the surface of the mycoplasmas (Zhang *et al.*, 1994).

In addition to direct adherence to host cell GAGs, the recruitment of sulphated glycosaminoglycans to the bacterial cell surface has been recognised as a mechanism for pathogenesis, whereby they may serve as universal binding sites for a diverse array of host heparin-binding proteins, including adhesive glycoproteins (such as fibronectin), and inflammatory or immunomodulatory intermediates (Henderson *et al*., 2011; Rabenstein, 2002).

Table 1.3. Inhibitory mechanisms of different competitors.

Inhibitor[a]	*% Inhibition*[b] *produced by pre-incubation with:*	
	Cilia	Mycoplasma
Laminin (200)	74.8 ± 0.6	6.2 ± 5.8
Mucin (200)	-6.6 ± 6.6	78.3 ± 3.1
Dextran Sulfate[c] (100)	0.5 ± 6.7	16.4 ± 20.2
Chondroitin sulphate B (200)	-1.3 ± 0.4	45.6 ± 15.5
Fucoidan (100)	27.0 ± 3.5	77.6 ± 6.5
Heparin (10 U/ml)	-2.2 ± 6.2	93.4 ± 11.5

[a] Concentrations are in milligrams per millilitre unless indicated otherwise.

[b] Reported as mean ± standard deviation in triplicate assays.

[c] Molecular weight, 500 000.

Reproduced from Zhang *et al.* (1994).

In 1989, the first heparin-binding motifs were identified; XBBBXXBX or XBBXBX where B represents basic residues and X represents any other residue, preferentially hydrophobic (Cardin and Weintraub, 1989). The motifs have recently been revised, following structural analysis of heparin binding proteins, to consist of a string of between five and eight amino acids containing a pattern of basic amino acids (preferentially lysine or arginine and also histidine) periodically distributed amongst the remaining 17 natural amino acids (Dempewolf *et al*., 2013). The identification of such motifs allows for the prediction of potential heparin-binding proteins or regions from protein sequences.

1.4.1.2. Fibronectin

M. hyopneumoniae has been shown to bind fibronectin in a dose-dependent manner and several of the adhesin paralogs have been shown to bind fibronectin, including Mhp271 (Deutscher *et al.*, 2010), Mhp107 (Seymour *et al.*, 2011) and Mhp684 (Bogema *et al.*, 2012), P102 and Mhp108 (Seymour *et al.*, 2010; Seymour *et al.*, 2012). While the biological significance of the interaction between *M. hyopneumoniae* and fibronectin has not yet been confirmed, other pathogenic bacteria utilize fibronectin as a bridging molecule to integrin β1 receptors, causing downstream signalling events leading to actin

and microtubule rearrangement that facilitates internalization and hence microbial invasion (Henderson and Martin, 2011).

Fibronectin is a host extracellular matrix glycoprotein and a eukaryotic adhesin, with the ability to bind an array of molecules, including collagen, fibrin, actin (through β1 integrins), DNA and heparin (Henderson and Martin, 2011). It is found in the interstitial connective tissues and ECM as well as in a soluble form in body fluid such as plasma. Fibronectin serves diverse biological functions including cellular adhesion, migration, tissue development and differentiation, blood clot stabilisation and wound healing (Carsons, 1989). Fibronectin is also affected by the fibrinolytic system; plasmin may degrade fibronectin (Wachtfogel *et al*., 1988) and plasminogen activators have been shown to degrade fibronectin in the absence of plasminogen (Quigley *et al*., 1987). Urokinase-type plasminogen activator and tissue-plasminogen activator (tPA) have also been shown to promote the expression of fibronectin, and tPA can promote the assembly of fibronectin in the extracellular matrix (De Petro *et al*., 2002).

It has been shown that the distribution of fibronectin on the surface of model porcine kidney epithelial-like cells (PK15 cells) correlates with the distribution of *M. hyopneumoniae* colonisation of the cell monolayer, indicating that fibronectin is recruited to the site where *M. hyopneumoniae* cells are attached (Figure 1.10) (Raymond *et al.*, 2014). It has been demonstrated that this is also the case *in vivo* using immunohistochemistry to show that *M. hyopneumoniae* cells co-localise with fibronectin in lung tissue samples of infected and healthy pigs (Raymond *et al.*, 2014).

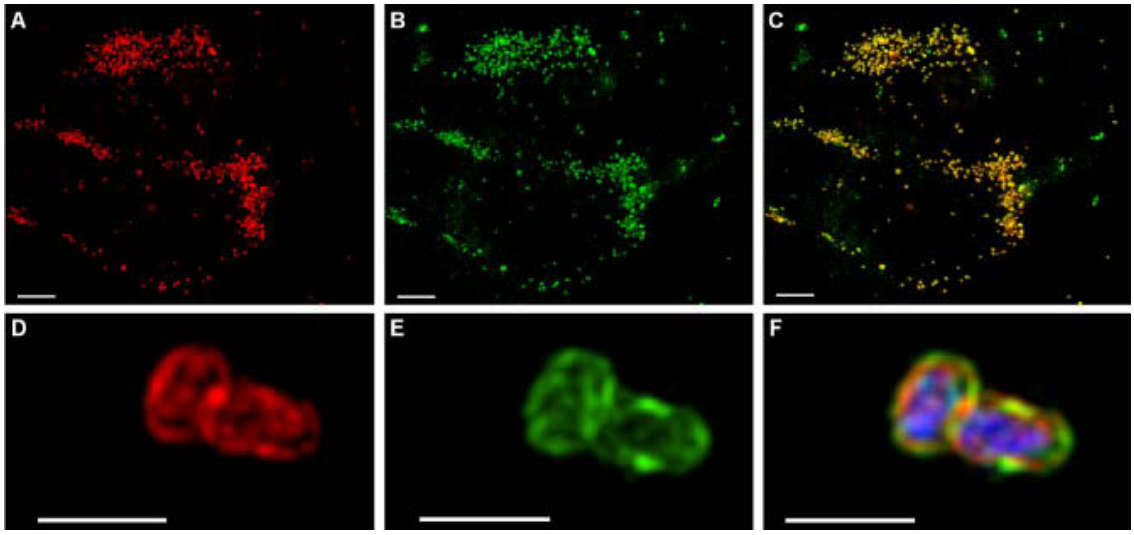


Figure 1.10. Fluorescence micrographs showing the distribution of fibronectin in *M. hyopneumoniae*-infected PK15 monolayers.

A–C. Confocal laser scanning microscopy images of *M. hyopneumoniae* cells colonizing PK15 monolayers. Panel A shows the location of *M. hyopneumoniae* cells on PK15 cell monolayers in red. Panel B shows the distribution of fibronectin on PK15 cell monolayers in green. Panel C shows a merge of *M. hyopneumoniae* cells and fibronectin showing their co-localization. Scale bar is 10 μm.

D–E. 3D-structured illumination microscopy images of *M. hyopneumoniae* cells bound to a PK15 cell monolayer. Panel D shows two *M. hyopneumoniae* cells adhering to the surface of a PK15 cell showing the distribution of P123J cleavage fragments containing R1 and R2 in red. Panel E shows the distribution of fibronectin on *M. hyopneumoniae* cells in green. Panel F shows an overlay of D and E. The merged image clearly depicts fibronectin co-localizing on the surface of *M. hyopneumoniae* cells. DNA was stained with DAPI and appears blue. Scale bar is 1 μm. Adapted from Raymond *et al.* (2014).

In *M. hyopneumoniae* infected pigs, bronchoalveolar lavage fluid (BALF) plasmin levels have been shown to be up-regulated and given that fibronectin is present on the ciliated epithelial border at the site of infection (Seymour *et al.*, 2012; Woolley *et al*., 2013), plasmin(ogen) and tPA in the lung may play a dual role in the degradation of extracellular matrix proteins and in initiating increased expression of fibronectin. This complex interplay between the host systems may contribute to the success of *M. hyopneumoniae* adherence and colonisation.

Fibronectin carries out a range of functions. In addition to cellular fibronectin which is recognised as an ECM component, there is also plasma fibronectin, which is found in blood, saliva and other body fluids, playing roles in wound healing and clot formation. Plasma fibronectin has been shown to be involved in inflammation, promoting chemotaxis, modulating leukocyte function and potentially activating macrophages (Henderson *et al.*, 2011). Fragments of fibronectin are found in the blood after injury (La Celle *et al*., 1991) and have been found in the fluid of chronic wounds. These fibronectin fragments have cell signalling actions, which appear to be important in infection and during inflammatory disease. Bacterial fibronectin binding proteins may target these fibronectin fragments and interfere with fibronectin fragment signalling. In addition, bacterial fibronectin binding proteins will often have actions other than acting solely as adhesins. This is reasonable considering that fibronectin has so many biological actions that interfering with it must have ramifications (Henderson *et al.*, 2011).

This draws similarities with the *M. hyopneumoniae* adhesins, which function primarily as adhesins, and many have been identified to bind fibronectin, however many have also been found to be multifunctional (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Wilton *et al.*, 2009).

1.4.1.3. Plasminogen

Plasminogen, whilst not an extracellular matrix component, is ubiquitous in the respiratory tract where it is available to be bound by cell surface proteins of *M.*

hyopneumoniae (Woolley *et al.*, 2013). Plasminogen is a circulating pro-enzyme, the precursor molecule to the broad-spectrum serine protease, plasmin. The mammalian plasminogen-plasmin system is central to fibrinolysis and extracellular matrix degradation (Lahteenmaki *et al*., 2005). It has also been shown to play an active role in tissue remodelling, acting on glycoproteins and proteoglycans of the ECM (such as fibronectin), and degradation of collagens through activation of latent matrix metalloproteases (Kucharewicz *et al*., 2003; Lahteenmaki *et al.*, 2005). This activation of matrix metalloproteases leads to cleavage of extracellular matrix proteins (Lahteenmaki *et al*., 2001). This, in turn, generates cleavage fragments of proteins and thus provides a pool of new N-termini that are substrates for microbial cell surface aminopeptidases, which in turn provide free amino acids that are required for growth. We have recently shown that a glutamyl aminopeptidase (Robinson *et al.*, 2013) and a leucine aminopeptidase (Jarocki *et al.*, 2015) are active as aminopeptidases on the surface of *M. hyopneumoniae*. This is particularly relevant for genome reduced organisms such as *Mycoplasma spp*., which lack many biosynthesis pathways (Maier *et al*., 2013).

The host plasminogen-plasmin system may also be hijacked by bacterial pathogens in order to promote binding of host structures, carry out tissue remodelling and facilitate invasion (Bergmann *et al*., 2013). In order to do this, the bacteria must sequester plasminogen to the cell surface in such a way as to allow it to be cleaved by host activation molecules to yield active plasmin, a serine protease with broad substrate specificity. Some bacterial species express endogenous plasminogen activators, facilitating plasmin generation independent of the host (Rabijns *et al*., 1997; Sodeinde *et al*., 1988; Wang *et al*., 1995). This process has been recognized as a common virulence mechanism displayed by a wide variety of pathogenic bacteria, parasites, fungi and several symbiotic organisms, with some examples provided in Figure 1.11 (Bergmann and Hammerschmidt, 2007; Lahteenmaki *et al.*, 2005; Lahteenmaki *et al.*, 2001).

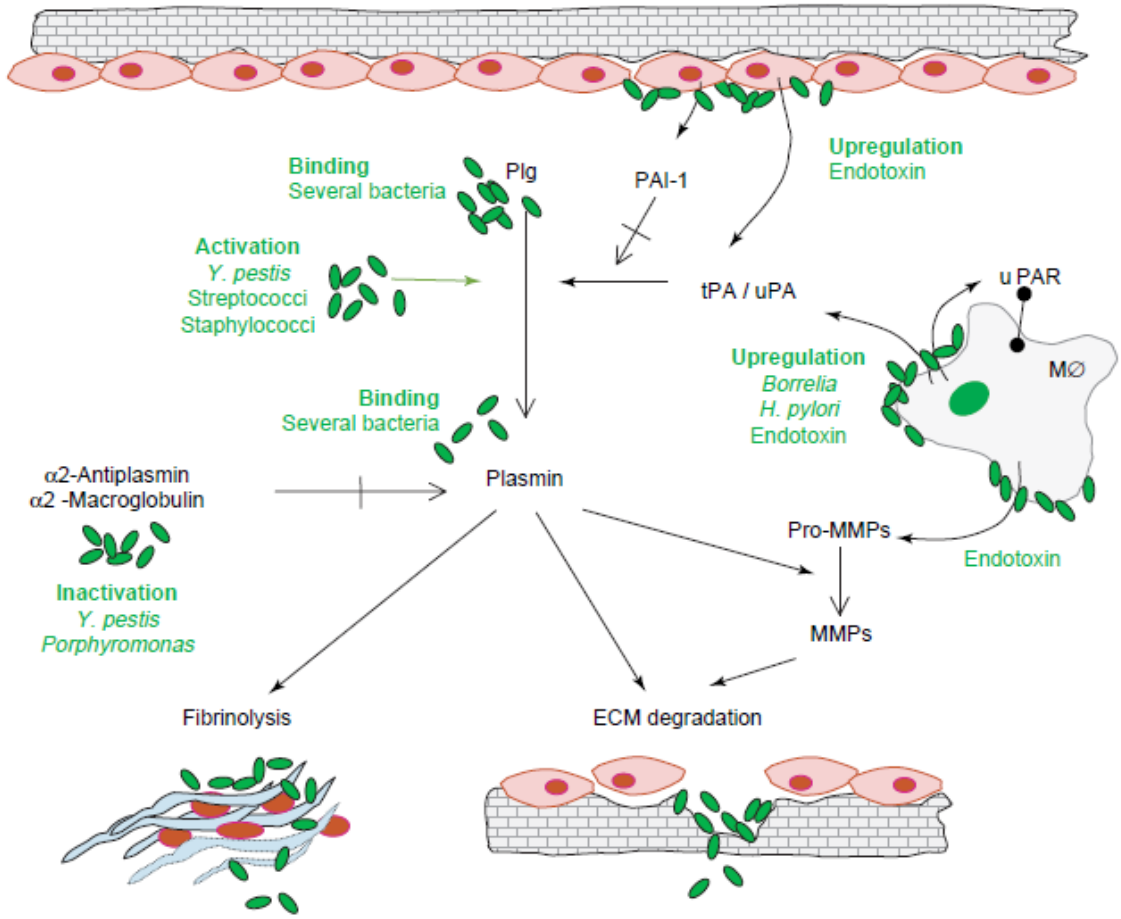


Figure 1.11. Overview of the mammalian plasminogen (Plg) system, its control and how pathogenic bacteria engage the system.

Bacteria intervene with the system (shown in green) by immobilizing Plg, plasmin and tPA on their surface, which enhances activation of Plg by tPA and protects plasmin from the main anti-protease, α2-antiplasmin. Some pathogens activate Plg by themselves, and a few species inactivate the antiproteases. Bacteria and endotoxin also influence secretion of tPA and PAI-1 from endothelium as well as uPA, uPAR and Pro-MMPs from macrophages (MØ). Generation of bacterium-bound plasmin and/or uncontrolled soluble plasmin can be targeted to fibrin clots and by bacterial adherence to BM and ECM, the degradation of which enables bacterial migration through tissue barriers. Reproduced from Lahteenmaki *et al.* (2005).

In vitro studies in *M. hyopneumoniae* have previously shown that the binding of plasminogen to the bacterial cell surface promotes activation to plasmin in the presence of host tissue-specific activator (tPA) (Seymour *et al.*, 2011; Seymour *et al.*, 2012). *M. hyopneumoniae* promotes the conversion of plasminogen to plasmin in the BALF bathing the respiratory tract of pigs. Elevated levels of plasmin can be detected in the BALF of pigs infected with *M. hyopneumoniae* (Seymour *et al.*, 2012; Woolley *et al.*, 2013) and a significant correlation has been described between bacterial load and plasmin activity in the BALF of experimentally challenged pigs (Woolley *et al.*, 2013). As mentioned previously, the ability of a pathogen to bind and activate plasmin(ogen) is linked with the ability to invade tissue site distal to the initial colonization site (Lahteenmaki *et al.*, 2005). *M. hyopneumoniae* has been cultured from the liver, spleen and kidneys of infected swine suggesting that it is capable of invading epithelial barriers and traversing to distal tissue sites (Le Carrou *et al*., 2006; Marois *et al*., 2007; Woolley *et al.*, 2013). The exact mechanism which allows *M. hyopneumoniae* to disseminate to these distal tissue sites is unknown; however, it is hypothesised to involve binding to glycosaminoglycans, plasminogen, fibronectin and/or actin (Raymond *et al.*, 2014).

1.4.1.4. Actin

Actin is a protein central to the eukaryotic cytoskeleton, exploitation of which is a common feature in host-pathogen interactions (Truong *et al*., 2014). Actin may be manipulated by invading pathogenic bacteria, as actin remodelling promotes bacterial invasion by non-phagocytic cells, survival within host cells, cell-to-cell spread and colonization at the interface of host epithelium (Hicks and Galan, 2013). Intracellular bacterial pathogens such as *Shigella*, *Salmonella*, *Listeria*, *Burkholderia*, *Rickettsia* and *Mycobacterium* species have been demonstrated to subvert cellular actin dynamics in order to facilitate movement within host cells and also infect neighbouring cells. This promotes intracellular survival and assists in immune evasion (Stevens *et al*., 2006). Intracellular pathogens may hijack the actin assembly machinery and display intracellular motility through the formation of membrane protrusions which form vacuoles in adjacent cells, leading to bacterial dissemination (Talman *et al*., 2014). Extracellular pathogens such as enteropathogenic *E.coli* have also been demonstrated to modulate the host actin cytoskeleton, by inducing motile, actin-rich pedestals. These possibly serve to facilitate the spread to nearby cells (Stevens *et al.*, 2006). Actin

binding may be direct, or indirect through binding fibronectin as discussed above. The ability of *M. hyopneumoniae* to directly bind actin has not yet been extensively investigated; however it may be a contributing factor to potential internalisation mechanisms. Functional screening of surface proteins of *M. hyopneumoniae* presented in Chapter 4 includes preliminary identification of putative actin-binding proteins.

1.4.1.5. Host epithelial cell surface proteins

Pathogen binding interactions may be mediated by other, as of yet unidentified protein-protein interactions between cell surface proteins of bacteria and host epithelial cells. In order to identify such interactions, an infection cell culture model has been established in the Djordjevic laboratory, utilising porcine kidney epithelial-like cells (PK15 cells). *M. hyopneumoniae* strains J, and 232 have previously been shown to adhere to PK15 cells in a receptor-dependent manner and their adherence was blocked by pre-treating *M. hyopneumoniae* cell suspensions with trypsin (Zielinski *et al.*, 1990), suggesting these interactions are mediated by protein. This was further investigated by examining *M. hyopneumoniae* strain J interactions with PK15 cells by scanning electron microscopy (Burnett 2006). It was found that *M. hyopneumoniae* adheres intimately to PK15 cells, and can be seen closely associated with microvilli on the surface of the monolayers (Figure 1.12), however, this adherence could be significantly reduced by pre-incubation of *M. hyopneumoniae* cells with a recombinant C-terminal 52 kDa fragment of the adhesin-like protein P159 (Figure 1.12B), and also by pre-treating *M. hyopneumoniae* cells with a saturating concentration of heparin (Figure 1.12C). SEM and binding studies using PK15 cells have also been performed using latex beads coated in recombinant *Mycoplasma* protein fragments to determine adherence and invasion capacity (Figure 1.12 D-F). (Burnett *et al.*, 2006).

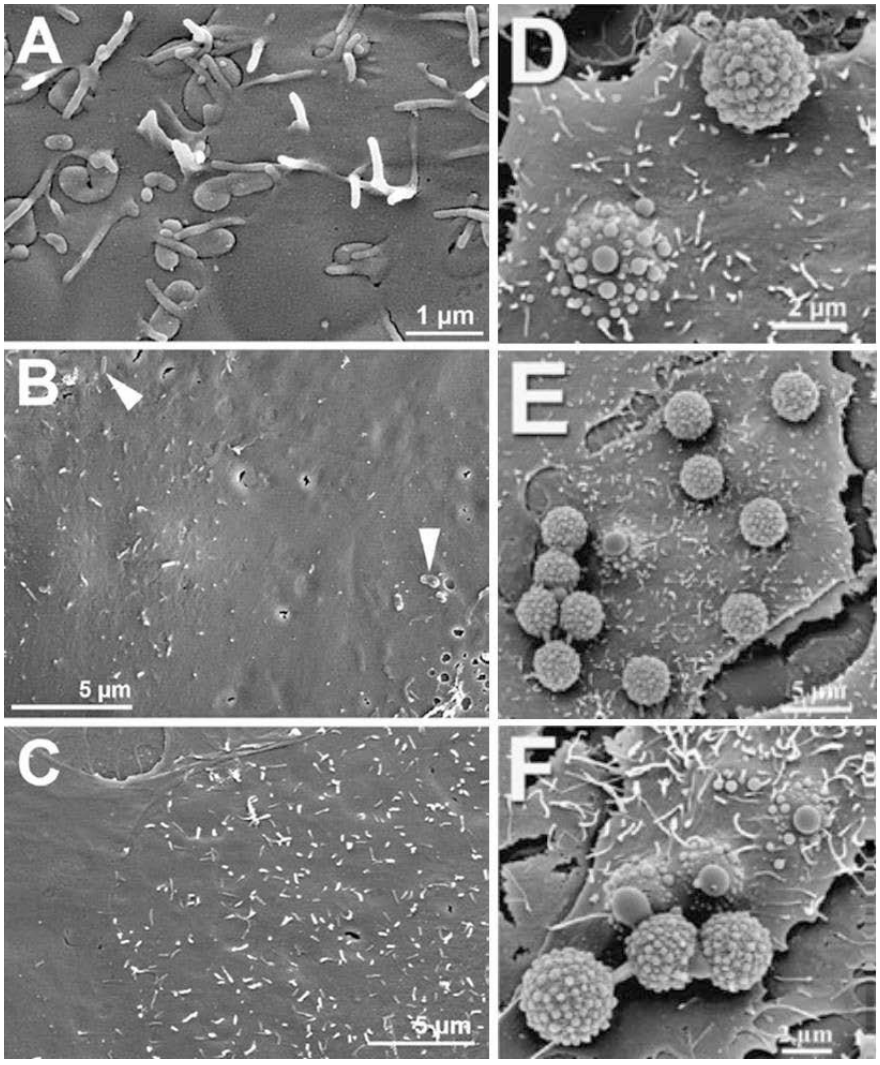


Figure 1.12. Scanning electron micrographs depicting the interaction of *M. hyopneumoniae* with PK15 cells.

A. *M. hyopneumoniae* interacting with PK15 cells. B. The inhibition of *M. hyopneumoniae* adherence due to the presence of recombinant C-terminal fragment of P159 ($F4_{P159}$) protein (1 µg). Adhering *M. hyopneumoniae* cells are indicated by an arrow. C. The inhibition of adherence due to the presence of heparin (500 µg ml^{-1}). D-F Scanning electron micrographs depicting the interaction of latex beads separately coated with recombinant P159 fragment ($F4_{P159}$) with PK15 cells. D. shows an $F4_{P159}$-coated bead adhering to; and another within, a PK15 cell 2 h after incubation. E and F. depict $F4_{P159}$-coated latex beads adhering to and residing inside PK15 cells 2 and 4 h after incubation respectively. Scale bars are given at the bottom of each image. Adapted from Burnett *et al.* (2006).

Collectively these interaction studies determined the C-terminal region of P159 (through binding studies with recombinant fragment $F4_{P159}$) to be a key factor in adhesion to host cells which may also play a role in facilitating invasion (Burnett *et al.*, 2006). Although these methods provide robust physical evidence for adherence or inhibition of binding, they are laborious and cannot be applied in a high-throughput manner.

To identify *M. hyopneumoniae* proteins that bind PK15 cell surface proteins in a global manner, we have developed a 2-step column affinity chromatography technique which utilises intact cell surface biotinylation of PK15 cells and subsequent avidin purification to create a PK15 surface protein column (modified from Nunomura *et al*. (2005)). This column is then incubated with *M. hyopneumoniae* cell lysate, washed and interacting proteins eluted (Raymond *et al.*, 2014; Raymond *et al.*, 2013; Tacchi *et al.*, 2014). This allows selective purification of low- and high-affinity binding proteins in a high-throughput, global manner that is useful as a screening tool prior to more thorough investigations.

1.5. Moonlighting proteins

The term "moonlighting" was coined to describe multifunctional proteins, capable of performing two or more distinct biological functions within a single polypeptide chain, excluding proteins that play the same role in different locations or splice variants (Jeffery, 1999). Since the term was defined, several hundred moonlighting proteins have been identified in a wide variety of organisms across all three domains of life (Henderson and Martin, 2014). Moonlighting proteins appear to be able to switch between functions due to a change in cellular localization, expression in a novel cell type, oligomeric state, and cellular concentration of a ligand, substrate, cofactor or product (Figure 1.13) (Jeffery, 1999).

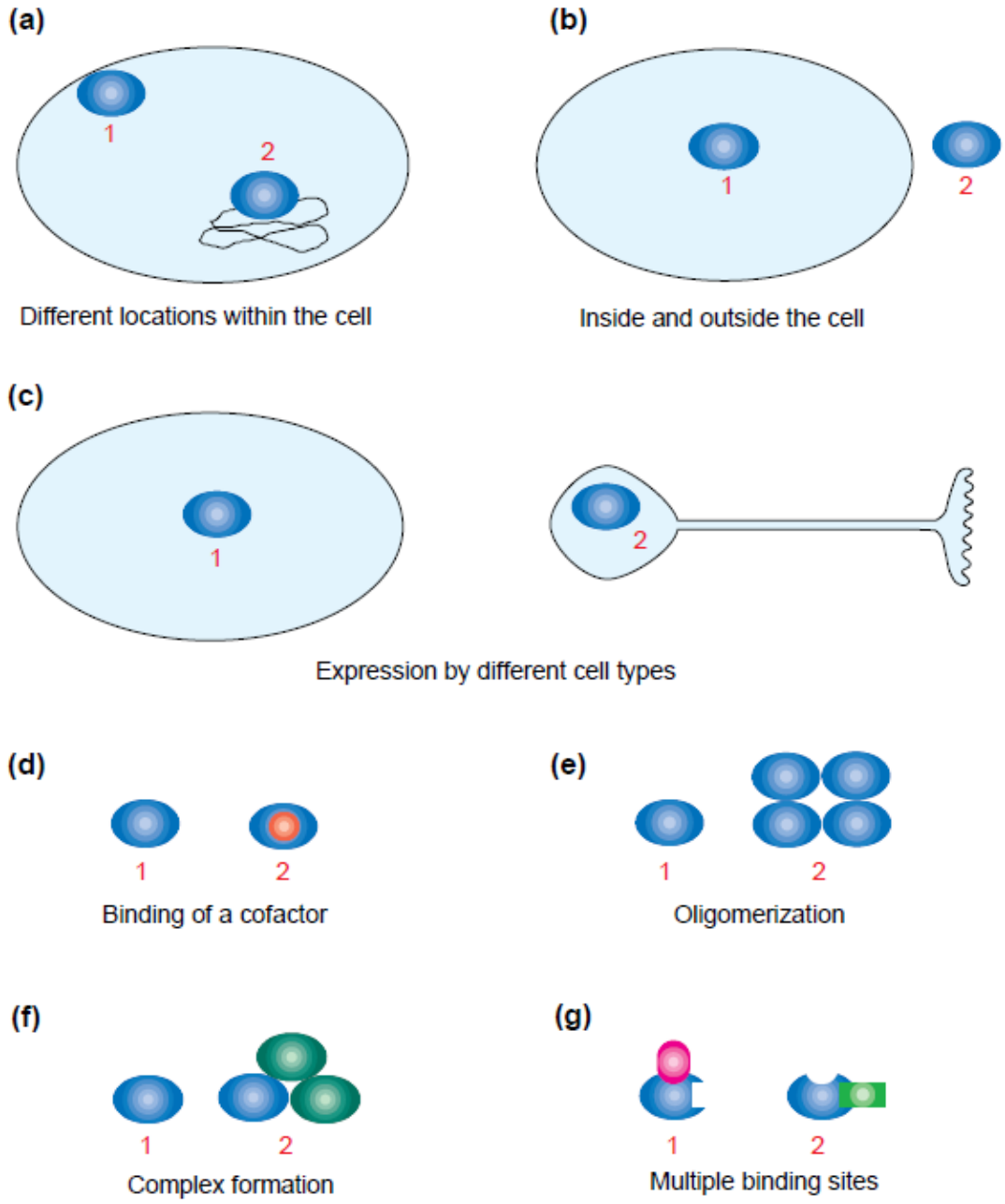


Figure 1.13. Examples of moonlighting functions.

Mechanisms for switching between functions (1 and 2). Reproduced from Jeffery (1999).

The discovery of moonlighting proteins in surface proteome analyses has been a contentious issue as they are not readily validated as truly surface-exposed without further investigation of individual proteins, which is not always feasible in large-scale analyses. Hence, the presence of potential moonlighting proteins at the cell surface has frequently been attributed to cell lysis in sample handling or normal cell turnover (de Miguel *et al*., 2010; Hempel *et al*., 2010; Tan *et al*., 2008). However there is groundswell of support for the concept of microbial moonlighting proteins (Henderson, 2014; Henderson and Martin, 2014; Wang *et al*., 2014).

Metabolic enzymes, previously assumed to be "housekeeping proteins", are now among the most commonly recognized examples of moonlighting proteins, particularly those involved in glycolysis (Henderson and Martin, 2011) (Figure 1.14). The structural and catalytic properties of glycolytic enzymes are well characterised in relation to their primary functions, and thus they can be easily recognised when carrying out alternative functions. Many glycolytic enzymes, despite lacking typical signal peptides and membrane anchoring mechanisms, are found on the surface of microbial pathogens and perform a variety of functions (Alderete *et al*., 2001; Chhatwal, 2002; Pancholi, 2001; Pancholi and Chhatwal, 2003; Sirover, 1999). These proteins have frequently been shown to act as virulence factors, contributing to host cell adhesion, invasion and immune modulation or evasion (Pancholi and Chhatwal, 2003).

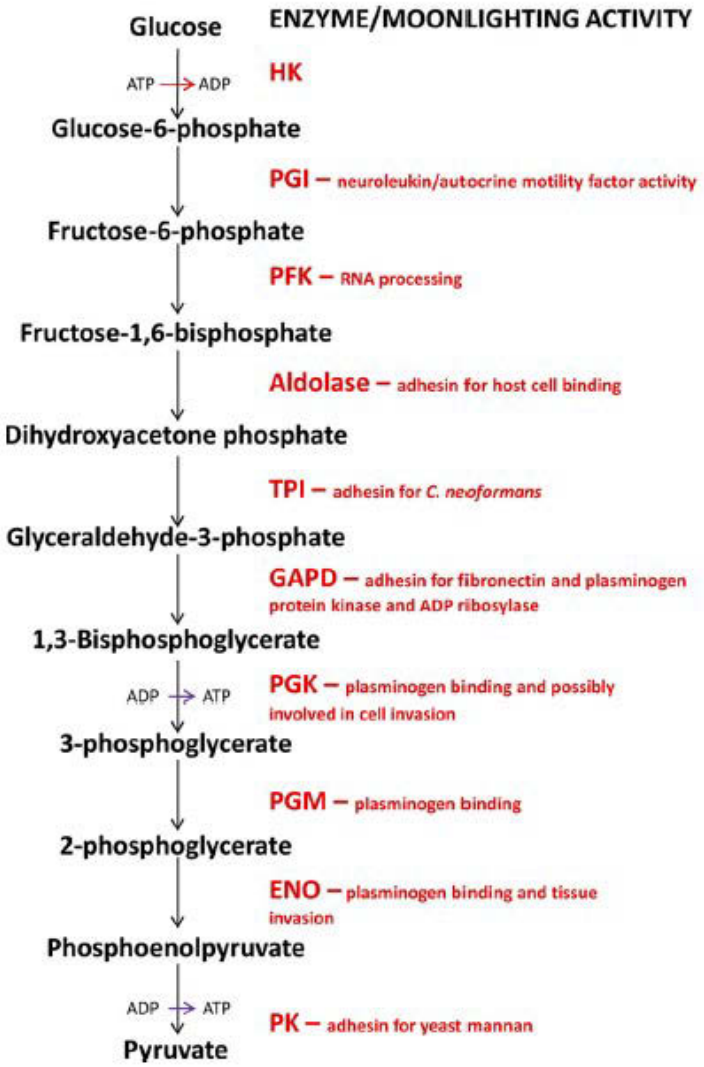


Figure 1.14. The glycolytic pathway and moonlighting actions of bacterial glycolytic proteins.

HK, hexokinase; PGI, phosphoglucose isomerase; PFK, phosphofructokinase; TPI, triose phosphate isomerase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, enolase; PK, pyruvate kinase. Reproduced from Henderson *et al.* (2011).

Identification of moonlighting proteins, much like identification of proteolytic cleavage, has previously relied on serendipitous empirical data (Jarocki *et al.*, 2014). Currently, there are no reliable methods to predict moonlighting proteins *in silico* as there are no apparent moonlighting motifs (Babady *et al.*, 2007) and no obvious links to sequence

homology between moonlighting and non-moonlighting proteins of the same class or between species (Zhao *et al*., 2009).

In *Mycoplasma pneumoniae*, elongation factor Tu (EfTu) and pyruvate dehydrogenase (PdhB) were identified as surface-exposed moonlighting proteins, through screening for fibronectin binding proteins by ligand blotting of whole cell lysates (Dallo *et al*., 2002). Immunogold labelling clearly showed localisation to the cell surface (Figure 1.15).

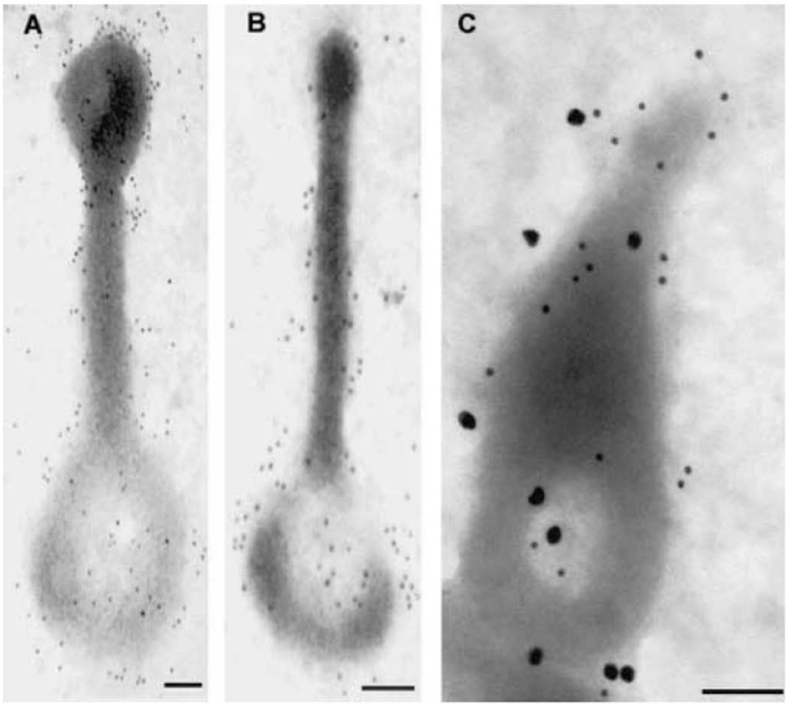


Figure 1.15. Immunogold electron microscopy detection of EfTu and PdhB proteins on *M. pneumoniae* cell surfaces.

Mycoplasmas were incubated with antisera generated against recombinant EfTu and/or recombinant PdhB and rabbit IgG conjugated gold particles (size 10 or 20 nm). Gold labelling of PdhB (panel A, 10 nm) showed both membrane and tip-associated localization. In contrast, gold labelling of EfTu (panel B, 20 nm) revealed random membrane distribution. Furthermore, gold particle double-labelling (panel C) confirmed the contrasting distribution of PdhB (10 nm) and EfTu (20 nm) on the *M. pneumoniae* membrane and tip surfaces (bar = 0.1 μm). From Dallo *et al.* (2002).

Further, in *M. hyopneumoniae*, two aminopeptidases, glutamyl aminopeptidase MHJ_0125 and leucine aminopeptidase MHJ_0461, were identified as surface-exposed moonlighting proteins, retaining proteolytic activity while also functioning as adhesins, binding host molecules such as plasminogen and heparin (Jarocki *et al.*, 2015; Robinson *et al.*, 2013). Moonlighting functions of proteases has been recently extensively reviewed (Jarocki *et al.*, 2014). While there are no accurate ways to predict moonlighting activity, global surface proteome analyses with an emphasis on functional interactions is likely to provide useful information on potential moonlighting proteins. The results of such an approach are presented in Chapter 4.

1.6. Conclusions and research aims

Mycoplasmas, despite possessing simple genomes have shown to have complex post-transcriptional and post-translational regulation, employing proteolytic processing and multifunctional or moonlighting proteins. The background presented highlights the need for a system-wide, non-hypothesis driven analysis of the proteome, with a specific focus on function in order to characterise the protein content of *M. hyopneumoniae* and potentially elucidate the mechanisms of pathogenesis.

The work presented in this thesis will be divided into three main chapters. Firstly the characterisation of the global proteome of *M. hyopneumoniae* is described, with a focus on protein-centric proteomic approaches to enable analysis of proteolytic processing. This is followed by a report of the surface-exposed proteome, examining moonlighting proteins with details of identified host-pathogen interactions. Finally, the in-depth characterisation of a prominent adhesin protein P216 is presented, focusing on post-translational processing and adherence to host substrates. A separate methods section will also be presented to provide complete descriptions of the methods. Changes or modifications to individual methods are described in the relevant chapters where appropriate.

This work has endeavoured to elucidate the expressed, functional proteome of *Mycoplasma hyopneumoniae*, and in doing so has revealed the complexity of this genome-reduced organism by uncovering the extent of post-translational modifications of proteins, that peptide-centric or targeted proteomics approaches would have failed to

identify. This has provided us with valuable insight into the way this organism interacts with its environment within the host, so we can better understand mechanisms of pathogenesis in order to intelligently design novel vaccines and therapeutics. In undertaking such a detailed proteomic study on a minimal organism, an aim is also to improve our understanding of fundamental cellular processes that may have wide-ranging implications in the field of cell biology in general and particularly in bacterial pathogenesis.

Chapter 2. Materials and methods

This section describes in full all methods carried out throughout the presented chapters. Specific details, including any modifications to techniques are included in the relevant chapters where appropriate.

Ethics statement: Antisera generation and collection was performed by trained staff at the Elizabeth Macarthur Agricultural Institute, following approval by the Animal Ethics Committee at the Elizabeth Macarthur Agricultural Institute (AEC project no. M12/11). Animal work was performed in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes.

2.1. Bacterial strains and growth conditions

2.1.1. Mycoplasma hyopneumoniae *culture*

Mycoplasma hyopneumoniae strain J was originally isolated in the United States, provided by T. Young (Iowa State University Veterinary Research Medical School, USA) (Bereiter *et al*., 1990). *M. hyopneumoniae* strain 232 was originally isolated in Great Britain, and subsequently obtained from A. Pointon (South Australian Research & development Institute, Australia) (Goodwin *et al*., 1965).

All *M. hyopneumoniae* cultures were grown in modified Friis medium (Appendix I), inoculated 1:50 with previously grown culture or cells from -80°C storage, to mid-log phase (pH 6.8-7.2 as indicated by colour change) at 37°C in rolling culture. Cells were harvested by centrifugation for 20 min at 10 000 × *g* and washed three times in phosphate buffered saline. Cell pellets were then used fresh, frozen at -20°C for later use or freeze-dried for longer term storage.

2.1.2. Escherichia coli *culture*

E. coli BL21™ Star (DE3) [Invitrogen, USA] was used for expression of hexahistidine (6x His) tagged recombinant proteins. Constructs were maintained in *E. coli* TOP10 [Invitrogen, USA].

All *E. coli* strains were grown in Luria Bertani (LB) broth (10% (w/v) peptone, 5% (w/v) yeast extract, 5% (w/v) NaCl) or on LB agar (10% (w/v) peptone, 5% (w/v) yeast

extract, 5% (w/v) NaCl, 12% (w/v) agar) supplemented with 50 mg.mL^{-1} Ampicillin (as required) at 37°C with aeration by shaking at 220 rpm.

2.1.3. PK15 cell culture

Porcine kidney epithelial like cells (PK15 cells) were used as a cell culture model for evaluation of *in vitro* binding capacity of *M. hyopneumoniae.* Adherent cells were cultured in flat-bottomed culture flasks at 37°C with 5% CO_2 in Dulbecco Modified Eagle Medium supplemented with 10% (v/v) foetal calf serum until 70-80% confluence was reached. Cells were subcultured or harvested by incubating with TrypLE Express [Life Technologies] for 10 min at room temperature. Cells were subcultivated at a ratio of 1:2.

2.2. Recombinant protein expression, purification and generation of antisera

2.2.1. Recombinant protein expression

The selected genes were synthesized and cloned into the expression vector PS100030 by Blue Heron Biotech (USA) removing any in-frame TGA codons by mutagenizing to TGG. In Mycoplasmas, the TGA codon encodes for tryptophan which results in truncated proteins when expressing Mycoplasma genes in *E. coli*. The recombinant constructs were transformed into BL21 Star™ (DE3) One Shot® chemically competent *E. coli* cells [Life Technologies] by heat shock. Overnight cultures (50 mL) in LB broth containing 100 µg.mL^{-1} ampicillin were grown and used to inoculate 1 L LB broth supplemented with 100 µg.mL^{-1} ampicillin. Once the OD_{600} of the cultures reached 0.5 indicating mid-log phase, cells were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Incubation was maintained for 3-4 h and the cells were harvested by centrifugation at 10 000 × *g* for 20 min.

2.2.2. Recombinant protein purification

6x His-tagged proteins were purified by nickel chromatography. *E. coli* cells were lysed by gentle stirring in Lysis buffer (100 mM NaH_2PO_4, 10 mM Tris-HCl, 8 M urea, pH 8.0). Unbroken cells and debris were pelleted by centrifugation at 10 000 × *g* for 20 min. Lysate was mixed with 2 mL Profinity™ immobilized metal affinity

chromatography (IMAC) nickel-charged resin [BioRad Laboratories] with gentle mixing overnight at 4°C. Slurry was then loaded into a column and washed with 8 mL wash buffer (100 mM NaH_2PO_4, 10 mM Tris-HCl, 8 M urea, pH 6.3). 6x His-tagged recombinant proteins were eluted with 12 mL (4 × 3 mL) elution buffer 1 (100 mM NaH_2PO_4, 10 mM Tris-HCl, 8 M urea, pH 5.9) and 12 mL (4 × 3 mL) elution buffer 2 (100 mM NaH_2PO_4, 10 mM Tris-HCl, 8 M urea, pH 4.5). Purity of elutions was assessed by 1D SDS-PAGE. Recombinant proteins were dialysed against 3 × 3 L PBS containing 1% (w/v) SDS for a minimum of 18 h with at least 3 h between buffer changes. Dialysed proteins were checked by 1D SDS-PAGE and verified by LC-MS/MS.

2.2.3. Generation of polyclonal antisera

Antisera against recombinant proteins were generated via primary and secondary intramuscular injections of antigen into New Zealand White rabbits at 2-week intervals. Antigens (approximately 0.5 mg) were prepared for injection by mixing equal volumes of purified protein (approximately 500 µl) and Freund's incomplete adjuvant [Sigma Aldrich, St. Louis, Missouri]. Pre-immune serum was collected prior to the primary injection and served as control antiserum. A trial bleeding was performed 10 to 14 days after the secondary injection of antigen, and immune responses to the antigen at this stage were assessed via immunoblotting. Positive sera were collected via cardiac bleeding, during which the rabbits were anesthetized and euthanized by exsanguination.

2.3. Sample preparation for proteomics

2.3.1. Preparation of Mycoplasma whole cell lysates for 2D gel electrophoresis

Washed mycoplasma cell pellets were resuspended in solubilisation buffer (7 M urea, 2 M thiourea, 1% w/v C7BzO) with 40 mM Tris-HCl, pH 8.8 and lysed by three rounds of sonication in 30 s bursts. Proteins were reduced and alkylated by the addition of 5 mM tributylphosphine and 20 mM acrylamide monomers and incubated for 90 min at room temperature. Samples were centrifuged at 16 000 × *g* for 10 min to pellet insoluble material and the soluble protein precipitated by 5 volumes of acetone at room temperature for 30 min. Protein was pelleted by centrifugation at 4 000 × *g* for 10 min and acetone decanted and the pellet air-dried. Pellets were resuspended in solubilisation

buffer (7 M urea, 2 M thiourea, 1% w/v C7BzO) and assayed to determine protein concentration.

2.3.2 Triton X-114 phase extraction of Mycoplasma proteins

Mycoplasma cell pellets were resuspended in 1% Triton X-114 buffer (1% TX-114, 10 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA) and extracted overnight (minimum 16 h) at 4°C on a rotating wheel. Insoluble material was pelleted and stored as "TX-114 insoluble phase". The soluble material was removed to a new tube and incubated at 37°C for 10 min. Phase separation was induced by centrifugation at 9000 × *g* for 5 min at room temperature. The TX-114 aqueous phase (top) was removed to a new tube, and TX-114 was added to a final concentration of 2%. To the TX-114 detergent phase (bottom) in remaining tube, 1% Triton X-114 buffer was added to a final concentration of 2% TX-114. Both samples were re-extracted for a further 4-6 h on a rotating wheel at 4°C. Phase partition was performed again and both aqueous (top) and detergent (bottom) fractions were pooled and precipitated with 5 volumes of ice-cold acetone for 30 min. Protein was pelleted by centrifugation at 4 000 × *g* for 10 min, the acetone decanted and the pellets air-dried. Aqueous, detergent and insoluble pellets were then subjected to extraction as described to obtain whole cell lysates above.

2.3.3. Preparation of native whole cell extracts

Washed mycoplasma cell pellets were resuspended in 0.1% Triton X-100 with the addition of Benzonase endonuclease [Sigma Aldrich] to degrade DNA. Samples were incubated for 10 min at room temperature before the addition of 1× cOmplete protease inhibitor solution [Roche]. Samples were centrifuged for 10 min at 16 000 × *g* to pellet any insoluble material.

2.3.4. Preparation of whole cell lysates for dimethyl labelling

All steps for dimethyl labelling preparation and purification were performed according to, or adapted from the protocol described by Kleifeld *et al.* (2011). Fresh washed *M. hyopneumoniae* cell pellets were resuspended in 4 M guanidine hydrochloride, 100 mM HEPES buffer at pH 8.0 and lysed by ultrasonication. Samples were diluted with 100 mM HEPES buffer and re-concentrated by passing through a 3000 MWCO filter. This

washing step was performed 3 times to remove any free amino acids or other small cellular contaminants which could interfere with the labelling. Sample was reduced and alkylated with 5 mM tributylphosphine and 20 mM of acrylamide monomers for 90 min at room temperature.

2.4. Cell surface analyses

2.4.1 Biotinylation

Cell surface biotinylation was carried out on intact cells using Sulfo-NHS-LC-biotin, combined with avidin column purification and/or blotting to purify or identify biotinylated surface proteins.

For surface biotinylation experiments, freshly harvested and washed *M. hyopneumoniae* cells were resuspended in PBS (pH 7.8) and biotinylated with 0.5 $mg.mL^{-1}$ EZ-Link Sulfo-NHS-LC-biotin [Thermo Scientific] for 30 s on ice. The reaction was then quenched with the addition of a final concentration of 50 mM Tris-HCl (pH 7.4) and incubated for 15 min. Cells were washed in three changes of PBS and pelleted by centrifugation as above.

2.4.2 Enzymatic cell surface shaving

Enzymatic cell surface shaving with trypsin was used to identify surface exposed proteins. Freshly harvested and washed *M. hyopneumoniae* cells were resuspended in PBS (pH 7.8) and pre-warmed with gentle mixing for 15 min at 37°C. A solution of 5 $mg.mL^{-1}$ cell culture grade trypsin [Sigma Aldrich] was pre-warmed along with the cells. A final concentration of 50 $\mu g.mL^{-1}$ trypsin was added to the cells and allowed to incubate with gentle mixing for 5 min. After 5 min cells were immediately placed on ice and pelleted by centrifugation at 4000 × *g* at 4°C. Supernatant containing liberated surface exposed proteins and peptides was removed and centrifuged to remove debris and any remaining intact cells at 10 000 × *g* at 4°C for 20 min. Supernatant was pH corrected with 100 mM ammonium hydrogen carbonate (NH_4HCO_3) to pH >8 and reduced and alkylated with 5 mM tributylphosphine (TBP), 20 mM acrylamide monomers for 90 min at room temperature. For analysis by LC-MS/MS, sample was diluted with five volumes 100 mM NH_4HCO_3 and 1 µg Trypsin Gold [Promega] added and digested overnight at 37°C with gentle mixing. Sample was cleaned up using solid

phase extraction as described in section 2.5.8 before analysis by LC-MS/MS as described in section 2.7.1.

2.5. Proteomic techniques

2.5.1. Isoelectric focusing – immobilised pH gradient strips (IPG IEF)

Samples prepared for isoelectric focusing were free from contaminating salts and conductive material. This was checked by monitoring conductivity of the sample to ensure it was < 200 μSi.cm^{-1}. Samples with higher conductivity were desalted prior to loading using a BioSpin column [BioRad] as per manufacturer's instructions.

Samples were loaded using either active or cup loading methods. For both methods, 11 cm immobilised pH gradient strips pH 3-10, 4-7 [BioRad] or 6-11 Immobiline Drystrips [GE Healthcare] were partially rehydrated in a hydrophobic rehydration tray [Proteome Systems] with 150-200 μL solubilisation buffer (7 M urea, 2 M thiourea, 1% w/v C7BzO) for ~1 h with the lid on to prevent evaporation.

For active loading, 100-150 μL of sample containing 150-300 μg of protein was loaded underneath the strip while in the focusing tray [Proteome Systems] under paraffin oil. The sample is drawn into the strip through hydrophilic attraction and when current is applied during focusing.

For cup loading, the focusing tray [BioRad] was assembled as shown in Figure 2.1. The loading cup was placed at the acidic end of the strip and 150-300 μg of protein in 50-100 μL was loaded into the cup and overlaid with paraffin oil. The sample is drawn into the strip when the current is applied and improves resolution of the basic proteins.

Focusing was performed using a Protean IEF [BioRad]. The current was limited to 50 μA per strip and isoelectric focusing was performed using the following method.

Standard 3-step program: slow ramp to 4000 V for 4 h, linear ramp to 10000 V for 4 h, then 10000 V until 120 kVh was reached.

TX-114 prepared samples were focused using a stepped program with rapid ramping: 100 V for 1 h, 300 V for 1 h, 600 V for 1 h, 1000 V for 1 h, 2000 V for 1 h, 4000 V for 40 kVh, 100 V for 12 h.

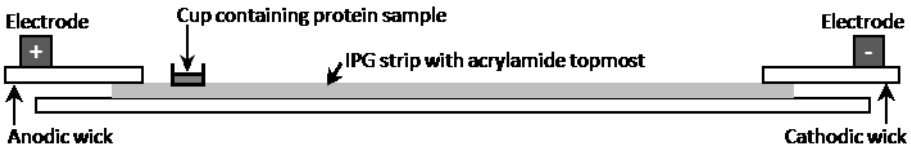


Figure 2.1. Schematic depiction of cup-loading for isoelectric focusing in IPG strips.

2.5.2. Liquid-phase isoelectric focusing (LP-IEF)

500 µg – 2 mg of protein sample was prepared to 2.5 mL in 7 M urea, 2 M thiourea, with or without 1% C7BzO as needed to maintain solubility for denaturing preparations or alternatively in 2.5 mL 0.05-0.1% Triton X-100 for native preparations. The conductivity was checked to ensure it was < 200 µSi.cm^{-1} and sample was desalted using MicroBioSpin [BioRad] columns if necessary. A final concentration of 0.2% w/v pH 3-10 BioLyte ampholytes [BioRad] were added and sample was loaded into the chamber of the BioRad MicroRotofor Isoelectric focusing cell. Focusing was carried out at using anode solution of 0.1 M NaOH and Cathode solution of 0.1 M H_2PO_4 at 1 W constant until the voltage stabilised. The run was allowed to continue for 30 min once voltage stabilised then fractions harvested (10 fractions total). IEF runs were typically completed in < 3 h.

2.5.3. One-dimensional sodium dodecyl sulphate –polyacrylamide gel electrophoresis (1D SDS-PAGE)

Samples to be run were mixed with 2× or 4× SDS sample buffer (Appendix - 0.25 M Tris-HCl pH 6.8; 0.25% w/v SDS; 10% glycerol and 0.0025% w/v bromophenol blue) and boiled for 5 min, or heated to 60°C for 10 min if the sample contained Urea in order to reduce carbamylation. All samples to be loaded were centrifuged to pellet debris for 5 min at 16 000 × *g* prior to loading. Precision Plus™ dual colour or unstained Molecular weight markers [BioRad] were loaded alongside samples.

Gel systems used were either Criterion Bis-Tris 4-12% pre-cast gels (1, 12 or 26 wells) with MES running buffer [BioRad], Criterion TGX 4-20% pre-cast gels (1, 12 or 26 wells) with TGX or TGS running buffer [BioRad] (Appendix).

Bis-Tris gels were run at 160 V for ~1 h and TGX gels were run at 300 V for ~25 min, or until the dye front reached the bottom of the gel.

All gels were fixed with fixing solution (40% Methanol/ 10% Acetic acid) for 30 min at room temperature with constant gentle agitation. Gels were stained to visualise protein with either Flamingo™ fluorescent stain [BioRad] for ~1 h in the dark with gentle agitation, or with Coomassie blue G250 (Appendix) overnight with gentle agitation. Gels stained with Flamingo were scanned with a Pharos FX™ Plus Molecular Imager [BioRad] and gels stained with Coomassie were scanned with an Epson Perfection 4870 Photo flatbed document scanner, set to film.

2.5.4. Two-dimensional sodium dodecyl sulphate –polyacrylamide gel electrophoresis 2D SDS-PAGE

Immobilised pH gradient strips were gently blotted to remove residual oil and equilibrated for 20 min in ~5 mL per strip SDS equilibration solution (2% SDS, 6 M urea, 250 mM Tris-HCl pH 8.5, 0.0025% w/v bromophenol blue) with gentle agitation. Strips were loaded into single well gels alongside Precision Plus™ dual colour or unstained Molecular weight markers [BioRad]. Alternatively, samples separated by LP-IEF were run in 12 well gels. Fractions were mixed with 2× or 4× sample buffer and heated to 60°C for 10 min and centrifuged to pellet insoluble material. Gels were run, fixed and stained according to the methods described in section 2.5.3 above.

2.5.5. Dimethyl labelling of bacterial proteins

Protein labelling was performed on 1 mg of *M. hyopneumoniae* strain J protein by the addition of 40 mM formaldehyde (ultra-pure grade) [Polysciences Inc., USA] in the presence of 20 mM sodium cyanoborohydride, buffered with 100 mM HEPES solution adjusted to pH 6 – 7 in a final volume of 1 mL, and incubated at 37°C for a minimum 4 h. The reaction was quenched by the addition of 100 mM ammonium bicarbonate and precipitated with 8 volumes of acetone and 1 volume of methanol at -20°C for 3 h. The precipitated protein was then pelleted by centrifugation at 14000 × *g* and washed with 5 volumes of methanol. The protein pellet was resuspended in 50 mM sodium hydroxide, pH 8.0 and digested with trypsin prior to peptide enrichment and analysis by LC-MS/MS.

2.5.6. Enrichment of labelled N-terminal peptides

To enrich for blocked N-terminal peptides, the high molecular weight aldehyde-derivatised polymer (HPG-ALD type II) [Flintbox Innovation Network] was added in a ratio 2 μg of polymer per 1 μg of peptides in the presence of 20 mM of sodium cyanoborohydride. The reaction was allowed to proceed at 37°C for a minimum of 4 h. Ammonium bicarbonate was added to 100 mM, adjusted to pH 6 – 7 and incubated for 30 min to quench the reaction. Samples were passed through a 10000 MWCO Ultrasep™ Advance Centrifugal Device by centrifugation at 10000 × *g*. Filtrates containing unbound N-terminal peptides, were acidified with 0.02% Trifluoroacetic acid and reduced to ~15 μL in a Vacufuge™ Concentrator 5301 [Eppendorf, Germany], ready for analysis by LC-MS/MS.

2.5.7. In-gel trypsin digestion

Protein spots or bands of interest were excised from gels using a scalpel. Large gel pieces were further cut to ~ 1 mm^3. Pieces were washed by incubation with wash solution (50 mM NH_4HCO_3, 50% acetonitrile (ACN)), for 10 min at room temperature with vortexing. Wash solution was removed and discarded and the process repeated at least 3 times or until all stain was removed from the gel. Gel pieces were dehydrated with the addition of 100% ACN and incubated for 10 min with vortexing. ACN was removed and discarded. For samples which had not been subjected to reduction and alkylation prior to electrophoresis, an in-gel reduction and alkylation step was required. Dehydrated gel pieces were rehydrated with 5 mM tributylphosphine, 20 mM acrylamide monomers in 100 mM NH_4HCO_3 and incubated at room temperature for 90 min. Residual liquid was removed and discarded and the washing procedure was repeated. Gel pieces were again dehydrated with 100% ACN and the gel pieces air-dried. Pieces were rehydrated with 12.5 $ng.\mu L^{-1}$ Trypsin gold made up in 100 mM NH_4HCO_3 at 4°C for 30 min. NH_4HCO_3 was added to cover the gel pieces and samples incubated at 37°C overnight (or a minimum of 6 h).

Tubes containing gel pieces in solution were incubated in a sonicating water bath for 10 min to encourage peptide extraction from the gel. Solution was removed to a new tube and 50% ACN, 0.2% TFA was added to cover the gel pieces. Sonication was repeated and peptide-containing solutions were pooled with the first extraction. This process was repeated twice. Solutions containing peptides were concentrated to ~15 μL in an

Eppendorf vacuum concentrator at 30°C. All samples were centrifuged for 10 min at 16 000 × *g* to pellet any insoluble material and the supernatant removed to an autosampler vial for analysis by LC-MS/MS. Alternatively, peptides were subjected to desalting using solid phase extraction (as described in section 2.5.8.).

2.5.8. Solid phase extraction (SPE)

Solid phase extraction was performed using C18 ZipTips [Millipore], or OASIS HLB SPE C18 columns, 1cc [Waters]. Columns or tips were pre-treated with 100% ACN (buffer B), equilibrated with 10 column/tip volumes of 98% MilliQ water 2% ACN (buffer A), then sample applied and allowed to bind by repeatedly flowing through the tip/column up to 10 times. Resin was washed with 5-10 column/tip volumes of buffer A and sample eluted in 75% buffer B, 0.2% formic acid. Eluted sample was dried to 25% original volume to remove ACN and sample reconstituted to required volume with MS loading buffer.

2.6. Detection of interactions

2.6.1. Blotting

Proteins were separated by SDS-PAGE and transferred to PVDF membrane using a semi-dry transfer method. Following gel electrophoresis, gels were rinsed with ddH_2O, before equilibration in 100 mL of buffer 1 (described below) for 5 min. PVDF membrane was cut to size and wetted completely with methanol for 2 min, before equilibration in buffer 2 for a minimum of 5 min.

Electroblotting was carried out in an Owl™ semi-dry horizontal transfer apparatus [Thermo Scientific] assembled according to Figure 2.2, with buffer compositions as follows:

- Buffer 1, 2 papers: 40 mM ε-aminocaproic acid, 25 mM Tris, 10% methanol, 0.0005% (w/v) SDS
- Buffer 2, 2 papers : 25 mM Tris, 20% methanol
- Buffer 3, 1 paper : 300 mM Tris

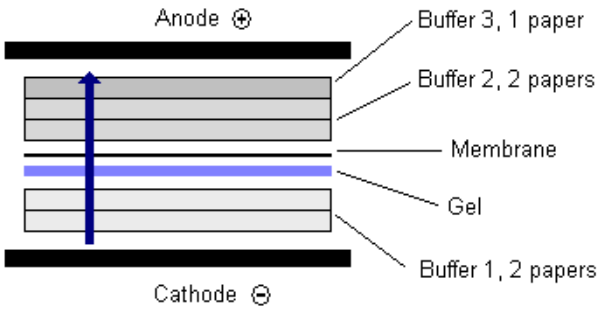


Figure 2.2. Electroblotting apparatus setup.

Transfer was conducted at 300 mA constant for 30 min. Following transfer, membranes were rinsed briefly with ddH_2O and stained with 0.0025% (w/v) Ponceau S in 1% acetic acid to check transfer. Membranes were fully destained in ddH_2O or 1% acetic acid before proceeding with probing.

Membranes were blocked with 5% (w/v) skim milk powder in PBS with 0.1% (v/v) Tween 20 (PBS-T) at room temperature for 1 h.

For detection of immunogenic proteins, membranes were probed with pooled convalescent sera from *Mycoplasma hyopneumoniae* infected pigs diluted 1:100 in PBS-T for 1 h, followed by incubation with peroxidase conjugated anti-pig antibodies diluted 1:3000 in PBS-T for 1 h.

For detection of fibronectin-binding proteins, membranes were probed with human[2] fibronectin [Calbiochem] 10 µg mL^{-1} in PBS-T for 1 h, then rabbit anti-fibronectin [MP Biomedicals, Inc.] diluted 1:3000 in PBS-T for 1 h, then peroxidase conjugated anti-rabbit antibodies diluted 1:1500 in PBS-T for 1 h.

For detection of plasminogen-binding proteins, membranes were probed with either purified biotinylated porcine plasminogen diluted 1:250 in PBS-T for 1 h, followed by incubation with ExtrAvidin-HRP diluted 1:5000 in PBS-T for 1 h; or alternatively probed with 10 µg mL^{-1} purified porcine plasminogen in PBS-T for 1 h, then rabbit anti-plasminogen diluted 1:3000 in PBS-T for 1 h, then peroxidase conjugated anti-rabbit antibodies diluted 1:1500 in PBS-T for 1 h.

[2] Human fibronectin displays sequence similarity to porcine fibronectin and was able to be obtained at higher purity than porcine fibronectin.

Between incubations, membranes were washed in three changes of PBS-T. Membranes were developed with SIGMA*FAST*™ 3,3´-Diaminobenzidine tablets [Sigma-Aldrich] as per manufacturer's instructions.

Control blots showed no naturally biotinylated proteins in *M. hyopneumoniae* cell lysates and there was no cross-reactivity with anti-rabbit antibodies.

2.6.2. Avidin purification of interacting proteins

One 175 cm^2 culture flask containing semi-confluent PK15 cells was washed with ice cold PBS pH 8.0 to remove medium components and was then labelled with 2 mM sulfo-NHS-LC-biotin at 4˚C for 30 min. PBS with 100 mM glycine was added to quench the reaction and remove excess biotin. Glycine was removed by washing in PBS. Cells were lysed in 1.0% Triton X-100 in Tris-HCl pH 7.6, 150 mM NaCl with protease inhibitors on ice with vortexing. After insoluble material was removed by centrifugation, the cleared lysate was added to avidin agarose [Thermo Scientific] and left to incubate on a rotating wheel for 1 h. The slurry was then packed into a column, the flow through collected, and any unbound proteins were removed by extensive washing with PBS.

M. hyopneumoniae cells from a 250 ml culture were pelleted, washed with PBS and gently lysed in 0.5% Triton X-100 in Tris-HCl pH 7.6, 150 mM NaCl with Complete protease inhibitors [Roche] on ice with vortexing and water bath sonication. Insoluble material was removed by centrifugation and the cleared lysate incubated with the biotinylated PK15-avidin agarose mixture overnight at 4˚C on a rotating wheel. The mixture was packed into a column, the flow through collected, and non-interacting proteins were removed by washing 6 times with 25 mM Tris-HCl and 150 mM NaCl, pH 7.4. Interacting proteins were collected by eluting 5 times with 100 mM Tris-HCl and 2 M NaCl, pH 7.4. A secondary elution in 30% Acetonitrile and 0.4% trifluoroacetic acid was performed in order to collect the biotinylated PK15 proteins and any strongly bound *M. hyopneumoniae* proteins which were not eluted by 2M NaCl. The salt and acidic elutions were concentrated separately through 3000 dalton cut-off filters and then acetone precipitated at -20°C for 30 min and finally pelleted by centrifugation at 25 000 × *g* at 4°C for 30 min. Protein was resuspended in SDS sample buffer and separated by 1D SDS-PAGE. Proteins were in-gel trypsin digested and analysed by LC-MS/MS.

2.6.3. Heparin column affinity chromatography

M. hyopneumoniae whole cell lysates were prepared by sonicating washed cell pellets in 10 mM sodium phosphate, pH 7 on ice for 3 rounds of 30 s bursts. Insoluble material was pelleted by centrifugation at 16 000 × *g* for 10 min. 300 µg of *M. hyopneumoniae* soluble protein was loaded into an autosampler vial on a Waters 2690 Alliance LC separations module, and loaded at 0.5 mL min^{-1} onto a 1 mL HiTrap™ Heparin HP column in binding buffer (10 mM sodium phosphate, pH 7). The standard elution program was run at 0.5 mL min^{-1} with continuous gradients as follows: Sample was loaded and washed with 100% binding buffer for 20 min, then 0-50% elution buffer (10 mM sodium phosphate, 2 M sodium chloride, pH 7) for 25 min, increasing 50-100% elution buffer over 10 min, 100% elution buffer for 5 min, then returning to 100% binding buffer. The elution profile was monitored at *A* 210-400 with a Waters 996 photodiode array detector and fractions were collected in 3 min intervals using a Waters fraction collector III.

2.6.4. Avidin purification of fibronectin binding proteins

1 mg of human fibronectin [Calbiochem] was biotinylated with EZ-link sulfo-NHS-biotin [Thermo Fisher Scientific] and then bound to avidin agarose [Thermo Scientific] overnight at 4°C on a rotating wheel. The slurry was then packed into a column, and the flowthrough was collected for monitoring by SDS-PAGE. Unbound biotinylated fibronectin was removed with 4 × 5 ml sequential washes with PBS. Freshly harvested *M. hyopneumoniae* cells were washed extensively (>3 times) in PBS and pelleted by centrifugation (11,000 × *g*, 10 min). Cells were resuspended in 10 ml of 1% (w/v) C7BzO [Sigma-Aldrich] in PBS (pH 7.8) and disrupted with three rounds of sonication at 80% power for 30 s bursts whilst on ice. After insoluble material was removed by centrifugation, the cleared lysate was added to the slurry and incubated again overnight at 4°C on a rotating wheel. The mixture was again packed into a column, the flow through collected, and non-binding proteins were removed by washing with 4 × 5 ml sequential washes with PBS. Fibronectin binding proteins were collected by eluting with 7 M urea, 2 M thiourea, 40 mM Tris and 1% (w/v) C7BzO. The elution was concentrated through 3,000 Da cutoff filters, then acetone precipitated at −20°C

overnight, and finally pelleted by centrifugation at 25,000 × *g* at 4°C for 30 min. Protein was resuspended in SDS sample buffer and separated by SDS-PAGE. Gel lanes were sliced into 10 sections and proteins in-gel digested before analysis by LC-MS/MS.

2.6.5. Avidin purification of plasminogen binding proteins

Plasminogen from human serum [Sigma-Aldrich] was labelled in 20-fold molar excess with Sulfo-NHS-LC-Biotin for 3 h at room temperature. Removal of excess biotin and buffer exchange into PBS was performed in a PD-10 Desalting Column [GE Healthcare, Life Sciences]. Biotinylated plasminogen was incubated with avidin agarose [Thermo Scientific] on a rotating wheel for 5 h. The slurry was packed into a column and the flow through collected. Unbound plasminogen was thoroughly washed with PBS.

M. hyopneumoniae cells from a 250 mL culture were pelleted, washed with PBS, and gently lysed in 0.5% Triton X-100/PBS. Insoluble material was removed by centrifugation at 16 000 × *g* for 10 min. The cleared lysate was incubated with biotinylated plasminogen-avidin agarose mixture overnight on a rotating wheel at 4 ˚C. The mixture was packed into a column and the flow through collected. Unbound proteins were thoroughly washed and collected in PBS, and interacting proteins were eluted with 30% ACN, 0.4% trifluoroacetic acid. The eluting proteins were concentrated using a 3000 Da cutoff filter and acetone precipitated at −20 °C for 30 min. Proteins were pelleted by centrifugation at 25 000 × *g* at 4 °C for 30 min. Protein was resuspended in SDS sample buffer for separation by 1D SDS-PAGE and/or GeLC-MS/MS.

2.6.6. Avidin purification of actin binding proteins

Actin from bovine muscle [Sigma-Aldrich] was solubilised in 8 M urea, 20 mM triethylammonium bicarbonate, pH 8.0. Cysteine residues were reduced and alkylated with 5 mM tributylphosphine and 20 mM acrylamide monomers for 90 min at room temperature. Actin monomers were labelled in 20-fold molar excess with Sulfo-NHS-LC-Biotin for 3 h at room temperature. Removal of excess biotin and buffer exchange into PBS was performed in a PD-10 Desalting Column [GE Healthcare, Life Sciences]. Biotinylated actin was incubated with avidin agarose [Thermo Scientific] on a rotating

wheel for 5 h. The slurry was packed into a column and the flow through collected. Unbound actin was thoroughly washed with PBS.

M. hyopneumoniae cells from a 250 mL culture were pelleted, washed with PBS, and gently lysed in 0.5% Triton X-100/PBS. Insoluble material was removed by centrifugation at 16 000 × *g* for 10 min. The cleared lysate was incubated with biotinylated actin-avidin agarose mixture overnight on a rotating wheel at 4 ˚C. The mixture was packed into a column and the flow through collected. Unbound proteins were thoroughly washed and collected in PBS, and interacting proteins were eluted with 30% ACN, 0.4% trifluoroacetic acid. The eluting proteins were concentrated using a 3000 Da cutoff filter and acetone precipitated at −20 °C for 30 min. Proteins were pelleted by centrifugation at 25 000 × *g* at 4 °C for 30 min. Protein was resuspended in SDS sample buffer for separation by 1D SDS-PAGE and/or GeLC-MS/MS.

2.6.7. Co-Immunoprecipitation using Dynabeads protein A

M. hyopneumoniae whole cell lysates were prepared by sonicating washed cell pellets in PBS, pH 7.4 with 0.05% TX-100 on ice for 3 rounds of 30 s bursts. Insoluble material was pelleted by centrifugation at 16 000 × *g* for 10 min. 50 µL pooled positive convalescent pig sera diluted with 0.1% Tween-20 in PBS (PBS-T) was interacted with 50 µL solution (corresponding to 1.5 mg) of Dynabeads protein A equilibrated with PBS for 20 min at room temperature with mixing. Beads with bound antibody were washed in two changes of PBS-T. 0.5-2 mg of *M. hyopneumoniae* lysate in 100-600 µL PBS-T was incubated with the beads for 20 min at room temperature with mixing. Unbound proteins were removed and the beads washed with three changes of PBS-T, 150 mM NaCl. The beads were resuspended in 100 µL PBS-T, 150 mM NaCl and transferred to a clean tube to prevent elution of proteins non-specifically adsorbed to the tube wall. Bound proteins were collected in two steps; a “native” elution was performed with the addition of 20 µL 1% acetic acid, mixing and incubation for 2 min, collecting into a new tube. A second “denaturing/reducing” elution step was performed with the addition of 1× SDS sample buffer with 20 mM DTT, mixing and incubation at 70°C for 10 min. All fractions, washes and elutions were collected and analysed by SDS-PAGE and/or LC-MS/MS.

2.7. Mass spectrometry and data analysis

2.7.1. 1D LC-MS/MS using QTOF

Using an Eksigent AS-1 autosampler connected to a Tempo nanoLC system [Eksigent, USA], 10 µL of sample was loaded at 20 µL min^{-1} with MS buffer A (2% ACN + 0.2% Formic Acid) onto a C8 Cap Trap column [Michrom, USA]. After washing the trap for 3 min, the peptides were washed off the trap at 300 nL min^{-1} onto an IntegraFrit column (75 µm × 150 mm) packed with ProteoPep II C18 resin [New Objective, Woburn, MA]. Peptides were eluted from the column and into the source of a QSTAR Elite hybrid quadrupole-time-of-flight mass spectrometer [Applied Biosystems/MDS Sciex] using the following program: 5−50% MS buffer B (98% ACN + 0.2% formic acid) over 30 min for gel slices or 15 min for gel spots, 50−80% MS buffer B over 5 min, and 80% MS buffer B for 2 min, 80−5% for 3 min. The eluting peptides were ionized with a 75 µm ID emitter tip that tapered to 15 µm [New Objective] at 2300 V. An Intelligent Data Acquisition (IDA) experiment was performed, with a mass range of 375−1500 Da continuously scanned for peptides of charge state 2+ to 5+ with an intensity of more than 30 counts/scan and a resolution >12000. Selected peptides were fragmented, and the product ion fragment masses were measured over a mass range of 100−1500 Da. The mass of the precursor peptide was then excluded for 120 s for gel slices or 15 s for gel spots.

2.7.2. 1DLC-MS/MS using Ion Trap

This technique was carried out by Professor Paul Haynes at the Australian Proteomics Analysis Facility, Macquarie University.

Each of the 16 SDS-PAGE fractions of triplicate sets was analysed by nanoflow LC-MS/MS (nanoLC-MS/MS) using a LTQ-XL linear ion trap mass spectrometer [Thermo, San Jose, CA]. In a fused silica capillary with an integrated electrospray tip, reversed-phase columns were packed in-house to approximately 7 cm (100 µm ID) using 100 Å, 5 µm Zorbax C18 resin [Agilent Technologies, CA, USA]. An electrospray voltage of 1.8 kV was applied *via* a liquid junction upstream of the C18 column. Samples were injected onto the column using a Surveyor autosampler, which was followed by an initial wash step with buffer A (5% v/v ACN, 0.1% v/v formic acid) for 10 min at 1 µL.min^{-1}. Then, peptides were eluted from the column with 0–50% buffer B (95% v/v ACN, 0.1% v/v formic acid) for 58 min at 500 nL min^{-1}. The column eluate was directed

into a nanospray ionization source of the mass spectrometer. Spectra were scanned over the range of 400–1500 amu and, using Xcalibur software (Version 2.06, Thermo) automated peak recognition, dynamic exclusion and MS/MS of the top six most intense precursor ions at 35% normalization collision energy were performed.

2.7.3. Protein extraction and digestion for strong cation exchange chromatography and MudPIT

To extract and alkylate proteins, lyophilised cells were resuspended in 8 M urea, 100 mM NH_4HCO_3, pH 9 and sonicated with an ultrasonic probe at 80% power for 3 × 30 s on ice. Reduction and alkylation of cysteine and precipitation was carried out as described above. The protein was resuspended in 8 M urea, 100 mM NH_4HCO_3, pH 9 and digested to peptides firstly by addition of 2.5 µg of Endoproteinase LysC [Roche, Switzerland] and incubating overnight at 37°C. The sample was then diluted to 1M Urea by adding 100mM NH_4HCO_3 prior to the addition of 2.5 µg of trypsin [Promega, USA] and incubated at 37°C for 16 h. Formic acid was added to a concentration of 1% and the peptides desalted and concentrated using an OASIS HLB SPE column 1cc [Waters] as per manufacturer's instructions, described in section 2.5.8. Bound peptides were eluted with 75% ACN, 0.2% formic acid. The ACN was removed by lyophilisation to a volume of 50 µl and 5 µl removed for 1DLC-MS/MS using QTOF as described above.

2.7.4. Peptide fractionation by cation-exchange chromatography

After addition of 20 mM KH_2PO_4, 20% ACN, pH 3 (SCX buffer A), the complex peptide sample was separated using a PolyLC polysulfoethyl A column (2.1 mm × 100 mm) pre-equilibrated in SCX buffer A while connected to an Agilent 1200 HPLC system. After sample injection and collection of unbound peptides, retained peptides were fractionated by an increasing gradient (0-50% B in 50 min) of 20 mM KH_2PO_4, 20% ACN, pH 3 + 0.5 M KCl (SCX buffer B). Eluting peptides were monitored at 214 nm and collected in a 96 well plate by peak detection mode. Peptide fractions were then transferred to microtubes and lyophilised to approximately 10 µl by lyophilisation. The peptide fractions were desalted using OMIX C18 SPE pipette tips as per the manufacturer's instructions. Desalted peptide fractions were briefly (2 min) lyophilised

to remove ACN and 10 µl of MS buffer A (2% ACN, 0.2% formic acid) added prior to analysis by 1DLC-MS/MS on the QTOF as described above.

2.7.5. Multidimensional Protein Identification Technology (MudPIT) analysis using FT ICR

This technique was carried out by Mark Raftery at the Bioanalytical Mass Spectrometry Facility, University of New South Wales.

Digested peptides were separated by nano-LC using an Ultimate 3000 HPLC and autosampler system (Dionex, Amsterdam, Netherlands). Ten microliters of sample was loaded onto a SCX micro column (0.76 × 15 mm) containing Poros S10 or S20 resin (Applied Biosystems, Foster City, CA). Peptides were eluted sequentially using 5, 10, 15, 20, 25, 30, 40, 50, 100, 250, 500, and 1000 mM ammonium acetate (20 µL). The unbound load fraction and each salt step were concentrated and desalted onto a micro C18 pre-column (500µm × 2 mm, Michrom Bioresources, Auburn, CA) with buffer A (H_2O:CH_3CN, 98:2; 0.1% formic acid) at 15 $\mu L.min^{-1}$. After a 4 min wash the pre-column was switched (Valco 10 port valve, Dionex) into line with a fritless nano column (75 µm × ~10 cm) containing C18 media (5µ, 200 Å Magic, Michrom). Peptides were eluted using a linear gradient of H_2O:CH_3CN (98:2, 0.1% formic acid) to H_2O:CH_3CN (64:36, 0.1% formic acid) at 350 $nl.min^{-1}$ over 60 min. High voltage 1800 V) was applied to low volume tee [Upchurch Scientific] and the column tip positioned ~ 0.5 cm from the heated capillary (250°C) of a LTQ FT Ultra [Thermo Electron, Bremen, Germany] mass spectrometer. Positive ions were generated by electrospray and the LTQ FT Ultra operated in data dependent acquisition mode (DDA). A survey scan m/z 350-1750 was acquired in the FT ICR cell (Resolution = 100,000 at m/z 400, with an accumulation target value of 1,000,000 ions). Up to the 6 most abundant ions (>3,000 counts) with charge states > +2 were sequentially isolated and fragmented within the linear ion trap using collisionally induced dissociation with an activation q = 0.25 and activation time of 30 ms at a target value of 30,000 ions. m/z ratios selected for MS/ MS were dynamically excluded for 30 s.

2.7.6. MS/MS Data Analysis

The MS/MS data files were searched using Mascot (provided by the Australian Proteomics Computational Facility, hosted by the Walter and Eliza Hall Institute for

Medical Research Systems Biology Mascot Server) against the LudwigNR database (comprised of the UniProt, plasmoDB and Ensembl databases (25,913,601 sequences; 9,279,221,698 residues) with the following parameter settings:

Default samples – Fixed modifications: none. Variable modifications: NQ deamidation, propionamide, oxidised methionine. Enzyme: semitrypsin. Number of allowed missed cleavages: 3. Peptide mass tolerance: 100 ppm. MS/MS mass tolerance: 0.2 Da. Charge state: 2+ and 3+.

Dimethyl-labelled samples – Variable modifications: Dimethylated lysine, Dimethylated N-terminus, NQ deamidation, propionamide, oxidised methionine. Enzyme: semitrypsin. Number of allowed missed cleavages: 5. Peptide mass tolerance: 100 ppm. MS/MS mass tolerance: 0.2 Da. Charge state: 2+ and 3+.

Biotinylated samples – Variable modifications: NHS-LC-Biotinylated lysine, NHS-LC-biotinylated N-terminus, NQ deamidation, propionamide, oxidised methionine. Enzyme: semitrypsin. Number of allowed missed cleavages: 3. Peptide mass tolerance: 100 ppm. MS/MS mass tolerance: 0.2 Da. Charge state: 2+ and 3+.

2.8. Bioinformatics

In-silico analyses were performed using protein sequence data obtained from UniProt (http://www.uniprot.org) with the use of several online resources (UniProt, 2015).

Basic Local Alignment Search Tool (BLAST) searches, performed using the non-redundant protein database at the National Center for Biotechnology Information (NCBI) or UniProt database, were used to identify proteins that share regions of identity or similarity.

Physical data such as theoretical p*I*, molecular weight, amino acid composition and protein stability were collected using the ProtParam tool at ExPASy (Wilkins *et al*., 1999).

Putative transmembrane domains (TMDs) were predicted using TMpred with parameters: TM helix length between 17 and 33 amino acids, score ≥ 500.

Signal peptide predictions were made using SignalP v. 3.0 and/or v. 4.1, organism group: Gram positive bacteria, default D-cutoff values (Bendtsen *et al*., 2004; Petersen *et al*., 2011). Note: Mycoplasmas lack a type 1 signal peptidase and as such do not have standard signal peptides, however this analysis was used to better understand prediction

algorithms and predict proteins that may be classed non-classically secreted, using SecretomeP.

Non-classically secreted proteins were predicted using SecretomeP v. 2.0, organism group: Gram positive bacteria (Bendtsen *et al*., 2005). Proteins were classed as non-classically secreted if they were predicted to be secreted with SecretomeP and were not predicted to possess a standard signal peptide by SignalP.

Lipoprotein prediction was performed using LipoP (http://www.cbs.dtu.dk/services/LipoP/) (Juncker *et al*., 2003; Rahman *et al*., 2008) and LIPO CBU (http://services.cbu.uib.no/tools/lipo) (Berven *et al*., 2006).

Motifs and consensus sequences were located in protein sequences using PattInProt.

Bioinformatic analysis used online resources: ProtParam (Wilkins *et al.*, 1999), PattInProt (Combet *et al*., 2000), TMpred (Wilkins *et al.*, 1999), COILS (Lupas *et al*., 1991) and PONDR VSL2 and VL-XT (Li *et al*., 1999; Romero *et al*., 1997; Romero *et al*., 2001).

The search for coiled-coils was conducted by means of the COILS2 algorithm with window widths of 14, 21, and 28 and the MTIDK matrix (Lupas *et al.*, 1991). The outputs from both weighted and unweighted scans were compared. In addition, the Paircoil2 algorithm, with a minimum window size of 21 and a P-score cutoff of 0.025, was used (McDonnell *et al*., 2006).

Chapter 3. Global protein-centric proteomic approaches highlight the widespread nature of post-translational protein processing in *Mycoplasma hyopneumoniae*

Sections of this chapter have been submitted for peer review to the journal Proteomics.

Jessica L. Tacchi[1], Benjamin B.A. Raymond[1], Paul A. Haynes[3], Iain J. Berry[1], Michael Widjaja[1], Daniel R. Bogema[1,4], Lauren K. Woolley[4,5], Cheryl Jenkins[4], F. Chris Minion[6], Matthew P. Padula[1,2], Steven P. Djordjevic[1,2*]

[1] The ithree institute, University of Technology, Sydney. PO Box 123, Broadway, NSW, 2007, Australia.
[2] Proteomics Core Facility, University of Technology, Sydney. PO Box 123, Broadway, NSW, 2007, Australia.
[3] Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW 2109, Australia.
[4] NSW Department of Primary Industries, Elizabeth Macarthur Agricultural Institute, Menangle, NSW 2568, Australia.
[5] School of Biological Sciences, University of Wollongong, Wollongong, NSW 2522, Australia.
[6] Department of Veterinary Microbiology and Preventative Medicine, Iowa State University, Ames, Iowa 50011, United States.

Author contributions:

Tacchi, J. L. Performed proteomic experiments, analysed data and wrote the paper.
Raymond, B. B. Performed TX114 biotinylation experiments, plasminogen binding and PK15 interaction analyses.
Haynes, P.A. Carried out ion trap mass spectrometry analyses.
Berry, I. J. Performed N-terminal dimethyl labelling experiments and modelling.
Widjaja, M. Performed fibronectin binding experiments.
Bogema, D.R. Expressed and purified P65 and generated polyclonal antisera.
Woolley, L.K. and Jenkins, C. Provided convalescent porcine sera.
Minion, F.C. Provided the construct for P65 expression and purification.

Padula, M. P. Carried out QTOF mass spectrometry and assisted data analysis and validation.

Djordjevic, S. P. Conceived the study and assisted in the writing of the paper.

3.1. Abstract

Mycoplasma hyopneumoniae is a genome-reduced, cell wall-less, bacterial pathogen with a predicted coding capacity of less than 700 proteins and is one of the smallest self-replicating pathogens. These features make it an attractive model to study proteolytic processing as a means to generate functional diversity. Here we present analyses of the global proteome of *M. hyopneumoniae* strain J using a multitude of techniques, including peptide-centric (2D LC-MS/MS and peptide isoelectric focusing) and protein-centric approaches (1D and 2D GeLC-MS/MS) which identified expressed protein products for 52% of the predicted proteome (347 proteins). Notably, our analyses highlight the prevalence of post-translational proteolytic processing of *M. hyopneumoniae* surface-associated proteins. Adhesins, lipoproteins, and proteins with well-characterised cytosolic functions that are also detected at the cell surface were shown to be proteolytically processed. Cleavage fragments were recovered from affinity chromatography assays using heparin, fibronectin, actin and host epithelial cell surface proteins as bait, suggesting these fragments retain biologically-important roles in pathogenesis. We hypothesise that protein processing is underestimated as a post-translational modification in genome-reduced bacteria and represents an important mechanism for creating cell surface protein diversity, potentially to compensate for a reduction in coding capacity at the genome level.

3.2. Introduction

The current trend in global proteomic analysis has been to use high-speed, ultra-sensitive mass spectrometers combined with orthogonal upfront chromatographic fractionation (i.e. 2DLC-MS/MS) in a peptide-centric manner to characterise proteomes. These high-throughput protocols rely on all proteins in a sample being digested with an efficient protease (e.g. trypsin) into peptides for downstream analysis. Peptide-centric or "bottom-up" approaches are widely used because peptides are more readily solubilised for fractionation and are amenable to chromatographic separation, and mass spectrometry is more sensitive when analysing peptides, rather than intact proteins (Mann *et al*., 2013). Conversely protein-centric approaches aim to preserve intact proteins throughout fractionation steps so that proteoform information may be retained (Smith *et al*., 2013) and then discrete proteins or fractions are digested to peptides and analysed individually by mass spectrometry. Protein-centric methods are thus not necessarily "top-down" approaches which aim to analyse individual intact proteins by mass spectrometry (Kelleher, 2004). Without selective enrichment, high-throughput peptide-centric approaches can fail to capture post-translational proteolytic modifications and can lead to an oversimplification of the complexity of the proteome.

Mycoplasma spp. are bacteria that evolved by a process of degenerative evolution from the low G + C Firmicutes. Mycoplasmas have lost genes for cell wall biosynthesis and many anabolic processes including a TCA cycle and are reliant on glycolysis for the production of cellular ATP (Razin *et al*., 1998; Woese, 1987). Mycoplasmas typically have small genomes of less than 1000 kbp and are dependent on the host for the supply of cholesterol for membrane biosynthesis, amino acids, nucleotides and other macromolecular building blocks for cell growth (Minion *et al*., 2004). As such, Mycoplasmas are excellent model organisms to examine the complexity of post-translational modifications in microbial proteomes. Despite an apparent wealth of genomic information until recently, there has been a notable paucity of proteomic information described for genome-reduced bacteria, with much of it particular to the human pathogen, *Mycoplasma pneumoniae* (Kuhner *et al*., 2009; Maier *et al*., 2013; van Noort *et al*., 2012).

Mycoplasma hyopneumoniae is a geographically-dispersed respiratory pathogen that inflicts significant losses to swine production, estimated at approximately a billion dollars per annum (Clark *et al*., 1991). In addition to these direct losses, antibiotics are used heavily to prevent or curtail outbreaks of mycoplasmal pneumonia (Cromwell, 2002; Jordan *et al*., 1998). Antibiotics used in intensive pig production systems are also released into soil and aquatic ecosystems and onto pastures used for plant production (Jechalke *et al*., 2014). As a consequence, the use of antibiotics in swine production drives the evolution of multiple antibiotic-resistant pathogen and commensal microbial populations in diverse environments (Faldynova *et al*., 2013; Vicca *et al*., 2004). Efficacious and cost-effective vaccines for the control of *M. hyopneumoniae* infections represent a key strategy to alleviate reliance on antibiotics.

Complete genome sequences of four geographically distinct strains of *M. hyopneumoniae* are available (Liu *et al*., 2011; Minion *et al.*, 2004; Vasconcelos *et al*., 2005), shedding light on the metabolic capacity, host specialisation and the evolutionary background of this minimal organism. Genomes range in size from 850-920 kb and encode approximately 700 open reading frames (ORFs). The *Mycoplasma hyopneumoniae* strain 232 genome contains 691 known proteins and 728 annotated genes. A recent proteome analysis of strain 232 identified 8607 unique peptide sequences (false discovery rate of 0.53%) confirming the expression of 70% (483) of the 691 predicted ORFs during culture in Friis broth. This included 171 of the 328 predicted hypothetical proteins (52%), 80% of the lipoprotein genes, and all the P97/P102 adhesin gene families. In the same study, proteogenomic analysis of strain 232 uncovered previously unidentified genes and 5′ extensions to several genes (Pendarvis *et al*., 2014).

Transcriptome studies indicate that 92% of predicted ORFs are transcribed in *M. hyopneumoniae* strain 7448 (Siqueira *et al*., 2014). Seventy-eight non-coding RNAs were also identified in the analysis. Genes with the highest expression levels primarily encoded proteins involved in basal metabolism, as well as chaperones, adhesins, surface proteins, transporters and RNase P. A number of uncharacterised proteins were also identified. The *M. hyopneumoniae* gene encoding the P216 adhesin protein was also

presented with a significant number of transcripts (RPKM – Reads Per Kilobase of transcript per Million mapped reads: 10,796.4) (Siqueira *et al.*, 2014).

During initial infection, *M. hyopneumoniae* strongly adheres specifically along the entire length of cilia of ciliated epithelial cells that line the trachea, bronchi and bronchioles in the upper respiratory tract of pigs. This association causes ciliostasis, loss of cilia and eventual epithelial cell death, which effectively perturbs mucociliary function. The P97 and P102 adhesin families are central to mediating attachment of *M. hyopneumoniae* to epithelial cilia (Bogema *et al*., 2012; Bogema *et al*., 2011; Deutscher *et al*., 2010; Deutscher *et al*., 2012; Hsu *et al*., 1997; Jenkins *et al*., 2006; Seymour *et al*., 2010; Seymour *et al*., 2011; Seymour *et al*., 2012; Wilton *et al*., 2009; Zhang *et al*., 1994). Notably, all members of the P97 and P102 adhesin families are processed post-translationally to the extent that it is difficult to find evidence of adhesin pre-proteins (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al*., 2004; Raymond *et al*., 2014; Raymond *et al*., 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2012; Tacchi *et al*., 2014; Wilton *et al.*, 2009). Most members of the P97 and P102 families are processed via highly efficient cleavage events typically at S/T-X-F↓-X-D/E sites, but also within stretches of hydrophobic amino acids and by numerous, less-efficient cleavage events often in a manner consistent with trypsin-like activity (Moitinho-Silva *et al*., 2013; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Tacchi *et al.*, 2014). Consequently the surface protein architecture of *M. hyopneumoniae* displays cleavage fragments derived via processing of the P97 and P102 adhesin families by several endopeptidases. Peptide-centric, high throughput mass spectrometric analyses largely fail to detect these critical post-translational events. What is unclear is how endoproteolysis alters the presentation of surface proteins not related to the P97 and P102 adhesin families including members of the lipoprotein family.

Cleavage fragments of the P97 and P102 adhesin families are the adhesive effector molecules on the surface of *M. hyopneumoniae* and are efficient binders of glycosaminoglycans that line the respiratory epithelia (Bogema *et al.*, 2011; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Raymond *et al.*, 2013; Tacchi *et al.*, 2014; Wilton *et al.*, 2009); fibronectin, a key component of host extracellular matrix (Deutscher *et al.*, 2010; Seymour *et al.*, 2010; Seymour *et al.*, 2012); and plasmin(ogen), a serine protease

that regulates fibrinolysis and other key proteolytic cascades in the host (Bogema *et al.*, 2012; Robinson *et al*., 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012). The recruitment of plasmin(ogen) to the cell surface of *M. hyopneumoniae* is significant in pathogenesis because the bound plasmin is capable of degrading host extracellular matrix and scavenging nutrients from the host. The lungs of pigs experimentally-infected with *M. hyopneumoniae* show higher bacterial loads, elevated levels of plasmin and inflammatory cytokines and more severe lung lesions compared to pigs vaccinated with a commercial bacterin formulation and experimentally-challenged with *M. hyopneumoniae* (Woolley *et al*., 2013). These data are consistent with the growing body of evidence linking the plasminogen/plasmin system and inflammation (Carmo *et al*., 2014; Raymond and Djordjevic, 2015; Syrovets and Simmet, 2004).

Here we used a combination of peptide- and protein-centric approaches to catalogue the global proteome of *Mycoplasma hyopneumoniae* strain J and assess the extent of proteolytic processing in the proteome of this species.

3.3. Methods

3.3.1. Preparation of M. hyopneumoniae *whole cell lysate*

M. hyopneumoniae (strain J) was grown in modified Friis broth (Friis, 1975) and harvested as described previously (Scarman *et al*., 1997). A 0.1 g pellet of *M. hyopneumoniae* cells was resuspended in 7 M urea, 2 M thiourea, 40 mM Tris-HCl pH 8.8, 1% w/v C7BzO and disrupted with four rounds of sonication at 50% power for 30 s bursts on ice. Proteins were reduced and alkylated with 5 mM tributylphosphine and 20 mM acrylamide monomers for 90 min. Insoluble material was pelleted by centrifugation and the remaining soluble protein was precipitated in five volumes ice cold acetone for 30 min and the pellet air dried.

For 1D SDS-PAGE, the pellet was resuspended in SDS sample buffer (0.25 M Tris-HCl pH 6.8; 0.25% w/v SDS; 10% glycerol and 0.0025% w/v bromophenol blue). For 2D-PAGE, protein pellets were resuspended in 7 M urea, 2 M thiourea, 1% w/v C7BzO. If solution conductivity was measured to be > 200 $\mu S.cm^{-1}$, samples were desalted and buffer exchanged into 7 M urea, 2 M thiourea, 1% w/v C7BzO using a microBioSpin column [Bio-Rad] according to manufacturer's instructions.

Preparation of whole cell lysate for N-terminal dimethyl-labelling of proteins and peptides was performed as described previously (Tacchi *et al.*, 2014).

3.3.2. TX-114 Extraction

M. hyopneumoniae cell pellets were resuspended in 1% Triton buffer (1% Triton X-114, 10 mM Tris-HCl pH 8.0, 150 mM sodium chloride, 1 mM EDTA) and extracted as previously described (Deutscher *et al.*, 2010; Jenkins *et al*., 2008). The detergent phase sample was mixed with SDS-sample buffer and separated as for GeLC-MS/MS.

3.3.3. Protein extraction and digestion for Strong Cation Exchange Chromatography

Harvested cell pellets were freeze-dried and lyophilised cells were resuspended in 8 M urea, 100 mM NH_4HCO_3, pH 9 and sonicated with an ultrasonic probe at 80% power for 3 × 30 s on ice. Reduction and alkylation of cysteine and precipitation was carried out as described above. The protein was resuspended in 8 M urea, 100 mM NH_4HCO_3, pH 9 and digested to peptides firstly by addition of 2.5 µg of Endoproteinase LysC [Roche, Switzerland] and incubating overnight at 37°C. The sample was then diluted to

1 M urea by adding 100 mM NH_4HCO_3 prior to the addition of 2.5 µg of trypsin [Promega, USA] and incubated at 37°C for 16 h. Formic acid was added to a concentration of 1% and the peptides desalted and concentrated using an OASIS HLB SPE column [Waters, 1cc] as per manufacturer's instructions. Bound peptides were eluted with 75% ACN, 0.2% formic acid. The ACN was removed by lyophilisation to a volume of 50 µl and 5 µl removed for 1DLC-MS/MS using Q-TOF.

3.3.4. Peptide fractionation by strong cation-exchange chromatography (SCX)

After addition of 20 mM KH_2PO_4, 20% ACN, pH 3 (SCX buffer A), the complex peptide sample was separated using a PolyLC polysulfoethyl A column (2.1 mm ID × 100 mm) pre-equilibrated in SCX buffer A while connected to an Agilent 1200 HPLC system. After sample injection and collection of unbound peptides, retained peptides were fractionated by an increasing gradient (0-50% B in 50 min) of 20 mM KH_2PO_4, 20% ACN, pH 3 + 0.5 M KCl (SCX buffer B). Eluting peptides were monitored at 214 nm and collected in a 96 well plate by peak detection mode. Peptide fractions were then transferred to microtubes and lyophilised to approximately 10 µl. The peptide fractions were desalted using OMIX C18 SPE pipette tips as per manufacturer's instructions. Desalted peptide fractions were briefly (2 min) lyophilised to remove ACN and 10 µl of MS buffer A (2% ACN, 0.2% formic acid) added prior to analysis by 1DLC-MS/MS using Q-TOF.

3.3.5. Peptide fractionation by isoelectric focusing (Peptide IEF)

Isoelectric fractionation of peptides was performed using an Agilent 3100 OFFGEL system as per the manufacturer's instructions with the following modifications. A 13 cm pH 3-10 IPG strip [GE Healthcare] was rehydrated with a solution of 7 M urea, 2 M thiourea, 1% C7BzO and 12% glycerol for 30 min before assembling with the OFFGEL frames (12 well). 100 µg of protein digest (prepared as for SCX) was desalted and concentrated using an OASIS HLB SPE column, lyophilised to remove ACN and the resulting peptide mixture was diluted to 1.8 mL with 7 M urea, 2 M thiourea, 1% C7BzO and 12% glycerol and equal volumes loaded into the 12 wells of the assembled OFFGEL frame. Isoelectric focusing was performed overnight using the following program: 150-3000 V over 3 h, 3000-10000 V over 5 h, 10000 V until 100 000 kVh was

reached. Fractions were then recovered into separate tubes and frozen at -20°C until further analysed.

3.3.6. Two dimensional polyacrylamide gel electrophoresis (2D-PAGE)

2D gels were run using 250 µg of whole cell lysate with 0.2% pH 3-10 carrier ampholytes [Bio-Rad]. Isoelectric focusing was performed using 11 cm pH 4-7 IPG strips [Bio-Rad] and 11 cm pH 6-11 immobiline drystrips [GE healthcare]. Focusing was carried out using a Protean IEF system [Bio-Rad] at a constant 20°C and 50 µA current limit per strip with a 3-step program: slow ramp to 4000 V for 4 h, linear ramp to 10000 V for 4 h, then 10000 V until 120 kVh was reached. Following IEF, the strips were equilibrated with 5 mL equilibration solution (2% SDS, 6 M urea, 250 mM Tris-HCl pH 8.5, 0.0025% (w/v) bromophenol blue) for 20 min before the second dimension SDS-PAGE. The second dimension gels were run using precast Bio-Rad TGX midi gels with TGS running buffer [Bio-Rad]. Reference gels were stained with Coomassie blue G250 overnight and destained with 1% acetic acid to remove background. All visible spots (180 from the pH 4-7 gel and 160 from the pH 6-11 gel) were manually excised from the gel and subjected to in-gel trypsin digestion, before analysis by LC-MS/MS.

3.3.7. 1D gel electrophoresis liquid chromatography tandem mass spectrometry (GeLC-MS/MS)

150 µg protein from any preparation were separated by SDS-PAGE and fixed and stained with Coomassie Blue G-250. Additionally, a high-load lane was run using 500 µg protein from whole cell lysates. Entire gel lanes were cut into 16 equal slices for whole cell lysates, 30 for the high-load lane or 15 for the TX-114 fraction. Gel slices were further diced into ~1 mm^2 cubes, destained, washed and digested in-gel with trypsin for analysis. Identification of proteins was performed following clean-up of peptide fractions using OMIX C18 SPE pipette tips, using one of the LC-MS/MS methods described below.

3.3.8. Expression of recombinant proteins and creation of polyclonal antisera

Expression of recombinant P65, creation of polyclonal antisera was carried out as described previously (Bogema *et al.*, 2012; Jenkins *et al.*, 2006; Minion *et al*., 2003).

3.3.9. Blotting

Proteins separated on narrow range pH 6-11 2D gels were transferred to PVDF membranes as previously described (Deutscher *et al.*, 2012). Blots were blocked with 5% (w/v) skim milk powder in PBS with 0.1% Tween 20 (v/v) (PBS-T) at room temperature for 1 h. For detection of immunogenic proteins, membranes were probed with pooled convalescent sera collected from low health status *Mycoplasma hyopneumoniae* infected pigs described previously (Bogema *et al.*, 2012) diluted 1:100 in PBS-T for 1 h, followed by incubation with peroxidase conjugated anti-pig antibodies diluted 1:3000 in PBS-T for 1 h. For detection of adhesin R1 cilium binding domains, membranes were probed with a antisera raised against the F3 recombinant fragment that spans the R1 cilium binding domain of MHJ_0194 ($F3_{P97}$); described previously (Jenkins *et al.*, 2006) diluted 1:100 in PBS-T for 1 h, then peroxidase conjugated anti-rabbit antibodies diluted 1:1500 in PBS-T for 1 h. For detection of P65 fragments, membranes were probed with antisera raised against recombinant P65 diluted 1:200 in PBS-T for 1 h, then peroxidase conjugated anti-rabbit antibodies diluted 1:2000 in PBS-T for 1 h. Membranes were washed in 3 changes of PBS-T between incubations and were developed with SIGMA*FAST*™ 3,3´-Diaminobenzidine tablets [Sigma-Aldrich] as per manufacturer's instructions.

3.3.10. Affinity chromatography for identification of protein interactions

Heparin affinity chromatography and avidin purification of fibronectin-binding proteins and PK15 cell surface protein interactors were performed as described previously (Raymond *et al.*, 2014; Raymond *et al.*, 2013; Tacchi *et al.*, 2014).

Avidin purification of actin- and plasminogen-binding proteins was carried out as follows. Actin from bovine muscle [Sigma-Aldrich] was solubilised in 8 M urea, 20 mM triethylammonium bicarbonate, pH 8.0. Cysteine residues were reduced and alkylated with 5 mM tributylphosphine and 20 mM acrylamide monomers for 90 min at room temperature. Actin monomers were labelled in 20-fold molar excess Sulfo-NHS-LC-Biotin for 3 h at room temperature. Plasminogen from human serum [Sigma-Aldrich] was labelled in 20-fold molar excess Sulfo-NHS-LC-Biotin for 3 h at room temperature. Excess biotin was removed by buffer exchange into PBS using a PD-10 Desalting Column [GE Healthcare, Life Sciences]. Biotinylated actin and plasminogen were incubated with avidin agarose [Thermo Scientific] on a rotating wheel for 5 h. The

separate slurries were packed into columns and the flow-through collected from each. Unbound ligand was thoroughly washed with PBS. *M. hyopneumoniae* cells were pelleted, washed with PBS, and gently lysed in 0.5% Triton X-100/PBS. Insoluble material was removed by centrifugation and the cleared lysate was incubated with biotinylated ligand-avidin agarose mixtures overnight on a rotating wheel at 4 ˚C. The mixtures were packed into columns and the unbound proteins were thoroughly washed and collected in PBS. Interacting proteins were eluted with 30% ACN, 0.4% trifluoroacetic acid. The eluting proteins were concentrated using a 3000 Da cutoff filter and acetone precipitated before pelleting by centrifugation. Elutions were subsequently subjected to 1D SDS-PAGE for transfer and detection by blotting or GeLC-MS/MS for protein identification.

Surface proteins were identified by enzymatic cell surface shaving using trypsin for 5 min at 37°C as previously described (Deutscher *et al.*, 2012) and cell surface labelling using Sulfo-NHS-LC-Biotin for 30 s at 4°C as previously described (Bogema *et al.*, 2011).

3.3.11. 1D LC-MS/MS using Q-TOF

These methodologies were performed as described previously (Raymond *et al.*, 2013; Tacchi *et al.*, 2014). Briefly, samples were loaded using an Eksigent AS-1 autosampler connected to a Tempo nanoLC system [Eksigent, USA] at 20 µl.min^{-1} onto a C8 trap column [Michrom, USA] before washing and elution at 300 nL.min^{-1} onto a PicoFrit column (75 µm × 150 mm) packed with Magic C18AQ resin [Michrom, USA]. Peptides were eluted and ionised into the source of a QSTAR Elite hybrid quadrupole-time-of-flight mass spectrometer [AB Sciex] at 2300 V using the following programs: 5−50% MS solvent B (98% ACN + 0.2% formic acid) over 30 min for gel slices or 15 min for gel spots, 50−80% MS B over 5 min, 80% MS B for 2 min, 80−5% for 3 min. An Intelligent Data Acquisition (IDA) experiment was performed, with a mass range of 350−1500 Da scanned for peptides of charge state 2+ to 5+ with an intensity of more than 30 counts/scan. Selected peptides were fragmented, and the product ion fragment masses were measured over a mass range of 50−1500 Da. The mass of the precursor peptide was then excluded for 120 s for gel slices or 15 s for gel spots.

3.3.12. 1D LC-MS/MS using Ion Trap

Peptide samples were analysed by nanoflow LC-MS/MS (nanoLC-MS/MS) using a LTQ-XL linear ion trap mass spectrometer [Thermo, San Jose, CA], using a fused silica capillary with an integrated electrospray tip (75 μm ID × 70 mm) packed with 100 Å, 5 μm Zorbax C18 resin [Agilent Technologies, CA, USA]. An electrospray voltage of 1800 V was applied via a liquid junction upstream of the C18 column. Samples were injected onto the column using a Surveyor autosampler, which was followed by an initial wash step with buffer A (5% v/v ACN, 0.1% v/v formic acid) for 10 min at 1 $\mu l.min^{-1}$. Peptides were eluted from the column with 0–50% buffer B (95% v/v ACN, 0.1% v/v formic acid) for 58 min at 500 $nL.min^{-1}$. The column eluate was directed into a nanospray ionization source of the mass spectrometer. Spectra were scanned over the range of 400–1500 amu and, using Xcalibur software (Version 2.06, Thermo) automated peak recognition, dynamic exclusion and MS/MS of the top six most intense precursor ions at 35% normalization collision energy were performed.

3.3.13. MS/MS data analysis

The MS/MS data files were searched using Mascot (provided by the Australian Proteomics Computational Facility, hosted by the Walter and Eliza Hall Institute for Medical Research Systems Biology Mascot Server) against the LudwigNR database, comprised of the UniProt, plasmoDB and Ensembl databases (vQ209. 8785680 sequences, 3087386706 residues) with the following parameter settings: fixed modifications: none; variable modifications: propionamide, oxidised methionine, deamidated asparagine, n-terminal pyroglutamic acid and carbamoylmethylcysteine cyclization; enzyme: semitrypsin; number of allowed missed cleavages: 3; peptide mass tolerance: 100 ppm or 2.0 Da for data generated by Q-TOF or ion trap instruments, respectively. MS/MS mass tolerance: 0.2 Da or 0.4 Da for data generated by Q-TOF or ion trap instruments respectively; charge state: 2+ and 3+.

Scaffold v3.00.02 [Proteome Software Inc., Portland] was used to validate and compare MS/MS based peptide and protein identifications. Peptide identifications were accepted if their calculated probability was greater than 95.0% with a false discovery rate of 1.27% and protein identifications were accepted if their calculated probability using the Peptide Proph*et al*gorithm was greater than 80.0% with a false discovery rate of 2.4%.

Protein probabilities were assigned by the Protein Proph*et al*gorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

3.3.14. In-silico analyses

Predicted MW and p*I* information for intact proteins and fragments was obtained using ProtParam via ExPASy bioinformatics resource portal (http://web.expasy.org/protparam/) (Wilkins *et al.*, 1999). Transmembrane domain predictions were made using TMPred (http://embnet.vital-it.ch/software/TMPRED_form.html) (Hofmann and Stoffel, 1993) with default minimum 17 and maximum 33 amino acid length of the hydrophobic portion of the transmembrane helix. The PONDR VSL2 algorithm was used to predict regions of protein disorder (http://pondr.com) (Obradovic *et al*., 2005). Molecular modelling was carried out using the UCSF Chimera software (http://www.cgl.ucsf.edu/chimera) (Pettersen *et al*., 2004). All analyses were performed using sequence data obtained from UniProt (http://www.uniprot.org/) (UniProt, 2015).

3.4. Results

3.4.1. The global proteome

GeLC-MS/MS was performed on 3 biological replicates of *M. hyopneumoniae* whole cell lysates, with technical replicates analysed by ion trap MS/MS and Q-TOF MS/MS (Figure 3.1A). 1D GeLC-MS/MS was also performed on a TX-114 detergent fraction and on a high-load lane of whole cell extract (where mass context was not reliably retained due to macromolecular crowding effects) and these were also analysed by Q-TOF MS. In addition, 2DLC experiments were performed in duplicate using SCX followed by 1D LC-MS/MS and peptide IEF was performed in duplicate as an additional upfront fractionation technique. Table 3.1 summarises the identification of proteins expressed in *M. hyopneumoniae* as detected by each of these methods.

Table 3.1. Overview of number of identifications by each method.

Method	Protein IDs	Peptide IDs	Unique Spectra	Proteins unique to method
2DLC	232	689	802	0
Peptide IEF	249	1147	1217	0
Ge Ion Trap	331	2774	3832	7
Ge Q-TOF	297	1701	1961	2
Ge high-load	331	1748	2093	6
Ge TX-114	206	846	897	5

The use of multiple techniques improved confidence in “one-hit wonders”; proteins identified by a single peptide in a single replicate. Adopting the approach of White *et al*. (White *et al*., 2011), if the same single peptide was identified in two or more replicates, the protein was considered to be present, rather than a “one-hit wonder”. Similarly, if a single peptide identified a protein in one replicate and a different single peptide identified the same protein in a separate replicate, the protein was considered to be expressed. Single peptide hits were only retained in the data set if, after being subjected to manual validation, the MS/MS spectra had a considerable sequence of b- and y-ions that were the dominant ions in the spectra. Five proteins were identified to be true one-

hit wonders, with the identifying spectra and fragmentation data shown in Supplementary file S3.2.

Three hundred and forty-seven unique *Mycoplasma hyopneumoniae* strain J proteins, representing ~52% of the predicted proteome, were identified from the combined experiments following analysis by Scaffold [Proteome Software] (Supplementary file S3.1). Interestingly, 2 uncharacterised proteins were identified mapping only to strain 232; an 8.8 kDa protein, Q5ZZV3 identified by 1 peptide in 2 runs on both ion trap and Q-TOF, and an 11.3 kDa protein, Q5ZZV5 identified by 2 peptides in 1 run from ion trap data. A BLAST search of the UniProt database shows these proteins are conserved among strains 232, 7448 and 168, however are not annotated to be present in strain J. Seventy-seven (22%) of the identified proteins are named in UniProt as "Uncharacterised protein", despite some sharing homology with proteins that are well characterised in the literature such as P97 and P102 paralogs, MHJ_0369 and MHJ_0368 (Q4A9W4 and Q4A9W5), homologs of Mhp385 and Mhp384 (Q600R9 and Q600S0) respectively, in *M. hyopneumoniae* strain 232 (Deutscher *et al.*, 2012).

Analysis by GeLC-MS/MS identified all proteins in the dataset and had the additional benefit of retaining protein mass context, enabling further interrogation of the data for post-translational processing events and/or potential genome misannotation. For example, MHJ_0009 encoding a 77.5 kDa uncharacterised protein (Q4AAU0) was identified consistently in slices 6, 13 and 14 of GeLC-MS/MS using ion trap and Q-TOF analyses from whole cell lysates (Figure 3.1B). Peptides identified from replicates of slice 6 mapped to the N-terminus and middle regions of the protein at approximately the predicted mass of the intact protein. Slices 13 and 14 however, are taken from regions of the gel with mass 10-15 kDa, and peptides identified from replicates of these slices mapped only to the C-terminal region of the protein. This C-terminal fragment may represent the product of post-translational proteolytic cleavage.

Predicting the true N-terminus of the C-terminal fragment at amino acid position 567 (M) would generate a protein with a mass of 12.5 kDa and p*I* of 5.47 as predicted by ProtParam. We identified the C-terminal fragment from 2D gels at the same approximate molecular mass, with p*I* ranging from ~5.5 – 6.2 (Figure 3.3). The C-

terminal fragment contains a thioredoxin-like domain, and a BLAST search of this fragment gives ~60% identity to thioredoxin from other *Mycoplasma* species (*M. bovoculi*: E-value: 2e-43, Score: 375, Identity: 62%), suggesting it may have the capacity to carry out thioredoxin-like function in the cleaved form. Notably, *M. hyopneumoniae* already possesses putative thioredoxin (MHJ_0380, Q4A9V3), suggesting this is not likely to be a simple misannotation of thioredoxin. MHJ_0009 was also identified from GeLC-MS/MS of samples following heparin affinity chromatography from a slice at molecular mass of ~10-12 kDa, in elutions carrying proteins with low heparin binding affinity (0-600 mM NaCl). This is consistent with the presence of putative heparin binding regions within the C-terminus. Eight putative heparin-binding regions were identified within MHJ_0009 similar to those previously described (Raymond *et al.*, 2013; Tacchi *et al.*, 2014) in both the N- and C-terminal fragments as denoted by grey underlined regions in Figure 3.1. The protein was identified by the same two C-terminal peptides identified from low molecular mass slices in GeLC-MS/MS as underlined in black in Figure 3.1.

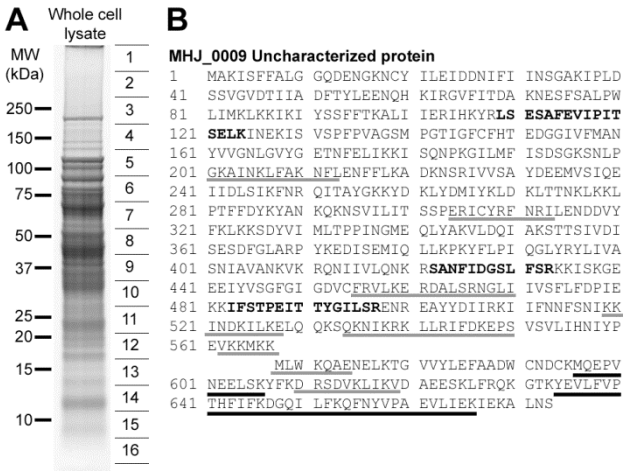


Figure 3.1. Analysis by GeLC-MS/MS.

Panel A shows a representative 1D gel of *M. hyopneumoniae* whole cell lysates. The gel lanes were cut into 16 slices (as shown), digested in-gel with trypsin and analysed by LC-MS/MS using ion trap and Q-TOF instruments, allowing protein mass context to be retained. Panel B shows identified peptides mapping to uncharacterised protein MHJ_0009 (Q4AAU0) in bold. Peptides in bold were identified from gel slice 6 at the approximate predicted intact mass (77 kDa). Peptides underlined in black were generated from proteins identified only from slices 13 and 14. Analysis of the C-

terminal cleavage fragment spanning amino acids 568-664 with ProtParam indicated that it was 12.5 kDa with a predicted p*I* of 5.47 (See also Figure 3.3). Putative heparin binding motifs are underlined in grey.

Overall, 35 proteins (10% of the identified proteome) showed convincing evidence of proteolytic processing through identification from GeLC-MS/MS or 2D gel experiments at molecular masses that were not in agreement with the predicted intact mass (Table 3.2, also denoted by asterisks in Supplementary file 3.1). Ten of these belong to the P97/P102 adhesin families. Consistent with these data, endoproteolytic processing events have been characterised in detail in MHJ_0194 (P97) (Djordjevic *et al.*, 2004; Raymond *et al.*, 2014), MHJ_0493 (P97 paralog P216) (Tacchi *et al.*, 2014; Wilton *et al.*, 2009), MHJ_0663 (P97 paralog P146) (Bogema *et al.*, 2012), MHJ_0369 (P97 paralog Mhp385) (Deutscher *et al.*, 2012), MHJ_264 (P97 paralog Mhp107) (Seymour *et al.*, 2011), MHJ_0195 (P102) (Djordjevic *et al.*, 2004; Seymour *et al.*, 2012), MHJ_0494 (P102 paralog P159/P76/P110) (Burnett *et al.*, 2006; Raymond *et al.*, 2013), MHJ_0662 (P102 paralog P135/Mhp683) (Bogema *et al.*, 2011), MHJ_0638 (P102 paralog Mhp384) (Deutscher *et al.*, 2012) and MHJ_263 (P102 paralog P116/Mhp108) (Seymour *et al.*, 2010). Other proteins showing evidence of cleavage include 5 uncharacterised proteins, 3 known surface antigens, 2 annotated proteases, multiple annotated cytosolic proteins and glycolytic enzymes, such as pyruvate dehydrogenase complex components A, B and D, and lactate dehydrogenase. It is important to note that this list is not exhaustive, as many other proteins were not identified with sufficient sequence coverage to be confirmed as cleavage fragments. For example, a single peptide identified from the middle of an ORF at an apparent mass a fraction of that of the predicted intact mass is insufficient to deduce if a single cleavage event has occurred, if multiple cleavages occur or if the protein is in the process of degradation from normal cellular turnover.

Table 3.2. Cleaved proteins identified from protein-centric analyses.

	Accession	Putative cleaved protein ID	Gene/locus	Intact MW	Identified Mass	Mass from 4-7 2D gels	Mass from 6-11 2D gels	Putative Fn binding	Heparin binding
Adhesins	Q4A925	Putative adhesin like-protein P146 (Bogema *et al.*, 2012)	MHJ_0663	147	120-70, 50-40, ~37-25D	94	35, 38, 45, 92	Intact	95-21
	Q4A926	Uncharacterised protein (Bogema *et al.*, 2011)	MHJ_0662	135	50-45	52, 74	47, 53, 55, 57	-	57-27, 21-16
	Q4A9J1	Putative p76 membrane protein (P159) (Raymond *et al.*, 2013)	MHJ_0494	161	~110-20	(Raymond *et al.*, 2013)	(Raymond *et al.*, 2013)	~20	153-27, 21-16
	Q4A9J2	Putative P216 surface protein (Tacchi *et al.*, 2014)	MHJ_0493	216	~120, 85-20	(Tacchi *et al.*, 2014)	(Tacchi *et al.*, 2014)	-	153-27, 21-16
	Q4A9W4	Uncharacterised protein (Deutscher *et al.*, 2012)	MHJ_0369	114	~30	25	23, 25	-	95-73, 21-16
	Q4A9W5	Putative Lppt protein (Deutscher *et al.*, 2012)	MHJ_0368	109	~50, 30-20D	57	-	-	57-44
	Q4AA66	Putative P97-like protein (Seymour *et al.*, 2011)	MHJ_0264	120	~25	-	100	-	-
	Q4AA67	Putative P102-like protein (Seymour *et al.*, 2010)	MHJ_0263	116	~25, 37, 20	-	19, 41	-	-
	Q4AAD5	Uncharacterised protein (P102) (Djordjevic *et al.*, 2004; Seymour *et al.*, 2012)	MHJ_0195	102	~60 (N), ~42 (C)	-	41-46	22-30, 37-70	73-21
	Q4AAD6	Uncharacterised protein (P97) (Djordjevic *et al.*, 2004; Raymond *et al.*, 2014)	MHJ_0194	123	~120-60, ~37, 30-20	(Raymond *et al.*, 2014)	(Raymond *et al.*, 2014)	(Raymond *et al.*, 2014)	198-153, 122-57, 27-21, 16-12
Surface Ags	P0C0J8	46 kDa surface antigen (p46)	p46 MHJ_0511	46	25-15, >10	44, 47 (multimers)	23, 45	Intact	44-27
	Q4A932	Putative prolipoprotein p65	MHJ_0656	71	71-50	50, Intact	-	Intact, 23-37	73-57
	Q4A981	ABC transporter xylose-binding lipoprotein	xylF MHJ_0606	50	~20D	~50	-	Intact	57-44

Annotated Cytosolic proteins	Q4A9G1	Elongation factor Tu (EfTu)	tuf MHJ_0524	44	44, 21	Fig 2.	Fig 2.	Intact, ~20 C	95-57, 44-27, 21-16
	P0C0J3	L-lactate dehydrogenase (L-LDH) (EC 1.1.1.27) (Immunogenic protein p36)	ldh ictD MHJ_0133	34	34-10, ~20	18, 32, 34	10, 17, 26, 28, 31, 35	Intact	34-21, 16-12
	Q4A9I0	Acetate kinase (EC 2.7.2.1)	ackA MHJ_0505	44	~10 (N)	44	43-45	-	44-34
	Q4A9I1	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	pdhD MHJ_0504	66	Fragments as Multimers	76, 84	-	Intact	95-44
	Q4A9P9	3-hexulose-6-phosphate synthase (EC 4.1.2.-)	sgaH MHJ_0436	25	~15 (C)	23	-	-	44-34
	Q4A9V3	Putative thioredoxin	MHJ_0380	13	10 (C)	10	10, 14	-	-
	Q4AAA3	Periplasmic sugar-binding protein	rbsB MHJ_0227	44	44-13D	40-44	45	Intact	44-34
	Q4AAB1	Putative methylmalonate-semialdehyde dehydrogenase (EC 1.2.1.27)	MHJ_0219	54	~25 (C)	-	51, 53	-	-
	Q4AAL7	Pyruvate dehydrogenase (EC 1.2.4.1)	pdhB MHJ_0112	37	37-15	35-40	31, 35	Intact	34-27
	Q4AAL8	Pyruvate dehydrogenase E1-alpha subunit (EC 1.2.4.1)	pdhA MHJ_0111	42	42-15 + multimers	40, 53, 97	45	Intact	44-34
	Q4AAL9	Adenine phosphoribosyltransferase (APRT) (EC 2.4.2.7)	apt MHJ_0110	19	<10D + multimers	21, 23, 38, 40	31, 35, 45	Intact	16-12
	Q4AAR4	Chaperone protein DnaK (HSP70)	dnaK MHJ_0063	66	66, <50	47, 68	-	Intact, ~40	73-57, 44-34
	Q4AAR8	Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12)	gap MHJ_0031	37	37-~10	35, 38	31, 35	Intact	-
	Q4AAV7	ATP synthase subunit beta (EC 3.6.3.14)	atpD MHJ_0049	52	20-10D (N)	52	26, 43	-	-
Proteases	Q4AAC8	ATP-dependent zinc metalloprotease FtsH (EC 3.4.24.-)	ftsH MHJ_0202	79	79-37	-	-	Intact	73-57, 44-34
	Q4A9G3	Oligoendopeptidase F (EC 3.4.24.-)	pepF MHJ_0522	71	~25, ~50 15D	-	-	-	73-57

Uncharacterised	Q4AA06	Uncharacterised protein	MHJ_0326	25	20-<10D	21	-	Intact	21-16
	Q4A974	Uncharacterised protein	MHJ_0613	95	20-15D (N)	-	100	-	-
	Q4A9G2	Uncharacterised protein	MHJ_0523	230	200-75D (C)	-	-	-	-
	Q4A9Q4	Uncharacterised protein	MHJ_0431	75	~25	-	-	-	-
	Q4AAB8	Uncharacterised protein	MHJ_0212	236	250, ~100	191	100 104, 197	-	198-122
	Q4AAU0	Uncharacterised protein	MHJ_0009	78	Intact, ~12	-	10 (C)	-	12-10

Intact MW shows the calculated mass of the predicted intact protein. Identified masses from 1D GeLC-MS/MS experiments and narrow range 2D gels also shown. Putative Fn binding and Heparin binding show mass ranges at which proteins were identified from fibronectin- and heparin-affinity chromatography GeLC-MS/MS experiments, respectively. Masses provided in kDa. D: identified from TX-114 detergent phase GeLC-MS/MS, (N) or (C) identified fragment mapping to N- or C-terminus of the protein. Shaded indicates not detected in cell surface analyses. References are provided where proteins have been previously characterised.

Only one of the cleaved proteins listed in Table 3.2, the uncharacterised protein MHJ_0523, has not also been identified in surfaceome studies using enzymatic shaving and/or cell surface biotinylation (See Chapter 4). MHJ_0523 encodes a 230 kDa putative lipoprotein and is predicted to possess a transmembrane domain at the N-terminus (TMPred score 1612) and three other putative transmembrane domains (Figure 3.2), which would indicate it should be bound to, or embedded in the cell membrane. Extraction of *M. hyopneumoniae* with TX-114 is likely to have concentrated MHJ_0523 into the detergent-soluble fraction, indicating that it may be surface-exposed but expressed at low levels, rendering it undetectable by our shaving/biotin labelling methods. Detection of MHJ_0523 in slice 1 indicates that the molecule is poorly soluble during SDS-PAGE or that it forms large mass multimeric structures. Fragments identified were from the C-terminus ranging from masses upwards of 75 kDa on the TX114 gel, with no coverage of the first 314 amino acids. Five putative S/T-X-F-X-D/E cleavage motifs were identified along the length of the protein but we were unable to confirm if processing is occurring at these sites.

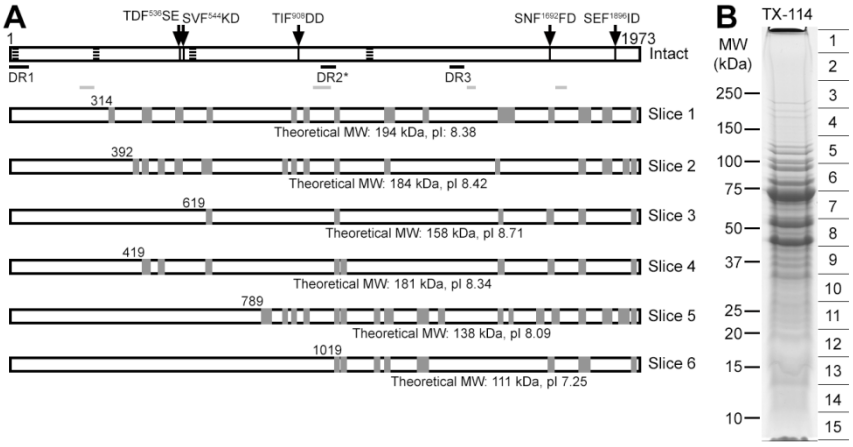


Figure 3.2. MHJ_0523 cleavage map.

Panel A shows general features and peptides mapping to MHJ_0523 identified from GeLC-MS/MS of a *M. hyopneumoniae* Triton X-114 detergent phase enrichment (panel B). Transmembrane domains are indicated by horizontally-striped regions and three disordered regions spanning more than 40 amino acids were detected (DR1-3). Five dominant S/T-X-F-X-D/E cleavage motifs are indicated by arrows. Peptide coverage identified from individual slices is indicated in grey and theoretical molecular weight and isoelectric points of protein fragments (according to peptide coverage).

3.4.2. 2D gel mapping - Processing of cilium adhesins

GeLC-MS/MS preserves the intact molecular weight of proteins and was a valuable strategy to identify cleavage events that affected the migration of members of the P97 and P102 adhesin families (Bogema *et al.*, 2011; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Tacchi *et al.*, 2014). Much finer resolution of cleavage fragments was achieved using narrow range 2D-PAGE. pH 4-7 and 6-11 gels were run using whole cell extracts of *M. hyopneumoniae* (Figure 3.3). Overall, 340 spots comprising 180 spots from a 4-7 isoelectric point gradient gel and 160 spots from a 6-11 isoelectric point gradient gel were resolved well enough to be excised and analysed by LC-MS/MS. Identifications were obtained for 302 spots (159 from p*I* 4-7 and 143 from p*I* 6-11) (Supplementary files S3.3 and S3.4). 130 unique proteins were identified from these 302 spots, representing 19% of the predicted proteome (37% of the identifiable proteome).

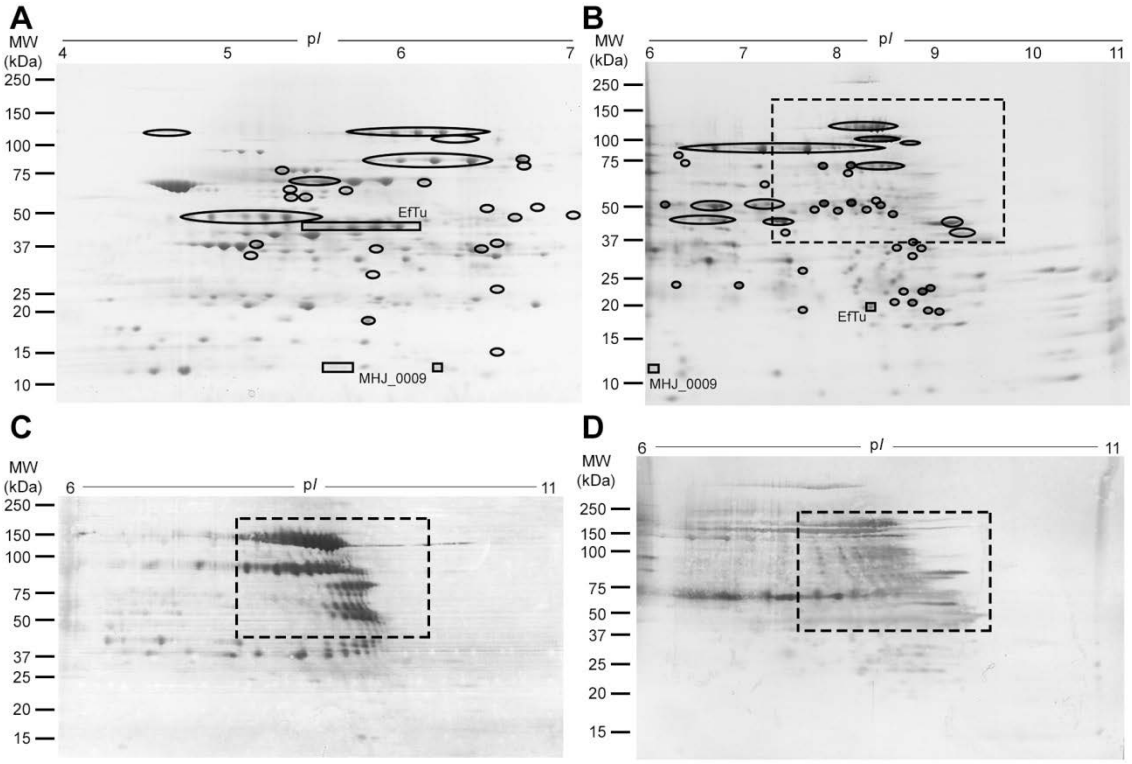


Figure 3.3. 2D gels and immunoblots.

Panel A shows 2D gel image (pH 4-7) with locations of relevant spots indicated. Panel B shows 2D gel image (pH 6-11) with locations of relevant spots indicated and "cloud region" boxed by a dashed line. Spots identified to contain protein cleavage fragments

are circled. Full description of all cut spots and identifications from pH 4-7 and pH 6-11 2D gels can be found in Supplementary files S3.3 and S3.4 respectively. Panel C shows a 2D blot probed with rabbit serum raised against the F3 recombinant fragment that spans the R1 cilium binding domain of MHJ_0194 ($F3_{P97}$). Strongly staining protein fragments carrying regions of R1 or R1-like fragments of the cilium adhesin P97 and Mhp271 show that proline-rich repeats are highly antigenic. Panel D shows a blot probed with pooled convalescent sera from sero-positive pigs. The "cloud regions" are also boxed, showing overlap between adhesin fragments and immuno-reactive regions of the blots.

Eighty-seven proteins were identified from multiple spots. Not all of these, however, could be attributed to processing events, with a significant number of proteins appearing as "spot trains" at a specific molecular weight that track along the p*I* gradient. This is likely to be the result of other post-translational modifications that affect p*I*, such as phosphorylation which has been previously documented in *M. hyopneumoniae* (Wilton *et al.*, 2009). Of particular interest was the presence of "cloud regions" where numerous spots could be detected, but could not be individually resolved (Figure 3.3 - Boxed). These cloud regions are significant, as similar patterns in the same region have been previously identified when *M. hyopneumoniae* proteins were separated over non-linear pH 6-11 gels using a different gel system, carried out in a different laboratory and are thus unlikely to be artefact of sample preparation or gel separation methods (Djordjevic *et al.*, 2004). We postulated that these low abundance cleavage fragments are generated by endoproteolysis of abundantly expressed members of the P97 and P102 adhesin families. A 2D blot probed with rabbit anti-$F3_{P97}$ serum (Jenkins *et al.*, 2006) showed that P97, P66, and a range of lower abundance fragments of MHJ_0194 are recognised (Figure 3.3C). Identical blots probed with a pool of convalescent sera sourced from pigs testing positive for infection with *M. hyopneumoniae* showed a strong reaction to the low abundance P97 and P102 adhesin cleavage fragments (Figure 3.3D). These observations are consistent with the highly immunoreactive nature of proteins carrying proline-rich repeats (Wilton *et al.*, 2009) such as those recognised by anti-$F3_{P97}$ serum.

3.4.3. Evidence that the P65 lipoprotein is processed on the surface of M. hyopneumoniae

P65, MHJ_0656 (Q4A932), comprises 627 amino acids and encodes a 71 kDa lipolytic lipoprotein with preference for short chain fatty acids. The N-terminal 29 amino acids comprise the signal sequence and are expected to be removed followed by lipid modification of the cysteine residue at position 30, generating a mature lipoprotein with a mass of 68 kDa and a p*I* 5.8. We identified P65 as a series of protein spots on a 2D gel with a mass of approximately 68 kDa and with p*I* of 5.8 (Figure 3.4, peptide coverage in black). This 68 kDa molecule was also identified in separate affinity-capture assays using heparin and biotinylated porcine epithelial-like surface proteins as bait (Figure 3.4, peptide coverage in red and blue respectively). P65 is predicted to display three regions of protein disorder from amino acids 189-228 (DR1), 340-418 (DR2) and 553-627 (DR3) according to the PONDR VSL2 algorithm. One of these, DR1, also overlaps with a coiled coil region (100% probability using the COILS algorithm) between amino acids 214-245 suggesting that this region may not be disordered (Ferron *et al*., 2006). Efficient cleavage events are known to occur in S/T-X-F↓X-D/E and related motifs that reside within acidic domains in the P97 and P102 adhesin families in *M. hyopneumoniae* (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2012; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2012; Tacchi *et al.*, 2014). We identified a S/T-X-F↓ X-D/E motif in P65 with sequence ^{360}T-N-F↓D-D^{364} that resides in DR2 and a cleavage site that cuts at phenylalanine with sequence ^{501}V-A-F↓F-A^{505} that is not located within a region of disorder. Both motifs reside within acidic regions that display a p*I* of 5 or less (Figure 3.4). Cleavage at ^{360}T-N-F↓D-D^{364} is expected to generate an N-terminal fragment of 38 kDa and a C-terminal fragment of 30 kDa. Tryptic peptides that mapped to the N-terminal 38 kDa (amino acids 30- 362) and to the C-terminal 30 kDa regions of P65 (amino acids 365-627) were identified when *M. hyopneumoniae* proteins were enriched by extraction with TX-114 and characterised by LC-MS/MS in gel slices representing proteins with masses between 35-45 kDa and 30-35 kDa respectively. Cleavage fragments with these masses were also identified by LC-MS/MS during affinity-capture experiments using fibronectin as bait (see Fragments 1 and 2 in Figure 3.4). Affinity capture experiments using fibronectin as bait also provided evidence that the 30 kDa C-terminal fragment was cleaved at the ^{501}V-A-F↓F-A^{505} site, where cleavage is expected to generate a fragment of 16 kDa. Consistent with

this, several tryptic peptides that mapped between amino acids 363-501 were identified in a gel slice containing *M. hyopneumoniae* proteins with masses between 15 and 23 kDa (Fragment 6 in Figure 3.4). Protein spots migrating with a mass of approximately 50 kDa on 2D gels produced tryptic peptides mapping to P65, consistent with a fragment that started at position 30 and ended at position 503 (Fragment 3 in Figure 3.4), providing further evidence that cleavage occurred at position 503 at the ^{501}V-A-F↓F-A^{505} site.

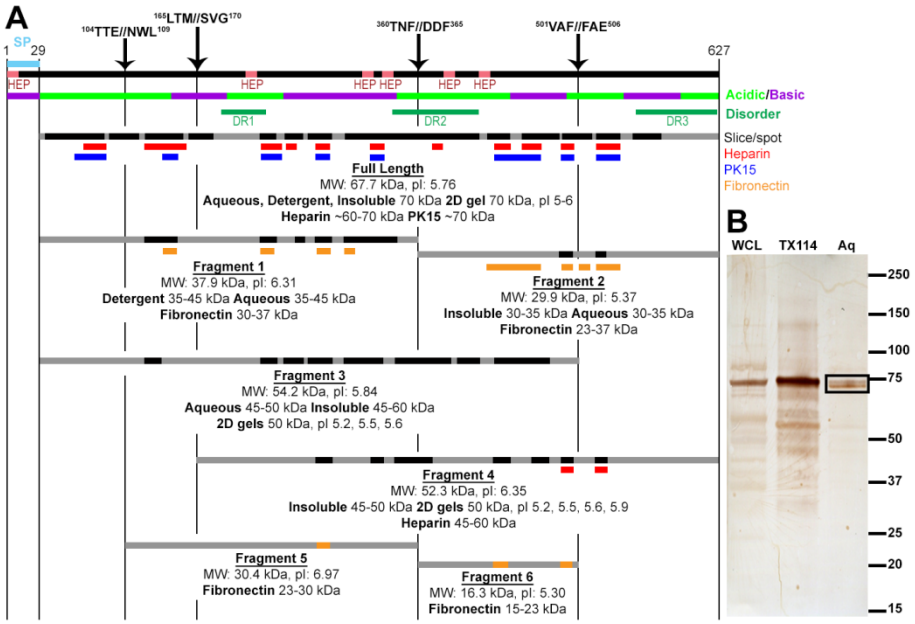


Figure 3.4. Cleavage map of P65.

Panel A shows the major features of P65 including a putative heparin binding site, acidic/basic regions and disordered regions with four proposed cleavage sites. Peptides mapping to protein fragments identified from multiple analyses are indicated. Black regions indicate peptides obtained from gel spot or GeLC-MS/MS slice data. Peptides identified by affinity chromatography using heparin, PK15 surface proteins and fibronectin coupled GeLC-MS/MS data are indicated in red, blue and orange, respectively. Panel B shows a Western blot of *Mycoplasma hyopneumoniae* proteins probed with antisera raised against recombinant P65. The lane labelled WCL contains *M. hyopneumoniae* whole cell lysate. Lanes labelled TX114 and Aq contain biotinylated surface proteins of *M. hyopneumoniae* strain J that partitioned to the detergent and aqueous phases, respectively. Biotinylated proteins were recovered from the Triton and aqueous phases by avidin chromatography prior to gel loading. Multiple cleavage

fragments of P65 were detected at masses lower than the abundant intact from. The boxed proteins in the aqueous phase extract at ~75 kDa could be attributed to the loss of the lipid anchor in the N-terminus, explaining its abundance in the aqueous phase.

Two additional cleavage sites were identified in the N-terminal third of P65. These were identified as ^{104}T-T-E↓N-W-L^{109} and ^{165}L-T-M↓S-V-G^{170}. Cleavage at position 167 is expected to generate two fragments spanning amino acids 30-167 (15.5 kDa; p*I* 4.87) and amino acids 168-627 (52.4 kDa; p*I* 6.35). A fragment with peptide coverage consistent with cleavage at this site was identified by LC-MS/MS in a series of protein spots with mass of approximately 50 kDa and with p*I*s ranging from 5.2-5.9. Peptides mapping to the same fragment were also identified from a gel slice containing Triton-X114 insoluble proteins with masses between 45-50 kDa, and again following heparin affinity purification, in a gel slice containing proteins with masses between 45-60 kDa (Fragment 4 in Figure 3.4). We were unable to find a 15.5 kDa protein spanning amino acids 30-167. Cleavage at amino acid position 106 is expected to generate an N-terminal 8.5 kDa protein (p*I* = 4.36) and a C-terminal 59.3 kDa protein (p*I* = 7.00). While neither of these cleavage fragments was identified in our studies, we did identify a single tryptic peptide in a gel slice spanning 30-35 kDa that contained *M. hyopneumoniae* proteins captured during affinity chromatography using fibronectin as bait (Fragment 5 in Figure 3.4). This fragment is consistent with cleavage at positions 106 and at 362 generating a protein with a mass of 30.4 kDa with a p*I* of 6.97.

3.4.4. Processing events identified in atypical cell surface proteins of M. hyopneumoniae.

Metabolic proteins such as elongation factor Tu (EfTu), pyruvate dehydrogenase complex components A, B and D, lactate dehydrogenase and glyceraldehyde phosphate dehydrogenase showed evidence of post-translational processing (Table 3.2). Evidence that EfTu is extensively processed is presented in Figure 3.5. The N-terminal methionine at position one was lost, presumably by methionine aminopeptidase or leucine aminopeptidase which has recently been shown to reside on the cell surface and is capable of removing methionine residues (Jarocki *et al.*, 2015). EfTu was identified at its predicted mass of 44 kDa and at multiple p*I* between 5.7 - 6.0 on the pH 4-7 2D gel (Figure 3.5A, peptide matches in black). However, we also identified tryptic peptides

spanning the C-terminal half of EfTu from a pH 6-11 2D gel spot at approximately 21 kDa and p*I* 8.5-9 (Figure 3.5A, peptide matches in black). The full length, 44 kDa EfTu protein was also identified in separate affinity-capture assays using heparin, biotinylated fibronectin, actin and porcine epithelial-like surface proteins as bait (Figure 3.5, peptide coverage in red, orange, green and blue respectively). In addition to dimethyl labelling data and the identification of semi-tryptic peptides within the EfTu protein sequence in our proteome datasets, peptide coverage of cleavage fragments in the affinity-capture assays provided strong evidence that EfTu is subjected to proteolytic processing. While further studies are needed to confirm biologically meaningful interactions between EfTu and these host molecules, the affinity-capture assays i) provide independent evidence that regions within EfTu bind host molecules and ii) enrich for lower abundance cleavage fragments and are useful in mapping cleavage events that occur in EfTu.

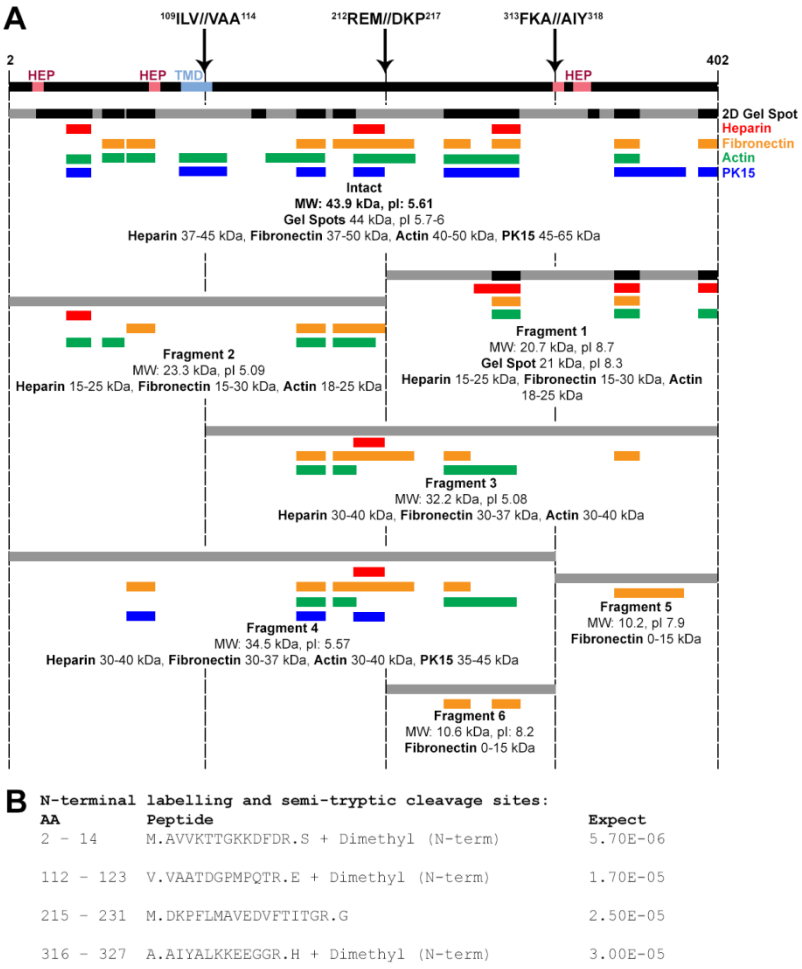


Figure 3.5. Cleavage map of Elongation Factor Tu.

Panel A shows the theoretical cleavage pattern observed in EfTu. Cleavage occurs at amino acid 112 within a sequence that resembles a motif previously identified for P159 (Raymond *et al.*, 2013). Five putative heparin binding domains and a single 18 amino acid transmembrane domain (TMD) between amino acids 98-115 predicted by TMPred, (score 505), are indicated in pink and pale blue regions, respectively. On pH 6-11 2D gels, a distinctive C-terminal fragment of ~21 kDa was identified from a gel spot migrating at p*I* ~8.3. The semi-tryptic peptide ^{215}DKPFLMAVEDVFTITGR^{231} provides evidence that cleavage occurs at position 214 in the sequence ^{212}REM↓DKP^{217}. Peptides mapping to fragments identified from affinity-capture coupled to GeLC-MS/MS are indicated in red, orange, green and blue using heparin, fibronectin, actin or surface proteins from PK15 cells as bait, respectively. Panel B shows N-terminally dimethyl-labelled peptides with confidence scores indicated. Spectra are provided in Supplementary file S3.5.

LC-MS/MS analysis of *M. hyopneumoniae* proteins separated on a 6-11 2D gel identified a distinctive C-terminal fragment of ~21 kDa in a gel spot migrating at p*I* ~8.3. Evidence that this cleavage fragment exists is supported by the identification of multiple semi-tryptic peptides in dimethyl-labelling experiments in the region between amino acids 213-223. Cleavage near this site is expected to have a profound effect on the p*I* of the cleavage fragment as a shift in cleavage upstream by as little as 12 amino acids (position 200) would putatively alter the p*I* from 8.72 (approximately where it was found by 2D-PAGE) to 6.25. Thus, cleavage at position 215 is likely. Tryptic peptides that map to a C-terminal cleavage fragment with a predicted mass of 20.7 kDa were also identified in separate affinity-capture studies using heparin (eluted with 100-600 mM NaCl indicating weaker binding), fibronectin and actin as bait (Figure 3.5, peptides indicated by red, orange and green respectively). Peptide coverage spanning amino acids 2-213 representing the N-terminal half of EfTu (~23 kDa) were also identified by separate affinity-capture studies using heparin, fibronectin, and actin as bait. We were however, unsuccessful in identifying the N-terminal 23 kDa fragment by characterising spots separated by 2D-PAGE.

An endoproteolytic cleavage event was also observed at position 111. Evidence that this event occurs comes from the identification of an N-terminally dimethyl-labelled peptide (V.VAATDGPMPQTR.E) that starts at position 112 (^{109}ILV ↓ VAA^{114}) by LC-MS/MS. This cleavage event produces an N-terminal 11.9 kDa fragment and a C-terminal 32.3 kDa fragment. Peptides mapping to this C-terminal 32 kDa fragment were identified by heparin, fibronectin and actin affinity capture coupled to GeLC-MS/MS. The cleavage event at position 111 in the N-terminal 23 kDa fragment theoretically generates two fragments each approximately 12 kDa in size which we were unable to detect by mass spectrometry methods.

Evidence for a third endoproteolytic cleavage site between amino acids 313-318 comes from the identification of sequential dimethylated and semi-tryptic peptides with sequence (FKA)AIYALKKEEGGR(H). This event would generate an N-terminal 34 kDa fragment and a C-terminal 10 kDa fragment. Affinity capture coupled to GeLC-MS/MS using heparin, fibronectin, actin and porcine kidney epithelial-like cell surface

proteins as bait identified an N-terminal EfTu fragment in gel slices containing proteins with masses between 30-40 kDa that generated tryptic peptides spanning amino acids 66 to 256, consistent with this cleavage event. Furthermore, the additional central cleavage at position 215 is also supported by the identification of the two C-terminal 10 kDa fragments of EfTu that were identified during separate affinity capture studies using fibronectin as bait (Figure 3.5, peptides in orange). Spectra of each of the key peptides that indicate that EfTu undergoes the cleavage events described are depicted in Supplementary File 3.5.

A 3D structural model of EfTu was constructed using Chimera version 1.9.1 based on the crystallized structure of whole, unmodified *Escherichia coli* EfTu (PDB; 1DG1) and is shown in Figure 3.6. A score of 538 (E-value $1e^{-153}$) was obtained for the sequence alignment to Q4A9G1 (MHJ_0524) and a Z-score of -10.07 was obtained for the overall quality of the model. Notably, the N-terminal half of the protein is enriched in helices and the C-terminal half comprises entirely of strands, which was confirmed by bioinformatic predictions of tertiary structure. The N-terminal quarter to the first cleavage site ^{109}ILV↓VAA^{114} is coloured in blue, the second quarter to cleavage site ^{212}REM↓DKP^{217} coloured in yellow, third quarter to ^{313}FKA↓AIY^{318} in purple and the C-terminal quarter in aqua. This 3D model indicates that the four quarters appear to be discrete subunits and the cleavage sites and putative heparin binding sites (highlighted in red) are surface accessible to proteases and substrates, supporting our hypothesis for multiple cleavage events in this molecule.

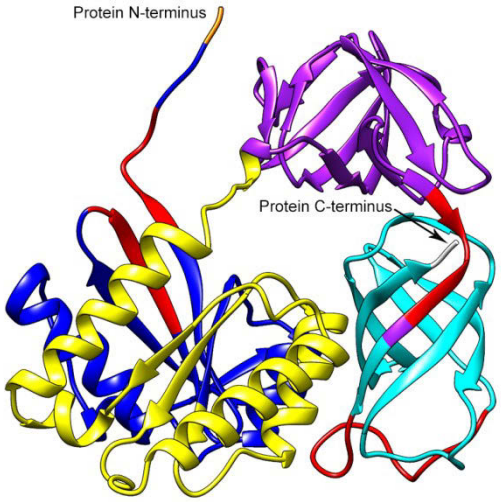


Figure 3.6. Structural model of EfTu.

Model based on whole, unmodified *Escherichia coli* EfTu (PDB; 1DG1). N- and C-termini of the protein indicated in orange and white respectively. The predicted fragments are represented by separate colours with putative heparin binding sites indicated by red overlay. The N-terminal quarter to the first cleavage site ^{109}ILV↓VAA^{114} is coloured in blue, the second quarter to cleavage site ^{212}REM↓DKP^{217} coloured in yellow, third quarter to ^{313}FKA↓AIY^{318} in purple and the C-terminal quarter in aqua. The first half of the protein is enriched in helixes, while the second half contains only strands.

3.4.5. Proteases

Eighteen ORFs have been annotated in the UniProt database (GO annotation) to have putative protease activity, eleven of which have been identified in our study. Identified proteins with annotated endoprotease activity include MHJ_0522 (Q4A9G3) Oligoendopeptidase F, MHJ_0525 (Q4A9G0) Lon protease, MHJ_0636 (Q4A952) tsaD, MHJ_0202 (Q4AAC8) ftsH, and MHJ_0568 (Q4A9B9) an uncharacterised protein. These proteases are likely to carry out the major proteolytic actions that give rise to adhesin fragments, as well as potentially processing other proteins.

Table 3.3. Putative proteases identified in the global proteome of *M. hyopneumoniae.*

Accession & locus	Identified Proteases	Surface	Gene ontology (GO)
Q4AAC8 MHJ_0202	ATP-dependent zinc metallo-protease FtsH (EC 3.4.24.-)	Y	cell division; integral component of membrane; metalloendopeptidase activity; zinc ion binding
Q4A9G0 MHJ_0525	Lon protease (EC 3.4.21.53) (ATP-dependent protease La)	Y	cellular response to stress; cytoplasm; serine-type endopeptidase activity
Q4AAK4 MHJ_0125	Putative aminopeptidase	Y	aminopeptidase activity
Q4A9G3 MHJ_0522	Oligoendopeptidase F (EC 3.4.24.-)	Y	metalloendopeptidase activity; zinc ion binding
Q4A9M4 MHJ_0461	Leucyl aminopeptidase (EC 3.4.11.1)	Y	aminopeptidase activity; manganese ion binding; metalloexopeptidase activity
Q4A929 MHJ_0659	XAA-PRO aminopeptidase (EC 3.4.11.9)	Y	aminopeptidase activity; metalloexopeptidase activity
Q4AAM9 MHJ_0098	ATP-dependent protease binding protein	N	ATP binding; nucleoside-triphosphatase activity; peptidase activity
Q4A952 MHJ_0636	tRNA N6-adenosine threonylcar-bamoyltransferase (EC 2.6.99.4)	N	cytoplasm; iron ion binding; metalloendopeptidase activity
Q4AAG1 MHJ_0169	Methionine aminopeptidase (MAP) (MetAP) (EC 3.4.11.18)	N	metal ion binding; metalloaminopeptidase activity; protein initiator methionine removal
Q4AAS7* MHJ_0022	Signal peptidase I (EC 3.4.21.89)	N	integral component of membrane; serine-type peptidase activity
Q4A9B9 MHJ_0568	Uncharacterised protein	N	serine-type endopeptidase activity

*Although MHJ_0022 has a signal peptidase I signature motif, existing biochemical data from amino-terminal sequence analysis of amino-terminal cleavage products indicates that this species lacks SPase I activity. “Surface” indicates proteins were (Y) or were not (N) identified in cell surface shaving or biotinylation experiments (Chapter 4).

3.5. Discussion

3.5.1. Protein- and peptide-centric approaches to mapping the M. hyopneumoniae *proteome*

The complementary nature of gel and gel-free protein separation approaches enabled us to identify 347 proteins representing 52% of the *M. hyopneumoniae* proteome (Lopez, 2007; Ly and Wasinger, 2011; Maillet *et al*., 2007; Reinders *et al*., 2006; White *et al*., 2011). The unidentified portion of the proteome consisted of 198 uncharacterised proteins, which may be of low abundance, have a high rate of turnover or may not be transcribed under the growth conditions used in our analyses. We also failed to identify 6 ribosomal proteins, cell division proteins FtsY and FtsZ, lipoprotein signal peptidase (MHJ_0027), and 12 putative ABC transporter ATP binding proteins. Twenty-four of fifty predicted lipoproteins were not identified by these analyses, along with 125 unidentified proteins that are predicted to reside within the cell membrane according to PSORT, indicating that solubility may be an issue. In addition, some ORF sequences that remain unidentified contain too many (or rarely too few) lysine and/or arginine residues making the tryptic peptides generated by digestion undetectable by the methods used.

The combination of protein-centric (GeLC-MS/MS and 2D-PAGE) and peptide centric (2DLC and peptide IEF) approaches increased protein coverage compared to any single method and increased confidence in identifications of low abundance proteins and proteins with poor sequence coverage. 2D-PAGE was able to resolve individual proteins and isoforms, providing information about post-translational modifications, while GeLC-MS/MS and gel-free methods are less biased against distinct classes of proteins, making them better suited for global protein identifications. The 1D GeLC-MS/MS run, particularly the high-load gel, performed on the Q-TOF yielded the greatest number of protein identifications consistent with studies by others (Reinders *et al.*, 2006). The protein-centric approaches used in our studies provided insight into the extent of protein processing in *M. hyopneumoniae*. Table 3.2 lists the proteins cleaved in *M. hyopneumoniae*. Notably, almost all of the proteins in Table 3.2 were identified in a comprehensive surfaceome analysis separately using cell shaving and surface biotinylation methodologies (Chapter 4). Our data suggests that protein processing is a

post-translational modification that occurs with greater frequency than is currently recognised and occurs in a wide range of functionally-diverse cell surface proteins.

Enrichment procedures such as TX-114 fractionation and affinity-capture chromatography techniques were useful for enriching the low-abundance proteome and provided clues to protein function. TX-114 extraction enriches for hydrophobic membrane proteins, which partition to the detergent phase (Wise and Kim, 1987b). As *M. hyopneumoniae* lacks a cell wall, the cell membrane is the mediator of contact between the bacteria and extracellular environment; hence membrane-bound proteins are potentially valuable targets for vaccine and therapeutic development. While the TX-114 GeLC-MS/MS protocol detected the fewest protein identifications, at 206, it contributed 5 unique proteins to the overall analysis, all of which were uncharacterised proteins described as lipoproteins and/or predicted to contain transmembrane domains using TMpred. Overall, 26 of 50 *M. hyopneumoniae* lipoproteins were identified by all methods, and LC-MS/MS analysis of TX-114 solublised proteins identified 22 of the 26.

While the precise functions of bacterial lipoproteins remain poorly understood there is mounting evidence to suggest they are PAMP (pathogen-associated molecular pattern) molecules on the surface of Gram-positive bacteria. PAMPs are recognised by Toll-like receptors that trigger innate immune responses (Hashimoto *et al*., 2006; Muhlradt *et al*., 1997; Ozinsky *et al*., 2000; Takeuchi *et al*., 2001). Most mycoplasma lipoproteins are surface exposed with acyl groups anchoring these proteins in the cell membrane, where they are thought to function as cytadhesins, transport proteins or virulence factors with immunomodulatory capabilities (Browning *et al*., 2011). P65 is an abundantly expressed, immunoreactive, and lipolytic lipoprotein that selectively partitions to the detergent phase during extraction with TX-114 (Kim *et al.*, 1990; Schmidt *et al*., 2004). Schmidt *et al*. showed that anti-P65 antibodies inhibit the lipolytic activity of P65 and growth of *M. hyopneumoniae* indicating that P65 performs a primary function on the external membrane surface by providing a source of essential lipids for growth (Schmidt *et al.*, 2004). It has also been suggested that P65 may alter surfactant properties in the lungs of pigs *in vivo* (Schmidt *et al.*, 2004). In our studies, P65 was recovered during affinity capture protocols using different host molecules as bait.

Although these are preliminary data that require quantitative studies to confirm a direct role for P65 in these interactions, this suggests that P65 displays motifs that facilitate binding to a diverse range of host molecules. Consistent with these preliminary observations, we show here for the first time that P65 is a target of several processing events that generate cleavage fragments which are selectively retained during affinity chromatography using porcine epithelial cell surface proteins, fibronectin or porcine heparin as bait. The ability of the cleavage fragments of P65 to bind the same bait proteins as P65 lends weight to the hypothesis that the interactions with host molecules are biologically relevant. Cleavage occurred at a number of sites in the P65 protein sequence including at a phenylalanine residue within a S/T-X-F↓X-D/E motif; a known processing site in the P97 and P102 adhesin families (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Tacchi *et al.*, 2014; Wilton *et al.*, 2009). An immunoblot of biotinylated cell surface *M. hyopneumoniae* strain J proteins fractionated using TX-114 that was probed with anti-P65 polyclonal antibodies identified a 65 kDa protein and numerous smaller mass fragments of P65 consistent with cleavage at several sites within the molecule. These data show that P65 and cleavage fragments of P65 reside on the surface of *M. hyopneumoniae*. Notably, there is clear evidence of a doublet at approximately 65 kDa (boxed in Figure 3.4, panel B) in the lane containing *M. hyopneumoniae* aqueous phase proteins. Previous studies have shown that P65 may undergo clipping at the N-terminus and be a target of further post-translational processing events (Kim *et al.*, 1990). Our data suggest that the doublet may represent forms of P65 that have lost the lipid anchor because they partitioned to the aqueous phase. If correct, these data suggest that a small lipopeptide similar to the macrophage-activating lipopeptide 2 (MALP-2) of *Mycoplasma fermentans* may be produced from P65.

Lipoproteins of mycoplasmal origin are known targets of post-translational processing events. The first 14 amino acids of MALP-404, a 41 kDa lipoprotein in *M. fermentans*, are removed by a post-translational cleavage event generating a 2 kDa MALP-2 lipopeptide. The C-terminal 39 kDa cleavage fragment (known as RF) that results from this cleavage event has been isolated from culture supernatants, but its function remains unknown (Davis and Wise, 2002). Unlike RF, both MALP-2 and MALP-404 are lipid-

modified and remain associated with the membrane of *M. fermentans*. MALP-2 is a potent immunomodulatory molecule that engages Toll-like receptor 2 (Muhlradt *et al*., 1997). MGA0674 is an 82 kDa lipoprotein in *Mycoplasma gallisepticum* whose expression is elevated in virulent strain R_{low} compared to the attenuated vaccine strain F, suggesting that it may play a role in pathogenesis. MGA0674 is a target of a processing event at position 225 that releases a C-terminal 57 kDa fragment from the anchored N-terminal 22 kDa lipoprotein (Szczepanek *et al*., 2010). There are other reports of processing events that target lipoproteins in *Mycoplasma pneumoniae* but their functions have remained poorly characterised (Regula *et al*., 2000).

3.5.2. The extent of proteolytic processing in M. hyopneumoniae

A significant number of the 35 proteins identified to be targets of post-translational processing were glycolytic enzymes and other metabolic proteins. Glycolytic enzymes are increasingly being identified as multitasking or moonlighting proteins in a wide range of organisms including parasites (Gomez-Arreaza *et al*., 2014), yeasts and fungi (Ikeda and Ichikawa, 2014), mammalian cells (Petit *et al*., 2013), plants (Zaffagnini *et al*., 2013), and bacteria (Wang *et al.*, 2014), and this is reflected in the range of entries seen in MultitaskProtDB (Hernandez *et al*., 2014). Processing events alter primary (enzymatic) function and are likely to profoundly influence how cleavage fragments interact with the mycoplasma membrane and host molecules (Sueyoshi *et al*., 2012).

Elongation factor Tu is well-characterised as a key part of the translation machinery in prokaryotes (Dallo *et al.*, 2002). It has also been shown to play important roles on the cell surface of many bacterial pathogens including *Streptococcus pneumoniae* (Mohan *et al*., 2014), *Mycoplasma pneumoniae* (Dallo *et al.*, 2002), *Pseudomonas aeruginosa* (Kunert *et al*., 2007), *Francisella tularensis* (Barel *et al*., 2008), *Listeria monocytogenes* (Archambaud *et al*., 2005), *Leptospira spp.* (Wolff *et al*., 2013) and *Burkholderia pseudomallei* (Nieves *et al*., 2010). EfTu is known to be a complement regulator binding protein and can simultaneously bind Factor H, FHL-1, CFHR1 and plasminogen in *S. pneumoniae* (Mohan *et al.*, 2014), and binds fibronectin in *M. pneumoniae* (Dallo *et al.*, 2002). EfTu and several cleavage fragments of EfTu were retained on affinity capture columns separately loaded with porcine epithelial cell surface receptors, heparin, fibronectin and actin as bait. Notably, we identified several

cleavage sites in EfTu. The masses of the fragments and tryptic peptide coverage of the fragments retained by immunoaffinity capture protocols are consistent with these cleavage events. Our data is consistent with EfTu being a target of several cleavage events on the cell surface of *M. hyopneumoniae*. While quantitative binding studies are needed to confirm the strength and specificity of binding interactions, our data suggests that EfTu is a multifunctional adhesin with discrete binding domains for a range of host molecules. EfTu was also identified to be reactive against convalescent swine serum, a result consistent with immunoblotting studies reported previously in *M. hyopneumoniae* strain 7448 (Pinto *et al.*, 2007). While we were unable to identify the N-terminal fragment that results from the cleavage event producing the C-terminal fragment identified on 2D gels, it was identified from GeLC-MS/MS of detergent phase TX-114 fractions and in several affinity-capture protocols using heparin respiratory epithelial cell surface proteins as bait.

3.5.3. Cleavage of uncharacterised proteins

Analyses of protein spots separated by 2D-PAGE identified a C-terminal cleavage event in the uncharacterised protein MHJ_0009 (Q4AAU0). The C-term cleavage fragment of the putative uncharacterised protein MHJ_0009 (Q4AAU0) shares strong sequence identity with thioredoxin. Further work is needed to confirm if the C-terminal cleavage fragment of MHJ_0009 displays oxidoreductase activity.

3.6. Conclusion

In conclusion, we identified 347 of 672 (52%) putative open reading frames predicted from the genome sequence of *M. hyopneumoniae* strain J. The proteome coverage from well-resolved 2D gels, while low, is unsurprising. The limitations of 2D gels are well documented, particularly considering the nature of sample preparation required, which limits the ability to retain and resolve very basic, acidic, small, large or hydrophobic proteins (Maillet *et al.*, 2007). However, 2D gels provide a technique complimentary to high-throughput 2DLC-MS/MS protocols by maintaining mass and p*I* context, allowing the identification of cleavage products and the extent of proteolytic processing. In addition, the use of LC-MS/MS for spot identification gave improved peptide coverage over more traditional MALDI-based spot identification methods and provided greater

confidence in the identification of cleavage fragments. Cleavage events will undoubtedly complicate efforts to correlate the transcriptome with the proteome in future studies (Gunawardana and Niranjan, 2013; Olivares-Hernandez *et al*., 2011).

Chapter 4. Proteomic and functional analyses elucidate the diversity of surface proteins of the genome-reduced organism *Mycoplasma hyopneumoniae*

This chapter has been written in preparation for submission for publication.

Jessica L. Tacchi[1], Benjamin B.A. Raymond[1], Michael Widjaja[1], Linda Falconer[3], Matthew P. Padula[1, 2], Steven P. Djordjevic[1, 2].

[1] The ithree institute, University of Technology, Sydney. PO Box 123, Broadway, NSW,
2007, Australia.
[2] Proteomics Core Facility, University of Technology, Sydney. PO Box 123, Broadway, NSW, 2007, Australia.
[3] Elizabeth Macarthur Agricultural Institute, NSW Department of Primary Industries, Menangle, NSW 2568.

Author contributions:

Tacchi, J. L. Performed proteomic experiments, analysed data and wrote the paper.
Raymond, B. B. Performed biotinylation experiments and PK15 interaction analyses.
Jarocki, V. M. Performed bioinformatics heparin motif analyses.
Berry, I. J. Performed N-terminal dimethyl labelling experiments.
Padula, M. P. Carried out mass spectrometry and assisted data analysis and validation.
Djordjevic, S. P. Conceived the study and assisted in the writing of the paper.

4.1. Abstract

Mycoplasma hyopneumoniae is an economically significant respiratory pathogen of swine, endemic in pig herds worldwide. Infection by this genome-reduced bacterium heavily relies upon cell surface proteins binding to host components in order to adhere to the cilia of respiratory epithelium, evade immune response and potentially invade host cells. We examined the surface proteome using complementary techniques of enzymatic cell surface shaving with trypsin and intact cell labelling with membrane-impermeable biotin followed by avidin affinity purification of labelled surface proteins. In order to understand the functional aspects of the surfaceome, we used column affinity chromatography techniques to screen for interactions of *M. hyopneumoniae* surface proteins with key host molecules; fibronectin, actin, plasminogen, host epithelial cell surface proteins and also screened for immunogenic proteins. This not only improved our confidence in surface protein localisation, but also began to assign function to previously uncharacterised proteins and elucidate the functions of putative moonlighting proteins. Our analyses identified 159 proteins to be surface exposed. While many important previously-characterised surface proteins were identified, a large proportion of the identified surfaceome was made up of proteins with no known secretion mechanism or transmembrane domains. 130 proteins identified to be surface exposed were also identified by affinity chromatography screening to bind one or more host molecules, or elicit an immune response, many of which are previously described or putative moonlighting proteins. Proteins that are functional at the cell surface, whether through primary or moonlighting functions, present as attractive potential targets for vaccine and therapeutic development.

4.2. Introduction

Pork is the most consumed meat per metric ton worldwide, with China now accounting for more than half of all swine production. The growing population has increased demand for animal-based protein sources and an improvement in the efficiency of raising meat-production animals. One of the biggest challenges facing pig farmers is the prevalence of economically-significant respiratory diseases. Up to 96% of swine herds worldwide experience outbreaks of mycoplasmal pneumonia (Escobar *et al.*, 2002), a chronic, high morbidity, low mortality disease which ultimately leads to reduced feed conversion efficiency. This results in reduced output of product to market and massive economic losses to the agricultural industry estimated in the billions of dollars per annum (Clark *et al.*, 1991). The primary etiological agent of porcine mycoplasmal pneumonia is the genome-reduced bacterium, *Mycoplasma hyopneumoniae*.

M. hyopneumoniae displays a strong affinity for cilia that line the trachea, bronchi and bronchioles in the porcine respiratory tract (Thacker, 2006). The ability to bind and colonise respiratory cilia is the most critical event during early pathogenesis as *M. hyopneumoniae* is not known to survive outside the host. Because *M. hyopneumoniae* lacks a cell wall, the first point of contact between the bacteria and the host is the repertoire of molecules displayed on the cell surface plasma membrane. These molecules function to overcome the mucociliary escalator by assisting with passage of *M. hyopneumoniae* through the respiratory mucous layers, avoiding the innate and acquired immune responses; adhering to and colonising cilia that beat with high frequency (~ 11-15 Hz) (Joki and Saano, 1994). These binding and colonising events lead to ciliostasis, cillial shedding and eventual ciliated epithelial cell death. By drastically compromising the mucociliary escalator in the respiratory tract of infected animals, *M. hyopneumoniae* also creates a favourable environment for infection by other bacterial and viral pathogens (Thacker, 2006), leading to further symptomatic infection. Attachment interactions are thought to be primarily mediated by proteins, as demonstrated by the abolition of adherence to cell monolayers and porcine ciliated epithelial cell suspensions following limited trypsin treatment of *M. hyopneumoniae* cells (Zielinski and Ross, 1992, 1993; Zielinski *et al.*, 1990).

A number of proteins have been implicated in the initial adherence events leading to pathogenesis, the majority of which belong to the now well-characterised P97 and P102 paralogous families of adhesins (Minion *et al.*, 2000; Zhang *et al.*, 1995). These encode large open reading frames, which are expressed and subsequently processed into smaller, functional proteins and presented at the cell surface where they retain the ability to bind key host molecules (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Wilton *et al.*, 2009). Recent research has revealed the prominence of proteolytic processing amongst cell surface proteins of *M. hyopneumoniae*, and the majority of these discoveries have been made using untargeted protein-centric approaches such as 1D and 2D gel electrophoresis. More recently it has been suggested that this surface processing could be linked to ectodomain shedding, a phenomenon well described in eukaryotes that releases protein fragments from the cell surface into the extracellular milieu via a proteolytic cleavage event (Horiuchi, 2013; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Tacchi *et al.*, 2014).

Proteomic studies that have described the surfaceome of Gram negative and Gram positive microbial pathogens have to some degree been hampered by the presence of a cell wall (Cordwell, 2006; Dreisbach *et al*., 2010; Hempel *et al*., 2011; Solis and Cordwell, 2011). A small genome size and the absence of a cell wall are attractive attributes in efforts to define the entire surface proteome of *M. hyopneumoniae* in a protein-centric manner. Although the P97/P102 family of adhesins are critical for colonisation of ciliated epithelial surfaces, few studies, many of which were performed before the availability of the completed genome sequence (Djordjevic *et al*., 1997; Djordjevic *et al*., 1994; Fagan *et al*., 1997; Fagan *et al*., 1996; Fagan *et al*., 2001; Scarman *et al.*, 1997; Schmidt *et al.*, 2004, 2007; Wise and Kim, 1987a, b) have contributed in identifying other surface accessible proteins to further understand host-pathogen relationships at the molecular level.

Recently, the surface proteome of *M. hyopneumoniae* strain 7448 was catalogued using biotin labelling of intact cells, identifying 59 protein species to be present at the cell surface (Reolon *et al.*, 2014). Here, we have utilised complementary methodologies of

biotin labelling and enzymatic cell surface shaving of intact cells, in protein- and peptide-centric approaches respectively, to improve coverage and confidence in identification. We also perform functional analyses at the protein level to assess the likelihood of proteins being genuinely surface-exposed through their ability to bind key host molecules and elicit and immunogenic response.

We present an overview of the expressed, detectable, surface proteome of *M. hyopneumoniae* strain J (in the context of the identifiable proteome presented previously – Chapter 3) and further investigate the proteins present with specific emphasis on evaluating functions that may aid pathogenesis.

4.3. Materials and methods

4.3.1. Culture conditions

M. hyopneumoniae (strain J) was grown in modified Friis broth (Friis, 1975) and harvested as described previously (Scarman *et al.*, 1997).

4.3.2. Surface biotinylation of M. hyopneumoniae *cells*

Freshly harvested *M. hyopneumoniae* cells were washed extensively (>3 times) in PBS and pelleted by centrifugation (4 000 x *g* 10 mins 4°C). Cells were resuspended in PBS (pH 7.8) and biotinylated with Sulfo-NHS-LC Biotin [Thermo Scientific] for 30 s on ice. The reaction was then quenched with the addition of a final concentration of 50 mM Tris-HCl (pH 7.4) and incubated for 15 mins. Cells were washed in three changes of PBS and pelleted by centrifugation.

4.3.3. Enzymatic cell surface shaving procedures

Enzymatic cell shaving using trypsin was performed at 37°C for 5 min as described previously (Deutscher *et al.*, 2010). Intact cells were pelleted by centrifugation and the supernatant containing liberated surface proteins collected on ice to cease trypsin activity. Surface proteins were analysed by 1D gel electrophoresis or further digested to peptides with trypsin prior to analysis by SCX and/or LC-MS/MS.

4.3.4. Preparation of M. hyopneumoniae *samples*

Whole cell lysates were prepared as described previously (Seymour *et al.*, 2012). Precipitated protein was pelleted, air dried and resuspended in buffers appropriate for downstream application. For 1D SDS-PAGE, the pellet was resuspended in SDS sample buffer (0.25 M Tris-HCl pH 6.8; 0.25% w/v SDS; 10% glycerol and 0.0025% w/v bromophenol blue). Alternatively, TX-114 extraction was performed as previously described on untreated or surface biotinylated samples to obtain detergent, aqueous and insoluble phases (Deutscher *et al.*, 2010; Jenkins *et al.*, 2008).

4.3.5. Avidin column affinity chromatography for the purification of biotinylated proteins

Performed as previously described (Nunomura *et al.*, 2005; Raymond *et al.*, 2013). Approximately 1 mg of biotinylated sample from whole cell lysates or TX-114 phase

extractions was applied to a column packed with 1 ml of immobilized monomeric avidin [Thermo Scientific] pre-treated with 30% CH_3CN in 0.4% trifluoroacetic acid and equilibrated with 2 M urea in 100 mM NH_4HCO_3, pH 8.5. After sequential washing with (i) 2 M urea in 100 mM NH_4HCO_3, (ii) 2 M urea in 100 mM NH_4HCO_3 containing 0.5 M NaCl, (iii) 2 M urea in 100 mM NH_4HCO_3 containing 30% CH_3CN, and (iv) 100 mM NH_4HCO_3, the bound peptides were eluted with 30% CH_3CN in 0.4% trifluoroacetic acid.

4.3.6. Column affinity chromatography for the detection of protein interactions

Heparin affinity chromatography, avidin purification of actin-, plasminogen- and fibronectin-binding proteins, and avidin purification of PK15 cell surface protein interactors were performed as described previously (Raymond *et al.*, 2014; Raymond *et al.*, 2013; Tacchi *et al.*, 2014).

4.3.7. Immunoprecipitation with convalescent sera

Convalescent porcine sera was obtained from low-health status finisher pigs from a herd infected with *M. hyopneumoniae* as described previously (Bogema *et al.*, 2012). Serum from sero-positive pigs was pooled for use in immunoprecipitation and blots. Washed cell pellets of *M. hyopneumoniae* were resuspended in PBS with 0.01% Triton X-100 and lysed by sonication at 50% power, 3 × 30 s bursts on ice. Insoluble material was pelleted. Immunoprecipitation was carried out using 50 µL (1.5 µg) of Dynabeads Protein A as per manufacturer's instructions. 50 µL pooled convalescent sera, diluted with 150 µL PBS-T was bound to the beads for 20 min at room temperature with mixing. Beads were washed with PBS-T before ~1.8 mg of *M. hyopneumoniae* protein in 600 µL PBS-T (0.2% tween 20) was incubated with the Protein A/Antibody complex-bound beads at room temperature for 30 min with mixing. Beads were washed four times with PBS-T with 150 mM NaCl to remove unbound protein. Bound proteins were eluted first by the addition of 20 µL elution buffer and incubated with mixing at room temperature for 20 min. Strongly bound proteins were eluted by the addition of 20 µL 1× SDS sample buffer with 20 mM DTT and incubated at 70°C for 10 min. Elutions were fractionated and analysed by GeLC-MS/MS.

4.3.8. Two dimensional gel electrophoresis

2D gels were run using 250 µg of whole cell lysate or purified biotinylated proteins, following avidin chromatography, in 7 M urea, 2 M thiourea, 1% w/v C7BzO with 0.2% pH 3-10 carrier ampholytes [BioRad]. Isoelectric focusing was performed over 11 cm pH 3-10 and 4-7 IPG strips [BioRad] and 11 cm pH 6-11 immobiline drystrips [GE healthcare]. Focusing was carried out and second dimension gels were run, fixed and stained with Coomassie Blue G-250 as described previously (Raymond *et al.*, 2013). All visible spots were manually excised from the gel and subjected to in-gel trypsin digestion and desalting using solid phase extraction by OMIX C18 SPE pipette tips, before analysis by LC-MS/MS.

4.3.9. Protein separation and blotting

Proteins separated by 1D or 2D SDS-PAGE were transferred to PVDF using a semi-dry transfer method as described previously (Kyhse-Andersen, 1984). Membranes were blocked with 5% w/v skim milk powder in PBS with 0.1% Tween 20 v/v (PBS-T) at room temperature for 1 h. Detection of biotinylated proteins was with ExtrAvidin peroxidise conjugate [Sigma-Aldrich] diluted 1:20 000 in PBS-T for 1 h. Detection of plasminogen-binding proteins was performed using biotinylated porcine plasminogen diluted 1:250 in PBS-T, followed by ExtrAvidin peroxidise conjugate [Sigma-Aldrich] diluted 1:1000 in PBS-T as previously described (Seymour *et al.*, 2012). Detection of Fibronectin binding proteins was performed as previously described (Deutscher *et al.*, 2010), or alternatively by using human fibronectin [Calbiochem] biotinylated with EZ-link sulfo-NHS-LC biotin, at 10 $\mu g.ml^{-1}$ in PBS-T for 1 h, followed by 1:5000 ExtrAvidin peroxidise conjugate [Sigma-Aldrich] in PBS-T for 1 h. Detection of immunoreactive proteins was with convalescent pig sera diluted 1:70 in blocking solution for 1 h, followed by incubation with anti-pig peroxidase conjugated antibodies [Sigma-Aldrich] diluted 1:1500 for 1 h. Between incubations, membranes were washed in three changes of PBS-T. Membranes were developed with SIGMA*FAST*™ 3,3′-Diaminobenzidine tablets [Sigma-Aldrich] as per manufacturer's instructions.

4.3.10. Gel electrophoresis liquid chromatography tandem mass spectrometry (GeLC-MS/MS)

Proteins from any preparation were separated by SDS-PAGE, fixed and stained with Coomassie Blue G-250. Entire gel lanes were cut into 16 equal slices, destained,

washed and digested with trypsin in-gel for analysis. Identification of proteins was performed following clean-up of peptide fractions using OMIX C18 SPE pipette tips, using Q-TOF LC-MS/MS methods as described previously (Raymond *et al.*, 2013). MS/MS data was searched using Mascot against the LudwigNR database. Scaffold was used to validate and compare MS/MS protein and peptide identifications as described previously (Raymond *et al.*, 2013).

4.3.11. Bioinformatics

In-silico analyses were performed using protein sequence data obtained from UniProt as described previously (Bogema *et al.*, 2011; Raymond *et al.*, 2013; Tacchi *et al.*, 2014). ProtParam was used to predict molecular mass and p*I* of intact or fragments of proteins as well as provide information about amino acid composition and protein stability. Subcellular localisation prediction was performed using PSORTb (v. 3.0) with parameters as follows; Organism type: Bacteria, Gram stain: Advanced (negative without outer membrane). Putative transmembrane domains (TMDs) were predicted using TMpred with parameters: TM helix length between 17 and 33 amino acids, score ≥ 500. Signal peptide predictions were made using SignalP v. 4.1, organism group: Gram positive bacteria, default D-cutoff values. Mycoplasmas lack a type 1 signal peptidase and as such do not have standard signal peptides, however this analysis was used to better understand prediction algorithms and predict proteins that may be classed non-classically secreted, using SecretomeP. Non-classically secreted proteins were predicted using SecretomeP v. 2.0, organism group: Gram positive bacteria. Proteins were classed as non-classically secreted if they were predicted to be secreted with SecretomeP and were not predicted to possess a standard signal peptide by SignalP. Lipoprotein prediction was performed using LipoP (http://www.cbs.dtu.dk/services/LipoP/) (Juncker *et al.*, 2003; Rahman *et al.*, 2008) and LIPO CBU (http://services.cbu.uib.no/tools/lipo) (Berven *et al.*, 2006).

4.4. Results

4.4.1. The surface proteome of M. hyopneumoniae

Overall, the *Mycoplasma hyopneumoniae* surface proteome identified by the combined methods was found to consist of 159 proteins (Table S4.1). This is 24.6% of the theoretical proteome of *M. hyopneumoniae* (Strain J: 671 ORFs), and 44.7% of the identified proteome described previously (347 identified proteins) (Chapter 3). 4 proteins were identified in surfaceome analyses which were not previously detected in global analyses (shaded in Table S4.1). Biotinylation followed by fractionation at the protein level gave 114 protein identifications (67 unique to biotin) while enzymatic shaving with trypsin identified 92 proteins (45 unique to trypsin), and of these 47 were identified by both methods (Figure 4.1A). Experiments were performed under very mild conditions, with biotinylation carried out for 30 s on ice and shaving performed for only 5 min at 37°C. Shaving controls were run in parallel where no trypsin was added and these samples were collected as "secreted proteins". To ensure cell lysis was minimised during cell handling procedures, a time-course experiment observing protein secretion was performed and analysed by SDS-PAGE alongside cell lysis controls (Figure S4.1).

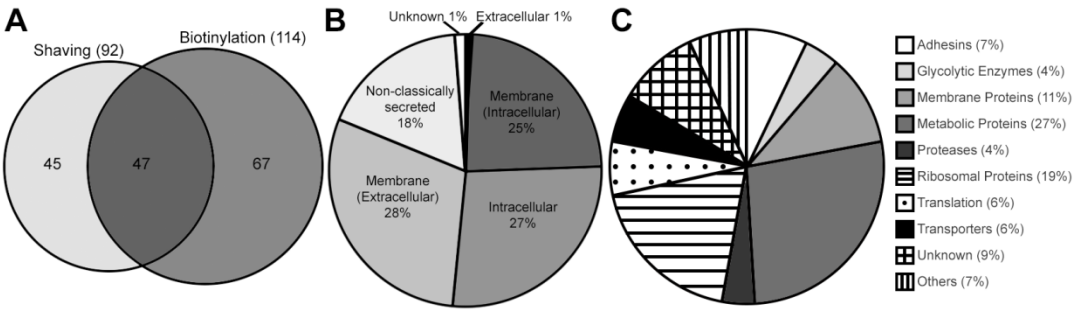


Figure 4.1. Overview of surface protein identifications.

Panel A shows a venn diagram comparing surface protein identifications by shaving and biotinylation methods. A total of 159 proteins were identified to be surface-exposed, with 45 unique to shaving methods, 67 uniquely identified by biotin labelling methods and 47 identified by both methods. Panel B shows the predicted subcellular localisation of all identified surface proteins by a combination of bioinformatic algorithms. C shows the functional assignment of all identified surface proteins. Individual protein identifications, localisations and functions described in Table S4.1.

Shaving methods led to the identification of 92 proteins with surface exposed lysine and arginine residues accessible to trypsin. 45 proteins were identified only by trypsin shaving procedures. This may be due to biotinylation incompatibility (the absence of an accessible lysine residue for labelling, poor ionisation of labelled peptides or inefficient enrichment), poor solubility of membrane proteins, better detection of further fractionated shaved peptides (by SCX), or these proteins may be shed into the extracellular milieu and not interact with the mycoplasma cell surface.

Biotinylation methods identified 114 proteins. More proteins were able to be detected than shaving methods due to the ability to enrich for low abundance proteins and the requirement for only one accessible lysine residue to enable labelling with biotin. A major advantage of this method is that protein mass context is preserved as the proteins are labelled and not enzymatically cleaved from the surface and additionally, this grants us the ability to further fractionate samples at the protein level by Triton X-114 phase extraction. Biotinylated surface proteins may be purified by avidin affinity chromatography and proteins may be then analysed by 1D or 2D GeLC-MS/MS (Figure 4.2). As such, one- and two-dimensional electrophoresis have also been used to investigate the presence of cleavage fragments at the cell surface as has been previously described for the P97/P102 paralog family of adhesins (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Djordjevic *et al.*, 2004; Jenkins *et al.*, 2006; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Tacchi *et al.*, 2014; Wilton *et al.*, 2009) as well as indicated for numerous other proteins (Chapter 3).

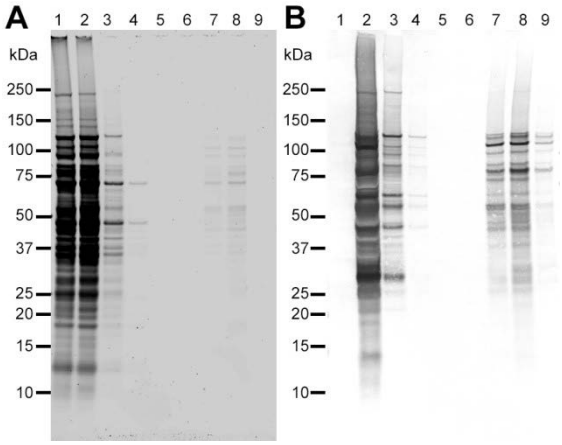


Figure 4.2. Avidin affinity purification of *Mycoplasma hyopneumoniae* surface biotinylated proteins.

Panel A shows coomassie stained reference gel and panel B shows the corresponding blot probed with ExtrAvidin HRP to detect biotinylated proteins. Lanes are as follows: 1. Non-biotinylated whole cell lysates, demonstrating no naturally biotinylated or cross-reactive proteins, 2. Unfractionated surface biotinylated whole cell lysates, 3. Column flow-through of unbound proteins, 4-6. Washes, 7-9. Elutions of biotinylated proteins in 30% ACN, 0.4% TFA.

4.4.2. Bioinformatics

When the global theoretical proteome of *M. hyopneumoniae* strain J is analysed using bioinformatics prediction algorithms, PSORTb 3.0.2 allocates 196 proteins to be membrane-associated, of which 17 (8.7%) were identified in the surfaceome (Yu *et al*., 2010); SignalP v4.1 indicates 52 proteins possess conventional signal peptides for secretion, of which we identified 30 (58%) (Bendtsen *et al.*, 2004); and SecretomeP 2.0 predicts 272 proteins to be non-classically secreted (i.e. not containing a signal sequence) of which 40 (15%) were identified in the surface proteome (Bendtsen *et al.*, 2005) (Table 4.1). The global proteome analyses of *M. hyopneumoniae* strain J (Chapter 3) however, failed to detect 48% of the predicted proteome, suggesting a large number of proteins are expressed below the limits of detection, or not expressed under culture conditions. In the detected global proteome, 71 proteins are predicted to be membrane associated by PSORTb, 26 proteins possess signal peptides and 108 proteins are predicted to be non-classically secreted. Thus, in the context of the detected global proteome, the surfaceome analyses identified 24% (17 of 71) of previously detected

membrane proteins (PSORTb); 77% (20 of 26) of signal peptide containing proteins (SignalP) and 50% (54 of 108) of previously detected non-classically secreted proteins (SecreteomeP).

Table 4.1. Summary of bioinformatics predictions of identified surface proteins.

Localisation	PSORTb	SecretomeP	SignalP	TMPred	Combined/Overall
Cytoplasmic	97	-	-	-	
Membrane	17	-	-	-	
Unknown	45	-	-	-	
Non-Classically Secreted	-	54	-	-	
Secreted	-	16	20	-	
Intracellular	-	89	139	-	
Contain TMDs	-	-	-	86	
Patterns of prediction					
Membrane (Intracellular)	Membrane/unknown/ cytoplasmic	Intracellular	+	+	40
Membrane (Extracellular)	Membrane/unknown/ cytoplasmic	non-classically secreted/secreted	+	+	45
Extracellular	Unknown	non-classically secreted/secreted	+	-	1
Unknown	Membrane	Intracellular	-	-	2
Intracellular	Cytoplasmic/unknown	Intracellular	-	-	43
Non-Classically Secreted	Cytoplasmic/unknown	Non-classically secreted	-	-	28

Number of proteins with corresponding predictions shown.

As it is well-documented that conventional subcellular localisation prediction algorithms struggle to accurately reflect atypical bacteria such as the mollicutes which lack a cell wall (Edman *et al.*, 1999), we utilised a combination of methods which are commonly used in these kinds of surface analyses to validate findings. The combination of prediction algorithms above were used to predict an overall subcellular location for each protein identified in the surfaceome analysis. Additionally, when examining *in silico* predictions of localisation, it must be taken into consideration that members of the

mollicutes do not possess a cell wall, and thus any proteins containing a transmembrane domain have the potential to be directly surface exposed.

Proteins were classified as intracellular, membrane (intracellular), membrane (extracellular), non-classically secreted, extracellular or unknown according to patterns of prediction described in Table 1. 71.7% (114 of 159) of proteins identified were thus classified as membrane-associated, extracellular or non-classically secreted (Table 4.1and Figure 4.1, see also Table S1).

4.4.3. "Cytosolic" proteins

Of the 159 identified surface proteins, 97 (61%) were predicted to be cytoplasmic using PSORTb. When analysed further by SecretomeP, 26 of these were revised to be classified as non-classically secreted proteins (i.e. proteins that are secreted through mechanisms other than the presence of a signal peptide). When the presence of transmembrane domains (predicted by TMPred) is taken into account, it is found that 41 of the predicted cytoplasmic proteins are predicted to be embedded in the membrane. As mycoplasmas lack a cell wall, the majority of proteins that pass through the membrane will be at least in part, exposed to the extracellular environment. With these restrictions, the number of predicted intracellular proteins drops from 98 to 43 (27%) (Figure 4.1). Seven of these have been identified by both methods, and the remainder identified in multiple runs or experiments; therefore even these cannot be completely excluded from the surface topography. Eight annotated cytoplasmic proteins are capable also of eliciting an immune response in the host indicated by immuno-affinity chromatography (Discussed later). Further, 84% (36 of 43) of predicted intracellular proteins are capable of interaction with host molecules such as fibronectin, actin, plasminogen, glycosaminoglycans or host cell surface proteins (Figure 4.3).

Of the predicted intracellular proteins identified in the surfaceome analyses, the prevalence of glycolytic enzymes is of particular note. Seven (of ten known) glycolytic enzymes predicted from the strain J genome sequence were detected at the surface of *M. hyopneumoniae*. The enzymes identified included 6-phosphofructokinase (MHJ_0107, Q4AAM2) (step 3), fructose-bisphosphate aldolase (MHJ_0014, Q4AAT5) (step 4), glyceraldehyde 3-phosphate dehydrogenase (MHJ_0031, Q4AAR8) (step 6), phosphoglycerate kinase (MHJ_0487, Q4A9J8) (step 7), 2,3 bisphosphoglycerate

independent phosphoglycerate mutase (MHJ_0592, Q4A992) (step 8), enolase (MHJ_0242, Q4AA88) (step 9) and pyruvate kinase (MHJ_0122, Q4AAK7) (step 10). Several enzymes with roles in pyruvate metabolism were also identified on the cell surface including all four components of the pyruvate dehydrogenase complex (PdhA-MHJ_0111, Q4AAL8; PdhB-MHJ_0112, Q4AAL7; PdhC-MHJ_0503, Q4A9I2; PdhD-MHJ_0504, Q4A9I1), acetate kinase (MHJ_0505, Q4A9I0), phosphate acetyltransferase (MHJ_0506, Q4A9H9), phosphoenolpyruvate-protein phosphotransferase (MHJ_0469, Q4A9L6), and lactate dehydrogenase (MHJ_0133, P0C0J3). Many of these glycolytic and metabolic "housekeeping" proteins were also found to interact with key host molecules (Table 4.2).

Table 4.2. Interactions of surface-associated glycolytic enzymes and selected metabolic proteins with host molecules.

	Locus	Protein Description	Accession	TMDs	Identified by	PK15	Hep	Actin	Fn	Plg	IP
Glycolytic enzymes	MHJ_0031	Glyceraldehyde 3-phosphate dehydrogenase	Q4AAR8	0	Both	X	X	X	X	X	-
	MHJ_0107	6-phosphofructokinase	Q4AAM2	1	Shaving	-	X	X	-	X	-
	MHJ_0014	Fructose-bisphosphate aldolase	Q4AAT5	1	Shaving	-	X	-	-	-	-
	MHJ_0487	Phosphoglycerate kinase	Q4A9J8	0	Shaving	X	X	X	-	X	-
	MHJ_0122	Pyruvate kinase	Q4AAK7	1	Shaving	-	X	X	-	X	-
	MHJ_0595	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	Q4A992	1	Shaving	-	-	-	-	-	-
	MHJ_0242	Enolase	Q4AA88	1	Both	X	X	X	X	X	-
Other metabolic proteins/enzymes	MHJ_0112	Pyruvate dehydrogenase (PdhB)	Q4AAL7	1	Both	X	X	X	X	X	X
	MHJ_0111	Pyruvate dehydrogenase E1-alpha subunit (PdhA)	Q4AAL8	1	Both	X	X	-	X	X	X
	MHJ_0504	Dihydrolipoamide dehydrogenase (PdhD)	Q4A9I1	2	Shaving	X	X	-	X	X	-
	MHJ_0503	Dihydrolipoamide acetyltransferase (PdhC)	Q4A9I2	2	Both	X	X	-	X	X	X
	MHJ_0506	Phosphate acetyltransferase	Q4A9H9	1	Shaving	X	X	-	-	X	-
	MHJ_0469	Phosphoenolpyruvate-protein phosphotransferase	Q4A9L6	1	Shaving	X	-	-	-	X	-
	MHJ_0133	L-lactate dehydrogenase	P0C0J3	2	Both	X	X	X	X	X	X
	MHJ_0505	Acetate kinase	Q4A9I0	0	Shaving	-	X	X	-	X	-
	MHJ_0110	Adenine phosphoribosyltransferase	Q4AAL9	1	Both	-	X	X	X	X	-
	MHJ_0254	Hypoxanthine phosphoribosyltransferase	Q4AA76	0	Biotin	-	X	X	X	X	X

NB. All proteins listed are predicted to be cytoplasmic by PSORTb, have no signal peptides detectable by SignalP, are not predicted to be non-classically secreted by SecretomeP nor are they predicted to be lipoproteins. The presence of transmembrane domains (TMDs) as indicated determines the overall predicted localisation to be membrane (intracellular), and the lack of TMDs determines intracellular overall localisation prediction. X indicates proteins were detected in biotinylated PK15 surface protein (PK15), heparin (Hep), actin, fibronectin (Fn) or plasminogen (Plg) affinity chromatography, or immunoprecipitation using convalescent sera (IP), coupled with GeLC-MS/MS.

30 ribosomal proteins were also identified to be surface exposed. Two of the identified ribosomal proteins possess TMDs, one of which (RpsH) is classed as membrane (intracellular), the other (RpsM) also is predicted to be non-classically secreted and is classed membrane (extracellular). RpsT is predicted by SignalP to possess a conventional signal peptide and is predicted to be extracellular, and RpsF is classed as unknown due to a PSORTb prediction of membrane localisation, despite lacking detectable TMDs. 11 ribosomal proteins were classified as intracellular with the remaining 15 classed as non-classically secreted. Despite their predicted localisations, all but one (30S ribosomal protein S12) of the ribosomal proteins identified at the surface were also identified to interact with host components (epithelial cell surface proteins, heparin, actin, plasminogen, fibronectin or eliciting an immune response), and of these, 26 were found to interact with more than one of these components, lending further credence to the validity of their surface localisation.

4.4.4. Interactions with host components

Surface proteins were investigated for the ability to bind host ECM components such as fibronectin, actin, plasminogen and the glycosaminoglycan-mimic heparin, as well as surface proteins expressed on host epithelial cells (Figure 4.3, see also Table S4.1). A total of 130 of the 159 surface proteins bound at least one of the tested host components or elicited an immune response. 40 of the surface proteins identified were found to be capable of binding fibronectin, 74 showed interaction with the cytoskeletal protein actin, 97 surface proteins were retained on a column using plasminogen as bait, and 87 surface proteins were found to interact with heparin, indicating a putative ability to bind glycosaminoglycans. Porcine kidney epithelial-like cells (PK15 cells) were used as a model for host cell binding. Surface proteins of PK15 cells were biotinylated and bound to an avidin-agarose column. Non-biotinylated *M. hyopneumoniae* whole cell extracts were passed through the column to identify interacting proteins and 52 of the identified surface proteins were found to bind. 15 proteins were identified to bind to all the host components tested, 11 of which also elicited an immune response (Table 4.3).

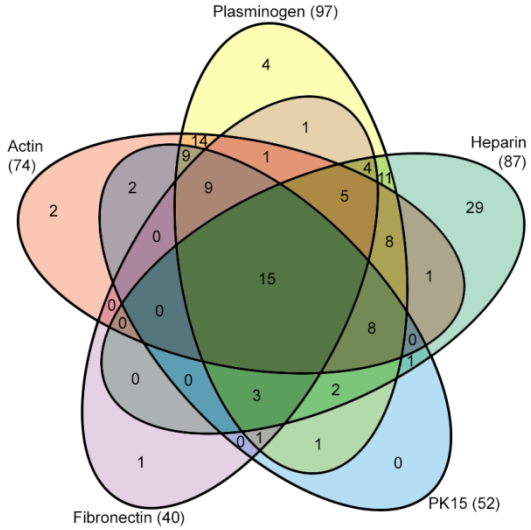


Figure 4.3. Surface proteins identified from affinity chromatography using key host molecules as bait.

Indicated are the numbers of surface proteins identified by affinity chromatography using biotinylated surface proteins of porcine kidney epithelial-like cells (PK15), heparin, actin, plasminogen or fibronectin as bait. A total of 130 surface proteins were identified to bind at least one key host component.

Table 4.3. Surface proteins of *M. hyopneumoniae* that bind heparin, fibronectin, plasminogen and PK15 surface proteins.

Locus	Protein names	Accession	Class
MHJ_0194	Uncharacterised protein (P97/P123)	Q4AAD6	Adhesin
MHJ_0195	Uncharacterised protein (P102)	Q4AAD5	Adhesin
MHJ_0494	Putative p76 membrane protein	Q4A9J1	Adhesin
MHJ_0662	Uncharacterised protein	Q4A926	Adhesin
MHJ_0663	Putative adhesin like-protein P146	Q4A925	Adhesin
MHJ_0511	46 kDa surface antigen (p46)	P0C0J8	Membrane Protein
MHJ_0656	Putative prolipoprotein p65	Q4A932	Membrane Protein
MHJ_0031	Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.-)	Q4AAR8	Glycolytic Enzyme
MHJ_0242	Enolase (EC 4.2.1.11)	Q4AA88	Glycolytic Enzyme
MHJ_0063	Chaperone protein DnaK (HSP70)	Q4AAR4	Metabolic
MHJ_0078	NADH oxidase (EC 1.6.99.3)	Q4AAP9	Metabolic
MHJ_0112	Pyruvate dehydrogenase (PdhB) (EC 1.2.4.1)	Q4AAL7	Metabolic
MHJ_0133	L-lactate dehydrogenase (Immunogenic protein p36)	P0C0J3	Metabolic
MHJ_0524	Elongation factor Tu (EF-Tu)	Q4A9G1	Translation
MHJ_0172	50S ribosomal protein L15*	Q4AAF8	Ribosomal Protein

Shaded indicates proteins that were not also found to be immunogenic by immunoprecipitation pulldown. *All proteins were identified by both biotinylation and shaving, except 50S ribosomal protein L15, which was only identified by biotinylation.

Surface proteins were also examined for immunogenicity through immunoprecipitation and immunoblotting using pooled positive convalescent sera (Figure 4.4 and Figure 4.5A and B). Reactivity with convalescent sera further acts as confirmation of *in vivo* expression. Table 4.4 lists the 36 immunogenic proteins identified by immuno-affinity chromatography.

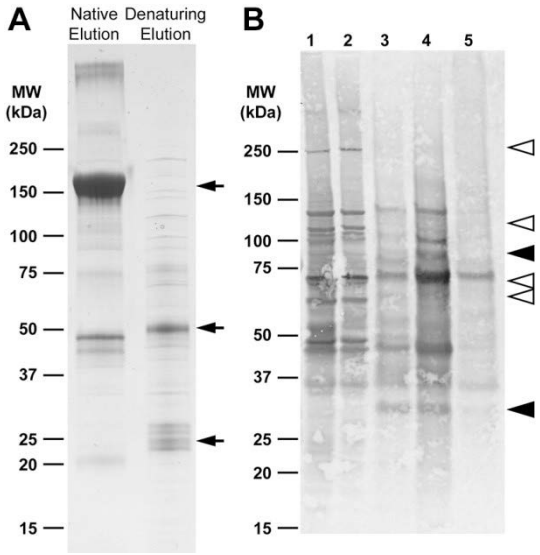


Figure 4.4. Immuno-affinity chromatography and convalescent sera blots showing immuno-reactive proteins of *M. hyopneumoniae*.

Panel A shows a 1D gel of sequential native and denaturing elutions following immuno-affinity chromatography using pooled positive convalescent sera. Arrows indicate intact immunoglobulin at ~150 kDa in the native elution and immunoglobulin fragments at ~50 and 25 kDa in denaturing the elution. Co-eluting protein A can also be seen at 42 kDa in the native elutions. Panel B shows a blot of immunoreactive proteins from unlabelled (1) and biotinylated (2) whole cell extracts and avidin-purified biotinylated TX114 phase extractions, aqueous (3), detergent (4) and insoluble fractions (5). Proteins not identified in purified surface extractions indicated by open arrows, proteins showing increased reactivity in the enriched fractions are indicated by closed arrows.

Table 4.4. Immunogenic surface proteins of *M. hyopneumoniae.*

Class	Locus	Protein Description	Accession	Lp	MW
Adhesins	MHJ_0194	Putative uncharacterised protein (P97)	Q4AAD6	N	126*
	MHJ_0195	Putative uncharacterised protein (P102)	Q4AAD5	N	102*
	MHJ_0368	Putative Lppt protein (Mhp384)	Q4A9W5	N	109*
	MHJ_0369	Putative uncharacterised protein (Mhp385)	Q4A9W4	N	114*
	MHJ_0493	Putative P216 surface protein (P216)	Q4A9J2	N	216*
	MHJ_0494	Putative p76 membrane protein (P159)	Q4A9J1	N	161*
	MHJ_0662	Putative uncharacterised protein (Mhp683)	Q4A926	N	135*
	MHJ_0663	Putative adhesin like-protein P146 (P146)	Q4A925	N	147*
Membrane Proteins	MHJ_0212	Putative uncharacterised protein	Q4AAB8	N	236*
	MHJ_0511	46 kDa surface antigen (p46)	P0C0J8	Y	45*
	MHJ_0606	ABC transporter xylose-binding lipoprotein	Q4A981	N	50*
	MHJ_0656	Putative prolipoprotein p65	Q4A932	Y	71*
Metabolic Proteins	MHJ_0063	Chaperone protein DnaK (HSP70)	Q4AAR4	N	66*
	MHJ_0111	Pyruvate dehydrogenase E1-alpha subunit (PdhA)	Q4AAL8	N	42*
	MHJ_0112	Pyruvate dehydrogenase (PdhB)	Q4AAL7	N	37*
	MHJ_0133	L-lactate dehydrogenase (Immunogenic p36)	P0C0J3	N	34*
	MHJ_0254	Hypoxanthine phosphoribosyltransferase	Q4AA76	N	20
	MHJ_0503	Dihydrolipoamide acetyltransferase (PdhC)	Q4A9I2	N	33
	MHJ_0617	DNA-directed RNA polymerase subunit beta'	Q4A970	N	159
Ribosomal Proteins	MHJ_0072	30S ribosomal protein S7	Q4AAQ5	N	18
	MHJ_0127	50S ribosomal protein L21	Q4AAK2	N	11
	MHJ_0165	30S ribosomal protein S11	Q4AAG5	N	15
	MHJ_0175	50S ribosomal protein L6	Q4AAF5	N	20
	MHJ_0178	50S ribosomal protein L5	Q4AAF2	N	20
	MHJ_0179	50S ribosomal protein L24	Q4AAF1	N	12
	MHJ_0189	50S ribosomal protein L4	Q4AAE1	N	24
	MHJ_0456	50S ribosomal protein L1	Q4A9M9	N	26
	MHJ_0577	30S ribosomal protein S4	Q4A9B0	N	24
Proteases	MHJ_0202	ATP-dependent zinc metalloprotease FtsH	Q4AAC8	N	78*
	MHJ_0525	Heat shock ATP-dependent protease (Lon)	Q4A9G0	N	99
Others	MHJ_0227	Periplasmic sugar-binding protein	Q4AAA3	N	44*
	MHJ_0592	Putative ATP-binding protein	Q4A995	N	124
	MHJ_0524	Elongation factor Tu (EF-Tu)[1]	Q4A9G1	N	44*
	MHJ_0213	Putative lipoprotein[2]	Q4AAB7	Y	102
	MHJ_0326	Putative uncharacterised protein[3]	Q4AA06	N	25*
	MHJ_0134	Uncharacterised protein[3]	Q4AAJ6	Y	107

MW: Calculated intact molecular weight in kDa. NB: Proteins marked with * and all adhesin proteins are known to be post-translationally cleaved into smaller, functional proteins (Chapter 3). Others class: [1]Translation, [2]Transporter, [3] Unknown. Lp: Predicted Lipoprotein.

Figure 4.3B shows immunogenic proteins from whole cell extracts and purified biotinylated phase fractionations. Some immunogenic proteins were not identified in purified surface extractions (open arrows), however many were enriched in these fractions, showing increased reactivity (closed arrows). The lower MW proteins (primarily ribosomal proteins) in the biotinylated purified samples are not easily detectable by blotting with convalescent pig sera; however in unfractionated samples (ie. not purified by avidin-affinity) these proteins are more readily detected. In addition these same proteins are repeatedly identified by mass spectrometry in analyses of purified surface proteins but are reproducibly not the dominant species. This may indicate that only a small proportion of these proteins are displayed on the cell surface, with the majority remaining within the cell. This may also suggest that these proteins may perform multiple functions occupying different cellular compartments, as moonlighting proteins.

2D immunoblots performed on pH range 4-7 and 6-11 gels of *M. hyopneumoniae* whole cell lystaes provide a visual overview of the immunoproteome (Figure 4.5A and B). It can be seen that many proteins form "spot trains" across a p*I* range, and particularly visible on the pH 6-11 blots are previously described "cloud regions" composed of minor cleavage fragments of adhesin proteins (Chapter 3). Ligand blots (Figure 4.5C-F) were also performed using fibronectin and plasminogen in order to validate some of the proteins identified by affinity chromatography methods and elucidate post-translational modifications such as proteolytic processing.

Only 10 fibronectin binding proteins identified by 2D ligand blots could be matched to spots on 2D gels as indicated in Figure 4.5. 25 of the 97 identified plasminogen binding proteins were validated by ligand blotting (Figure 4.5).

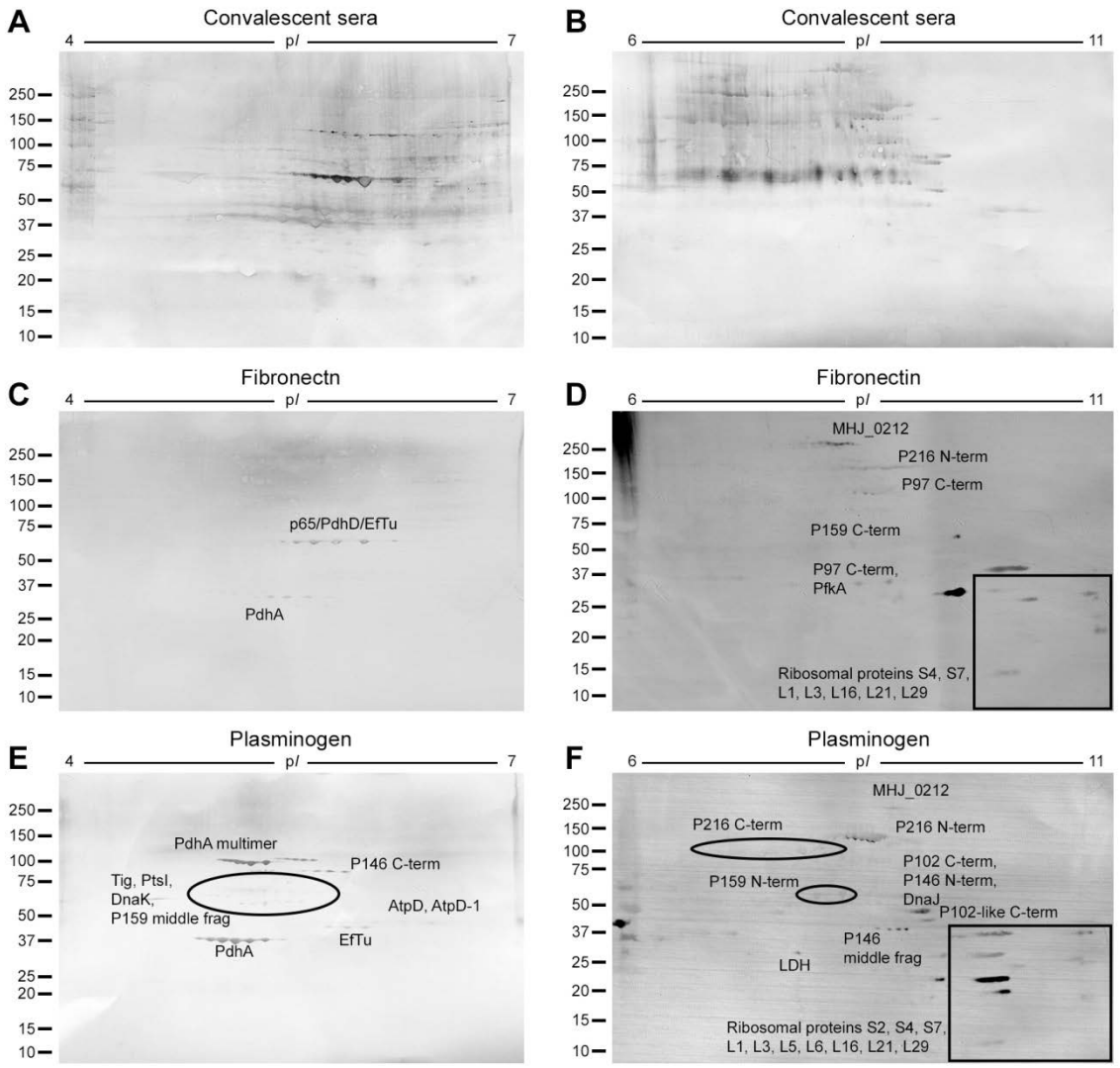


Figure 4.5. Immuno- and ligand-blots of pH 4-7 and 6-11 2D gels of *M. hyopneumoniae* whole cell lysates.

Panels A, C and E show pH 4-7 2D blots, and panels B, D and F show pH 6-11 2D blots. Panels A and B were probed with convalescent sera showing immunogenic proteins, panels C and D were probed with (biotinylated) fibronectin and panels E and F were probed with plasminogen. Where spots could be identified, reactive proteins are indicated, ribosomal proteins identified at low masses and basic p*I* are boxed. Reference gels and protein identifications have been described previously (Chapter 3).

26 of the 159 identified surface proteins were not found to interact with heparin, fibronectin, actin, plasminogen or PK15 surface proteins in column chromatography, nor were they determined to be immuno-reactive. Of these, only one (Elongation factor G) was identified by both methods, although 19 of the 26 proteins are predicted to be

surface exposed or membrane bound by combined bioinformatics predictions. Interestingly, all four proteins that were not detected in previous global proteome analyses were found to be "non-binding" proteins. A high number of lipoproteins (6 of 26) and proteins with putative transmembrane domains (15 of 26) amongst these "non-binding" proteins may also suggest that limits of solubility prevent these proteins from being readily detected in affinity-capture methods. Alternatively, these proteins may carry out other, as of yet undescribed functions at the cell surface.

4.4.5. P97 and P102 paralogs and proteases

All members of the P97 and P102 paralog families encoded by *M. hyopneumoniae* strain J, and Mhp494 (P159), a molecule with sequence similarity to P102, were identified on the surface using one or both approaches. This is consistent with previous finding that members of these two families are prominent proteins present on the surface of *M. hyopneumoniae.* Endoproteolytic cleavage is a hallmark of the P97 and P102 paralog families. In most instances, each of these proteins undergoes one or two major cleavage events. More recent evidence suggests that cleavage efficiency varies at some sites, and major cleavage fragments are targets of further proteolytic cleavage by different proteases (Raymond *et al.*, 2014; Raymond *et al.*, 2013; Tacchi *et al.*, 2014). These processes profoundly alter how proteins are expressed on the cell surface and provide a mechanism by which *M. hyopneumoniae* regulates surface topography (Bogema *et al.*, 2012). Proteolytic cleavage of the P97 and P102 paralogous families of adhesins has been previously discussed in detail by several protein-specific publications and will not be recapped here (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Wilton *et al.*, 2009). Additionally, an overview of proteins that undergo cleavage can be found in Chapter 3.

The *M. hyopneumoniae* genome encodes 17 annotated proteases, 11 of which have been previously identified to be expressed under normal culture conditions (Chapter 3). Here we have identified 6 annotated proteases at the cell surface (Table 4.5). Three proteases were annotated to have endoprotease activity while the remaining three were annotated

aminopeptidases. These endoproteases are likely to carry out the major proteolytic actions that give rise to adhesin fragments, as well as processing other proteins.

All identified proteases were also found to interact with heparin, while none were retained by actin column affinity chromatography. MHJ_0125 has been previously described in detail to be a surface-exposed moonlighting protein of *M. hyopneumoniae* (Robinson *et al.*, 2013).

Table 4.5. Identified surface proteases

Locus	Protein Description	Accession	Activity	PK15	Fn	Plg	IP
MHJ_0125	Putative aminopeptidase	Q4AAK4	Aminopeptidase	-	-	-*	-
MHJ_0461	Leucyl aminopeptidase	Q4A9M4	Aminopeptidase	-	-	X	-
MHJ_0659	XAA-PRO aminopeptidase	Q4A929	Aminopeptidase	-	-	-	-
MHJ_0202	ATP-dependent zinc metalloprotease FtsH	Q4AAC8	Metalloendoprotease	-	X	X	X
MHJ_0522	Oligoendopeptidase F	Q4A9G3	Metalloendoprotease	X	-	-	-
MHJ_0525	Lon protease	Q4A9G0	Serine endoprotease	X	-	X	X

X indicates proteins were detected in biotinylated PK15 surface protein (PK15), fibronectin (Fn) or plasminogen (Plg) affinity chromatography, or immunoprecipitation using convalescent sera (IP), coupled with GeLC-MS/MS. All proteases interacted with heparin, whilst none were retained by actin affinity chromatography. *MHJ_0125 has been previously described to bind plasminogen (Robinson *et al.*, 2013).

4.5. Discussion

The P97 and P102 paralogous families of adhesins in *M. hyopneumoniae* are among the best characterised proteins of this bacterium, and the most prominent class of surface proteins (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Jenkins *et al.*, 2006; Raymond *et al.*, 2014; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Wilton *et al.*, 2009; Wilton *et al*., 1998). It is unsurprising then, that they are well-represented in the analysis of surface proteins presented here, with all members expressed at the cell surface. These proteins have been previously described to play important roles in initial adherence events leading to colonisation of the host. Endoproteolytic fragments derived from the two families display multifunctional capabilities on the surface of *M. hyopneumoniae* where they bind directly to cilia, sulphated glycosaminoglycans, fibronectin and plasminogen (Adams *et al.*, 2005; Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Jenkins *et al.*, 2006; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Zhang *et al.*, 1995).

M. hyopneumoniae has a reduced coding capacity and therefore the ability to modify proteins post-translationally may be a conservative evolutionary step to maintain function and diversity, particularly for surface-exposed proteins. Proteolytic processing has been described previously for 11 members of the P97/P102 paralog family (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Wilton *et al.*, 2009). At this stage it is hypothesised that multiple proteases are responsible for cleavage, and the majority of cleavage events are carried out at the cell surface. The adhesin molecules possess a stretch of hydrophobic amino acids at the N-terminus that enable the protein to be inserted into the membrane and unfold to leave the majority of the length exposed to the extracellular environment (Djordjevic *et al.*, 2004). However, the known cleavage fragments of the adhesins do not possess any known mechanism to become surface exposed or remain attached to the surface and this is an area of interest for future studies. Therefore proteolytic processing is likely to be

carried out at the cell surface as the adhesins 'unloop' through the membrane. Consistent with this, we identified six proteases present at the cell surface, three with endoprotease activity that present as potentially responsible for carrying out dominant cleavage events in the adhesin proteins. These are currently the targets of further investigations.

In a recent study, the surfaceome of *M. hyopneumoniae* strain 7448 was found to consist of 59 proteins, when analysed by biotin labelling of intact cells (Reolon *et al.*, 2014). A BLASTP search of the identified proteins allowed identification of the homologous proteins in strain J, so that the surfaceomes of the two strains could be compared.
47 homologous proteins were identified in both the surfaceome of *M. hyopneumoniae* strain J and strain 7448. These proteins were perhaps unsurprisingly, largely made up of the adhesin proteins, which have been shown to be conserved between strains (Hsu and Minion, 1998b; Zhang *et al.*, 1995), membrane proteins and lipoproteins, but also moonlighting proteins described here and in other species to be involved in host-pathogen interactions. Lactate dehydrogenase, pyruvate dehydrogenase complex components, elongation factor Tu, chaperone protein DnaK and glyceraldehyde phosphate dehydrogenase were amongst the proteins identified to be surface exposed in both strains of *M. hyopneumoniae*, suggesting their moonlighting activities may also be conserved between strains. 16 proteins identified in the 7448 surfaceome were not identified in the analyses presented here (by homology). Of these, half (8) of the J strain homologs were not identified in strain J global analyses (Chapter 3), suggesting there may be strain-specific differences in expression levels.

A striking feature of this study has been the high number of predicted intracellular (27%) and non-classically secreted proteins (18%); that is, proteins with known or annotated functions in the cytosol, lacking any predicted transmembrane domains or traditional mechanisms for transport to the cell surface. Interpretation of results from bioinformatic analyses can be problematic as mycoplasmas are an unusual class of organisms, descending from Gram-positive firmicutes but lacking a cell wall and complete metabolic pathways (Dybvig and Voelker, 1996; Edman *et al.*, 1999). With regards to cell structure, the Mollicutes are more akin to mammalian cells, lacking a cell wall and being bounded by only a single plasma membrane, however export

mechanisms are currently assumed to be more closely related to Gram positive organisms through evolution (Dybvig and Voelker, 1996). It must be recognised that localisation prediction algorithms are not trained on, and therefore not optimised for atypical organisms such as the mycoplasmas (Edman *et al.*, 1999). Signal peptides of mollicutes have been found to be significantly shorter than those of Gram positive bacteria, while they are relatively longer than those of *E. coli*. Signal peptides of mollicutes have also been found to differ from other bacteria in their N-terminal charge, peptide length and periodicity of side chain hydrophobicity (Edman *et al.*, 1999). Signal peptides for lipoproteins were also found to be longer than for any other bacteria, which may be an adaptation to the thicker, cholesterol-rich membrane (Edman *et al.*, 1999). These may be important factors in the secretion systems for mycoplasmas.

Even in recognising shortcomings of *in silico* predictions of protein subcellular localisations, a large number of proteins with well-characterised intracellular functions (or homologs in other species) were still identified to be surface exposed. The identification of annotated cytosolic proteins in surface proteome experiments has previously been a contentious issue, often attributed to cell lysis or regarded as artefact and dismissed, however cytoplasmic proteins in some cases have been described to have multiple functions dependent on cellular location and are in such cases referred to as "moonlighting" proteins (Solis and Cordwell, 2011). In some cases, such as for the large proportion of uncharacterised proteins in *M. hyopneumoniae*, functions have not yet been ascribed, but the genuine presence of these proteins at the cell surface is supported by recognition by immune sera or interaction with host molecules.

In a review by Dreisbach *et al.*, the results of five independent studies of the surfaceome of *Staphylococcus aureus* were compared (Dreisbach *et al*., 2011). The studies covered several different strains and utilised methods of biotinylation (Hempel *et al.*, 2010), surface shaving with trypsin (Dreisbach *et al.*, 2010; Solis *et al.*, 2010; Ventura *et al*., 2010) and a subtractive proteome analysis or "SUPRA" (Glowalla *et al*., 2009). It was concluded that surface labelling using membrane-impermeable biotin identified the most proteins, however this was likely to be confounded by the ability of the biotin to label proteins buried within the cell wall and not truly surface exposed. This is not a problem for the cell wall-less mycoplasma. In addition, the lack of a cell wall means,

proportionately, a larger repertoire of proteins will be surface exposed compared to walled bacteria and proteins with transmembrane domains will likely be in direct contact with the external environment. In the comparison of the Staphylococcus surfaceomes between 40 and 60% of proteins identified were predicted to be cytosolic, or had known functions in the cytosol; this is a common feature of many surfaceome analyses (Dreisbach *et al.*, 2011). Surface proteome analyses of Gram-positive bacteria such as *Bacillus subtilis* (Tjalsma *et al*., 2004; Tjalsma *et al.*, 2008), and *Streptococcus pyogenes* (Severin *et al*., 2007), as well as Gram negative bacteria such as uropathogenic *E. coli* (Wurpel *et al*., 2014) have shown similar trends, and our results suggest this phenomenon extends also to Mycoplasmas. More recently, it has been reported that *S. aureus* biofilm extracellular matrix is largely composed of cytoplasmic proteins and that their binding to the cell surface is pH dependent (Foulston *et al*., 2014).

To begin to ascribe functions to surface proteins of *M. hyopneumoniae*, we screened surface proteins for the ability to bind key host molecules: heparin, fibronectin, actin, plasminogen and surface proteins of model host epithelial cells. By investigating function, we can obtain a more accurate representation of the interplay between host and bacterial systems, particularly those which may assist in colonisation, immune evasion or invasion of the host. A better understanding will also enable us to exploit these interactions in the process of rational vaccine or drug design.

Heparin was chosen as a structural mimic for highly sulphated glycosaminoglycans that line the cilia in the respiratory tract and are opportune targets for initial adherence events. Many proteins of *M. hyopneumoniae* have previously been described to bind heparin, including members of the adhesin paralogous families (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Hsu *et al.*, 1997; Jenkins *et al.*, 2006; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Wilton *et al.*, 2009; Zhang *et al.*, 1994), a leucine aminopeptidase (MHJ_0461) (Jarocki *et al.*, 2015), and a glutamyl aminopeptidase (MHJ_0125) (Robinson *et al.*, 2013). These proteins have also been described previously to bind plasminogen, the precursor to the serine protease, plasmin which is important for the degradation of tissue barriers,

and hence cell migration. Bacterial pathogens intervene with this system in a number of ways, usually to facilitate in host extracellular matrix degradation for the purpose of cellular invasion or recycling of amino acids through activation of latent matrix metalloproteases (Lahteenmaki *et al.*, 2005). Both glutamyl and leucine aminopeptidases of *M. hyopneumoniae* have been described to degrade ECM components and these have been hypothesised to play roles in the destruction of mucocilliary function during adhesion to porcine respiratory cilia and have been implicated in the acquisition of host amino acids through the systematic degredation of N-terminal amino acids (Jarocki *et al*., 2014; Robinson *et al.*, 2013). Interestingly, the glutamyl aminopeptidase was previously identified to bind plasminogen in binding assays using recombinant proteins (Robinson *et al.*, 2013), however it was not identified in the global affinity capture protocol used here. This may be attributed to abundance issues and protein dynamic range in the sample. It should be noted that traces of actin was detected in the commercially-obtained plasminogen used in affinity chromatography, thus proteins that are listed as binding actin and plasminogen may actually preferentially interact with actin. The ability to bind actin, a host cytoskeletal protein, is another common feature of bacterial invasion and cell-to-cell spread (Gruber and Sperandio, 2014). As *M. hyopneumoniae* has been isolated from the liver and spleen of contact-infected pigs, suggesting rapid colonisation occurs and invasion is likely to be an underexplored feature of infection (Marois *et al.*, 2007). This internal dissemination is likely to result in reinfection, following treatment, even if external sources of infection are removed (Le Carrou *et al.*, 2006), which is a factor that must be taken into consideration in the design of vaccines and therapeutics. Another key host protein exploited for adhesion and invasion by bacterial pathogens is fibronectin, a glycoprotein component of the host extracellular matrix. Fibronectin binding proteins promote adhesion to the surface of host cells and the subsequent internalisation (Foster *et al*., 2014). Finally, we screened for direct host-pathogen protein-protein interactions, using biotinylated surface proteins from a model porcine epithelial-like cell line (PK15), in an indirect avidin-affinity chromatography purification.

The ability to bind each of these components and/or elicit an immune response *in vivo* necessitates that these proteins are exposed to the host; however the mechanism by which they become exposed remains elusive. One hypothesis is that proteins may be

secreted by undiscovered secretion mechanisms or pathways (Boel *et al*., 2005; Foulston *et al.*, 2014; Yang *et al*., 2011), which is possible, given the relatively understudied nature of the mollicutes due to previous difficulties in genetic manipulation. Another hypothesis is cell lysis, either through normal turnover or through altruistic lysis or fratricide in bacterial communities, releases these proteins into the extracellular milleu, whereby they bind back to the Mycoplasma cell surface and/or affect their (moonlighting) functions extracellularly (Foulston *et al.*, 2014). In a genome-reduced organism that has previously been shown to recycle amino acids (Jarocki *et al.*, 2015; Robinson *et al.*, 2013) and post-translationally modify/cleave large proteins to obtain multifunctional fragments (Raymond *et al.*, 2014; Raymond *et al.*, 2013; Tacchi *et al.*, 2014) (Chapter 3), this would be yet another mechanism to maximise the coding capacity.

The concept of moonlighting proteins is not novel. They are described as proteins capable of performing two or more distinct biological functions within a single polypeptide chain, excluding proteins that play the same role in different locations or splice variants (Jeffery, 1999). Here, we specifically discuss topographical/geographical moonlighters; i.e. proteins with a different function depending on their cellular localisation. Moonlighting proteins can be found in a range of organisms, from pathogenic microbes (Jeffery, 1999), to eukaryotic pathogens such as pathogenic fungi as well as parasitic protozoa and multicellular parasites (Karkowska-Kuleta and Kozik, 2014) as well as mammalian cells (Jeffery, 2003; Jung *et al*., 2014; Petit *et al*., 2014). Glycolytic enzymes and chaperones, which are highly conserved, are amongst the most commonly identified examples of moonlighting proteins in all species.

We identified 7 annotated glycolytic enzymes of *M. hyopneumoniae* strain J to be surface exposed, 6 of which were also found to bind various host molecules (Table 2). Plasminogen-binding was a prominent feature of the moonlighting glycolytic enzymes identified, and is common throughout many species (Henderson and Martin, 2011). Examples of plasminogen-binding moonlighting proteins can be found in enolase of *Borrelia burgdorferi* (Toledo *et al*., 2012), *Bacillus anthracis* (Agarwal *et al*., 2008) and *Streptococcus pneumoniae* (Bergmann *et al*., 2001), as well as glyceraldehyde phosphate dehydrogenase of *Streptococcus pneumoniae* (Bergmann *et al*., 2004) and

triosephosphate isomerase of *Staphylococcus aureus* (Furuya and Ikeda, 2011). Phosphoglycerate kinase binds plasminogen in *Streptococcus pneumoniae* (Fulde *et al*., 2014) and in Group B Streptococcus, it also has the capacity to bind actin (Boone *et al*., 2011). Other well-characterised moonlighting functions of metabolic proteins include fibronectin binding and host cell adhesion. Notably, in the porcine pathogen *Streptococcus suis,* enolase, fructose-bisphosphate aldolase, phosphoglycerate mutase, pyruvate kinase and pyruvate dehydrogenase E1 component alpha subunit have been identified to bind fibronectin (Zhang *et al*., 2014). Fructose-1,6-bisphosphate aldolase in *Neisseria meningitidis* plays a role in host cell adhesion (Tunio *et al*., 2010).

In *Mycoplasma* species, moonlighting proteins have previously been identified in the human pathogen *M. pneumoniae.* The proteins elongation factor Tu and pyruvate dehydrogenase E1-beta subunit (PdhB) have been described to localise to the surface and play a role in binding host molecules such as fibronectin (Balasubramanian *et al*., 2008; Dallo *et al.*, 2002). This is consistent with our findings that both Pyruvate dehydrogenase (PdhB) and Elongation factor Tu of *M. hyopneumoniae* bind fibronectin, plasminogen, heparin and host epithelial cell surface proteins. We also report EfTu to be immunogenic. Additionally, we have previously reported that both PdhB and EfTu in *M. hyopneumoniae* undergo post-translational cleavage events and a detailed analysis of motifs that enable fragments of EfTu to bind to various host components can be found in Chapter 3. This evidence of cleavage is relevant to surface-exposed multifunctional proteins in two ways. Firstly we have hypothesised that proteolytic processing of the P97/P102 adhesin proteins occurs at the cells surface at specific motifs and regions within the proteins (Bogema *et al.*, 2011; Deutscher *et al.*, 2012; Raymond *et al.*, 2013), requiring proteins to be translocated across the cell membrane before they can be cleaved. Secondly, these proteins have well characterised structural properties that enable them to carry out their primary functions. Cleavage would inevitably alter the structure, which in turn would affect the function. If this post-translational processing is necessary for these proteins to affect their secondary functions, they can no longer be classed as true moonlighting proteins (Jeffery, 1999). However, the current evidence in the literature suggests homologous proteins in other species can affect their alternate functions in the absence of post-translational modifications.

The multi-functionality of ribosomal proteins has also previously been suggested by work performed on *Mycoplasma pneumoniae*. Size exclusion chromatography found some high abundance ribosomal proteins not associated with the ribosome (Maier *et al*., 2011). In addition, genome-wide screening or protein-protein interactions using tandem affinity purification coupled to mass spectrometry, indicated that several ribosomal proteins of *M. pneumoniae* are associated with different protein complexes, suggesting the potential to effect different functions in monomeric states, in complexes with other proteins and as components of the intact ribosome (Kuhner *et al.*, 2009; Maier *et al.*, 2011). Our analyses identified 30 ribosomal proteins as putative surface proteins, and while further investigation is needed to determine if these proteins are present at the cell surface as monomeric subunits or in a complex, several have shown preliminary evidence for binding to host plasminogen, fibronectin, actin, heparin and/or epithelial cell surface proteins, and nine elicit an immune response.

Given the reduced coding capacity of the mycoplasma genomes, it is not unfeasible that multifunctional proteins could be of higher abundance/importance in these bacteria in order to carry out functions that make these organisms successful host colonisers and pathogens.

4.6. Conclusion

What distinguishes this study from other surfaceome analyses is that we have taken a conceptually unbiased approach, coupled with high-throughput screening techniques to begin to assign biological function to surface-exposed proteins. The data presented demonstrates the lack of congruency between bioinformatics predictions and experimentally obtained data and serves to highlight how easily experimental data that does not conform to the model system may be disregarded, without further investigation into potential biological functions. We identified a number of potential moonlighting proteins, some previously described and some novel, which were present on the cell surface and showed evidence of binding to heparin, actin, fibronectin or plasminogen by affinity chromatography, and/or elicited an immune response. The immunoproteome begins to describe the proteins that are “seen” by the host immune system, capable of eliciting an immune response and most importantly, validating the expression and localisation of these proteins *in vivo*. Further studies are required to confirm the binding

ability and determine kinetics of individual protein-ligand or protein-protein interactions. By improving our understanding the surface proteome, we can better understand the nature of host-pathogen interactions and contribute to rational design of novel vaccines and therapeutics to combat infection.

Chapter 5. Cilium adhesin P216 (MHJ_0493) is a target of ectodomain shedding and aminopeptidase activity on the surface of *Mycoplasma hyopneumoniae*

The following chapter has been subjected to peer-review and published in the Journal of Proteome Research, May 2014.

Author contributions:

Tacchi, J. L. Performed proteomic experiments, analysed data and wrote the paper.
Raymond, B. B. Performed biotinylation experiments and PK15 interaction analyses.
Jarocki, V. M. Performed bioinformatics heparin motif analyses.
Berry, I. J. Performed N-terminal dimethyl labelling experiments.
Padula, M. P. Carried out mass spectrometry and assisted data analysis and validation.
Djordjevic, S. P. Conceived the study and assisted in the writing of the paper.

5.1. Abstract

MHJ_0493 (P216) is a highly expressed cilium adhesin in *Mycoplasma hyopneumoniae.* P216 undergoes cleavage at position 1074 in the S/T-X-F↓-X-D/E-like motif ^{1072}T-N-F↓Q-E^{1076} generating N-terminal and C-terminal fragments of 120 kDa (P120) and 85 kDa (P85) on the surface of *M. hyopneumoniae*. Here we show that several S/T-X-F↓X-D/E-like motifs exist in P216 but only ^{1072}T-N-F↓Q-E^{1076} and ^{1344}I-T-F↓A-D-Y^{1349} were determined to be *bona fide* processing sites by identifying semi-tryptic peptides consistent with cleavage at the phenylalanine residue. The location of S/T-X-F↓-X-D/E-like motifs within or abutting regions of protein disorder greater than 40 consecutive amino acids is consistent with our hypothesis that site access influences the cleavage efficiency. Approximately twenty cleavage fragments of P216 were identified on the surface of *M. hyopneumoniae* by LC-MS/MS analysis of biotinylated proteins and 2D SDS-PAGE. LC-MS/MS analysis of semi-tryptic peptides within P216 identified novel cleavage sites. Moreover, detection of a series of overlapping semi-tryptic peptides that differed by the loss a single amino acid at their N-terminus is consistent with aminopeptidase activity on the surface of *M. hyopneumoniae*. P120 and P85 and their cleavage fragments bind heparin and cell surface proteins derived from porcine epithelial-like cells indicating that P216 cleavage fragments retain the ability to bind glycosaminoglycans.

5.2. Introduction

Advances in genome sequencing technologies in the past decade have led to an explosion in the number of microbial genomes in public databases (NCBI Microbial genomes resources http://www.ncbi.nlm.nih.gov). The vast majority of microbial genome sequences are computationally annotated without manual curation, with the assumption that gene predictions accurately represent the coding capacity of the genome. In reality, annotation pipelines are unable to accurately predict the structural nature and location of posttranslational chemical modifications and are unable to predict most proteolytic processing events (Ansong *et al*., 2011). Protein-centric proteomics approaches provide an avenue to i) interrogate the accuracy of computationally predicted protein-coding sequences, ii) identify novel ORFs, iii) confirm start sites, iv) identify the proportion of proteins that retain or lose N-terminal methionine residues and v) mine peptide spectral libraries and characterise post-translational modifications. We have successfully used 2D SDS-PAGE and protein-centric approaches to unravel the complex proteolytic events that shape how surface proteins are processed into mature effector molecules in *Mycoplasma hyopneumoniae* (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 1994; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Wilton *et al.*, 2009), *Mycoplasma gallisepticum* (Szczepanek *et al*., 2010), *Streptococcus pyogenes* (Cole *et al*., 2005) and *Campylobacter jejuni* (Scott *et al*., 2010).

M. hyopneumoniae is a major respiratory pathogen that inflicts severe economic losses to swine production globally. As a consequence of genome reduction, *M. hyopneumoniae* has lost the genetic capacity to synthesize amino acids, nucleic acids and essential lipids for membrane construction and is reliant on successful colonisation of the porcine ciliated epithelium for survival. Colonisation of respiratory ciliated epithelium relies on the expression of two functionally-redundant adhesin families that are paralogs of P97 and P102. Members of the P97 and P102 paralog families are endoproteolytically processed, generating a complex combinatorial library of functionally redundant endoproteolytic cleavage fragments that bind glycosaminoglycans, fibronectin, and plasminogen (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et*

al., 2004; Jenkins *et al.*, 2006; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Wilton *et al.*, 2009). P97- and P102-related adhesins are thought to be modified by the action of several endoproteases because cleavage occurs at structurally unrelated motifs with sequences S/T-X-F↓X-D/E, T-N-T↓N-T-N and L-X-V↓X-V/A-X (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004). Recent evidence suggests that *M. hyopneumoniae* also displays surface accessible trypsin-like activity that plays a role in endoproteolytic processing of surface adhesins (Raymond *et al.*, 2013). Cleavage is not confined to adhesins that belong to the P97 and P102 paralog families. P159 (MHJ_0494), a unique cilium and glycosaminoglycan-binding adhesin, is also extensively processed at these sites (Raymond *et al.*, 2013). Cleavage fragments remain bound on the cell surface by as of yet an unknown mechanism(s).

P216 protein sequences from the geographically diverse strains 232 (USA), J (United Kingdom), 7448 (Brazil), Beaufort (Australia), and 168 (China) are highly conserved. P216 is endoproteolytically-cleaved between amino acids 1041 – 1089 generating N-terminal 120 kDa (P120) and C-terminal 85 kDa (P85) fragments (Wilton *et al.*, 2009). The cleavage site was mapped to the sequence ^{1072}TNF↓QE^{1076} which conforms to the S/T-X-F↓X-D/E motif (Bogema *et al.*, 2011). S/T-X-F↓X-D/E motifs reside in disordered regions of P97 and P102 paralog family members suggesting that a relaxed, unstructured conformation allows proteolytic cleavage to occur (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2012; Seymour *et al.*, 2012). Recombinant fragments $F1_{P216}$ – $F3_{P216}$ were used to determine the ability of regions within P216 to bind cilia and glycosaminoglycans and to generate mono-specific polyclonal antiserum (Wilton *et al.*, 2009). While immunoblot experiments demonstrated that P216 is predominantly represented on the cell surface as N-terminal 120 kDa and C-terminal 85 kDa fragments, several lower abundance cleavage fragments were detected suggesting that cleavage occurs at multiple sites in P216 (Wilton *et al.*, 2009). These putative cleavage fragments have not been experimentally validated.

Here we characterised the full gamut of endoproteolytically-derived minor cleavage fragments of P216 generated in strains 232 and J, mapped endoproteolytic cleavage

sites by LC-MS/MS analysis and examined the ability of cleavage fragments to bind glycosaminoglycans and epithelial cell surface proteins.

5.3. Materials and Methods

5.3.1. Culture conditions

Mycoplasma hyopneumoniae (strain J and/or 232) was grown in modified Friis broth (Friis, 1975) and harvested as described previously (Scarman *et al.*, 1997).

5.3.2. Cell surface analyses

Cell surface biotinylation of freshly cultured *M. hyopneumoniae* cells was carried out as described previously (Bogema *et al.*, 2011) using E-Z link Sulfo-NHS-LC-biotin [Thermo Fisher Scientific]. Biotinylated surface proteins were purified by avidin chromatography and 2D gel electrophoresis (Bogema *et al.*, 2011; Nunomura *et al.*, 2005; Raymond *et al.*, 2013). Enzymatic cell surface shaving with trypsin was carried out as described previously and tryptic peptides were identified by LC-MS/MS (Deutscher *et al.*, 2010; Raymond *et al.*, 2013).

5.3.3. Preparation of M. hyopneumoniae whole cell lysate

M. hyopneumoniae cells were lysed by sonication in solubilisation buffer (7 M urea, 2 M thiourea, 40 mM Tris, 1% (w/v) C7BzO) as described previously (Raymond *et al.*, 2013; Seymour *et al.*, 2012). Proteins were reduced and alkylated and insoluble material was pelleted by centrifugation. Soluble protein was precipitated in acetone and the resulting protein pellet was resolubilised in 7 M urea, 2 M thiourea, 1% (w/v) C7BzO as previously described (Raymond *et al.*, 2013; Seymour *et al.*, 2012). For 2D separations, sample was supplemented with 0.2% (w/v) 3-10 BioLyte ampholytes [BioRad].

5.3.4. Triton X-114 phase extraction

Intact, surface-biotinylated *M. hyopneumoniae* cells were subjected to Triton X-114 phase extraction as described previously (Bordier, 1981). In addition to soluble aqueous and detergent fractions, a Triton X-114 insoluble pellet was obtained by centrifugation of the detergent fraction. This pellet was re-solubilised in solubilisation buffer before reduction and alkylation as above. Biotinylated surface proteins were purified by avidin chromatography and separated by 1D SDS-PAGE (Bogema *et al.*, 2011; Nunomura *et*

al., 2005; Raymond *et al.*, 2013). Proteins were in-gel trypsin-digested and peptides analysed by LC-MS/MS.

5.3.5. Two dimensional gel electrophoresis

250 µg of protein was cup-loaded onto partially rehydrated 11 cm pH 4 – 7 IPG strips [BioRad] or 11 cm pH 6 – 11 Immobiline drystrips [GE Healthcare]. Focusing was performed in a BioRad Protean IEF cell at a constant 20˚C and 50 µA current limit per strip using a 3-step program: slow ramp to 4000 V for 4 h, linear ramp to 10000 V for 4 h, then 10000 V until 120 kVh was reached. Following focusing, the strips were equilibrated in 2% SDS, 6 M urea, 250 mM Tris-HCl pH 8.5, 0.0025% (w/v) bromophenol blue for 20 min before SDS-PAGE as described below.

5.3.6. Protein separation and blotting

Proteins were separated by SDS-PAGE using BioRad midi TGX™ gels and TGS running buffer, fixed, stained with Coomassie Blue G250 overnight and destained with 1% acetic acid to remove background. Alternatively, gels were blotted to PVDF membrane as described previously (Raymond *et al.*, 2013). Proteins on PVDF membranes were blocked with 5% (w/v) skim milk powder in PBS with 0.1% Tween 20 (v/v) (PBS-T) at room temperature for 1 h. For P216 immunoblots, membranes were incubated with F1 – $F3_{P216}$ polyclonal antisera individually or pooled (diluted 1:200 in PBS-T) for 1 h at room temperature. Blots were washed and exposed to HRP conjugated anti-rabbit antibodies (diluted 1:2000 in PBS-T) for 1 h at room temperature. Membranes were routinely washed in three changes of PBS-T and developed with SIGMA*FAST*™ 3,3´-diaminobenzidine tablets [Sigma-Aldrich] as per manufacturer's instructions.

5.3.7. Heparin column affinity chromatography

Heparin affinity chromatography of *M. hyopneumoniae* whole cell lysates was carried out using 1 mL HiTrap™ Heparin HP columns [GE Healthcare] on a Waters 2690 Alliance LC separations module as described in detail previously (Raymond *et al.*, 2013). Heparin binding fractions were collected and pooled into low and high affinity interactions based on the elution profile over an increasing salt gradient, concentrated and separated further by 1D SDS-PAGE.

5.3.8. Avidin purification of interacting proteins

To identify *M. hyopneumoniae* proteins that bind to surface proteins on porcine kidney (PK15) epithelial cells a protocol was adapted from (Chen *et al*., 2011; Nunomura *et al.*, 2005) as described in detail previously (Raymond *et al.*, 2013). Briefly, PK15 cells were washed and cell surface proteins were labelled with membrane impermeable EZ link sulfo-NHS-LC-biotin [Thermo Fisher Scientific] for 30 min at 4˚C. The reaction was quenched, the cells washed and lysed with 1.0% Triton X-100 with protease inhibitors on ice for 30 min with vortexing. Surface biotinylated proteins were allowed to bind to avidin agarose slurry for 1 h at room temperature, before packing into a column and unbound proteins removed by washing with 6 volumes of PBS.

Cleared *M. hyopneumoniae* cell lysates in 0.5% Triton X-100 were incubated with the biotinylated PK15-avidin agarose mixture overnight at 4˚C on a rotating wheel, before packing into a column and removing non-interacting proteins with six washes of 25 mM Tris-HCl, 150 mM NaCl, pH 7.4. Interacting proteins were collected in five elutions of 100 mM Tris-HCl, 2 M NaCl, pH 7.4 and strongly bound proteins and the biotinylated PK15 proteins were eluted with 30% ACN and 0.4% trifluoroacetic acid. Fractions were concentrated, precipitated and separated by 1D SDS-PAGE before in-gel trypsin digestion and LC-MS/MS.

5.3.9. Liquid chromatography and tandem mass spectrometry (LC-MS/MS)

Entire gel lanes of interest were sliced into 16 equal pieces for in-gel trypsin digestion and analysis by LC-MS/MS. All clearly visible spots from 2D gels, with the exception of some spots forming spot trains, were excised manually, digested in-gel with trypsin and analysed by LC-MS/MS. Mass spectrometry and subsequent data analysis was carried out as described previously (Raymond *et al.*, 2013). Briefly, sample was loaded a C8 Trap column [Michrom, USA] connected to a NanoLC system [Eksigent, USA]. After washing, the peptides were eluted off the trap (300 nL min^{-1}) onto an PicoFrit column (75 μm × 150 mm) packed with Magic C18AQ resin [Michrom, USA]. Peptides were eluted from the column and into the source of a QSTAR Elite quadrupole-time-of-flight mass spectrometer [AB Sciex] at 2300V using the following program: 5−50% MS solvent B (98% ACN +0.2% formic acid) over 30 min for gel slices or 15 min for gel spots, 50−80% MS B over 5 min, 80% MS B for 2 min,

80−5% for 3 min. An Intelligent Data Acquisition (IDA) experiment was performed, with a mass range of 350−1500 Da scanned for peptides of charge state 2+ to 5+ with an intensity of more than 30 counts/scan. Selected peptides were fragmented, and the product ion fragment masses were measured over a mass range of 50−1500 Da. The mass of the precursor peptide was then excluded for 120 s for gel slices or 15 s for gel spots.

5.3.10. MS/MS data analysis

The MS/MS data files were searched using Mascot (hosted by the Walter and Eliza Hall Institute for Medical Research Systems Biology Mascot Server) against the LudwigNR database (composed of the UniProt, plasmoDB and Ensembl databases (vQ312. 19 375 804 sequences; 6 797 271 065 residues)) with the following parameter settings. Fixed modifications: none. Variable modifications: propionamide, oxidized methionine. Enzyme: semitrypsin. Number of allowed missed cleavages: 3. Peptide mass tolerance: 100 ppm. MS/MS mass tolerance: 0.2 Da. Charge state: 2+ and 3+. For biotinylated samples, variable modifications also included biotinylated lysine and N-terminal biotinylation. Scaffold (v3.00.02, Proteome Software Inc., Portland) was used to validate and compare MS/MS based peptide and protein identifications. Peptide identifications were accepted if their calculated probability assigned by the Protein Proph*et al*gorithm was greater than 95.0% and protein identifications were accepted if their calculated probability was greater than 80.0%. All semi-tryptic and cleavage-defining peptide identifications were manually inspected for quality of spectra (Figure S5.1).

5.3.11. Bioinformatic analysis of P216

Bioinformatic analysis of P216 used online resources: ProtParam (Wilkins *et al.*, 1999), PattInProt (Combet *et al.*, 2000) TMpred (Wilkins *et al.*, 1999), COILS (Lupas *et al.*, 1991) and PONDR VSL2 and VL-XT (Li *et al.*, 1999; Romero *et al.*, 1997; Romero *et al.*, 2001) as previously described (Bogema *et al.*, 2011; Deutscher *et al.*, 2012; Liu *et al.*, 2011; Raymond *et al.*, 2013).

5.3.12. Preparation of whole cell lysates for dimethyl labelling

All steps for dimethyl labelling preparation and purification were performed according to, or adapted from the protocol described by Kleifeld *et al.* (2011). Fresh washed *M. hyopneumoniae* cell pellets were resuspended in 4 M guanidine hydrochloride, 100 mM HEPES buffer at pH 8.0 and lysed by ultrasonication. Samples were diluted with 100 mM HEPES buffer and re-concentrated by passing through a 3000 MWCO filter. This washing step was performed 3 times to remove any free amino acids or other small cellular contaminants which could interfere with the labelling. Sample was reduced and alkylated with 5 mM tributylphosphine and 20 mM of acrylamide monomers for 90 min at room temperature.

5.3.13. Dimethyl labelling of bacterial proteins

Protein labelling was performed on 1 mg of *M. hyopneumoniae* strain J protein by the addition of 40 mM formaldehyde (ultra-pure grade) [Polysciences Inc., USA] in the presence of 20 mM sodium cyanoborohydride, buffered with 100 mM HEPES solution adjusted to pH 6 – 7 in a final volume of 1 mL, and incubated at 37°C for a minimum 4 h. The reaction was quenched by the addition of 100 mM ammonium bicarbonate and precipitated with 8 volumes of acetone and 1 volume of methanol at -20°C for 3 h. The precipitated protein was then pelleted by centrifugation at 14000 × g and washed with 5 volumes of methanol. The protein pellet was resuspended in 50 mM sodium hydroxide, pH 8.0 and digested with trypsin prior to peptide enrichment and analysis by LC-MS/MS.

5.3.14. Enrichment of labelled N-terminal peptides

To enrich for blocked N-terminal peptides, the high molecular weight aldehyde-derivatised polymer (HPG-ALD type II) [Flintbox Innovation Network] was added in a ratio 2 μg of polymer per 1 μg of peptides in the presence of 20 mM of sodium cyanoborohydride. The reaction was allowed to proceed at 37°C for a minimum of 4 h. Ammonium bicarbonate was added to 100 mM, adjusted to pH 6 – 7 and incubated for 30 min to quench the reaction. Samples were passed through a 10000 MWCO Ultrasep™ Advance Centrifugal Device by centrifugation at 10000 × g. Filtrates containing unbound N-terminal peptides, were acidified with 0.02% Trifluoroacetic acid and reduced to ~15 μL in a Vacufuge™ Concentrator 5301 [Eppendorf, Germany], ready for analysis by LC-MS/MS.

5.4. Results

5.4.1. Structural and chemical features of P216

In strain J of *M. hyopneumoniae*, the *mhj_0493* gene encodes a 216 kDa protein ($P216_J$) comprising 1878 amino acids. The P216 protein sequence is conserved among geographically distinct strains of *M. hyopneumoniae*, sharing 97% identity in strains J, 232, 7448 and 168. Alignments with P216 homologs from these strains show that sequence differences are largely restricted to variation in lengths of poly-Q regions at amino acid positions 1258 and 1375 in strain J (Wilton *et al.*, 2009). Alignment of P216 sequences from strains J and 232 shows 97.081% identity with 1829 identical amino acids and 31 single nucleotide polymorphisms (SNPs) generating similar amino acids. TMpred analysis identified a single putative transmembrane domain spanning amino acids 7 – 30 (TMpred score: 2090) (Figure 5.1). LC-MS/MS data identify N-terminal tryptic peptides for most members of the P97 and P102 paralog families indicating that signal sequences are not processed by signal peptidases and remain intact (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004). Consistent with this, the most N-terminal tryptic peptide identified in P216 spans amino acids 10 – 31 (see later) suggesting that N-terminal cleavage fragments remain tethered to the *M. hyopneumoniae* membrane (Wilton *et al.*, 2009).

Previous studies indicate that members of the P97 and P102 adhesin families and the P159 (*mhj_0494*) adhesin display uneven charge density generating modular domains (Bogema *et al.*, 2012; Burnett *et al.*, 2006; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Wilton *et al.*, 2009). $P216_J$ comprises 264 positively charged residues (195 K, 56 R, 13 H) and 240 negatively charged amino acids (124 E, 116 D) with an overall p*I* of 8.6. Four acidic domains were identified at positions 295 – 354 (p*I* = 3.76), 541 – 600 (p*I* = 4.37), 781 – 840 (p*I* = 4.54) and 1039 - 1320 (p*I* = 5.03). The remaining regions of P216 including amino acids 10 – 294 (p*I* = 9. 17), 355 – 540 (p*I* = 9.43), 601 – 780 (p*I* = 9.40), 841 – 1038 (p*I* = 9.14) and 1321 – 1878 (p*I* = 9.09) are strongly basic. As was observed in P159, the dominant cleavage site (^{1072}TNF↓QEE^{1077}) at position 1074 resides within a large acidic domain spanning 281 amino acids with a p*I* of 5.03 (Bogema *et al.*, 2011). The VSL2 (PONDR) algorithm predicts the presence of seven disordered regions spanning at least 40 consecutive amino acids in the $P216_J$ sequence at positions 334 – 377 (44 amino acids), 501 – 583 (83 amino acids), 758 – 805 (48

amino acids), 1002 – 1106 (105 amino acids), 1246 – 1292 (46 amino acids), 1365 – 1406 (42 amino acids) and 1503 – 1560 (58 amino acids). Coiled-coil analyses were also performed as some protein disorder predictors interpret coiled-coils as disordered regions (Ferron *et al.*, 2006). COILS identified two coiled-coil regions at positions 502 – 527 and 1252 – 1282 in $P216_J$, one of which was also identified as a disordered region using PONDR VSL2 and VL-XT algorithms. The modular structure that defines $P216_J$ is depicted in Figure 5.1.

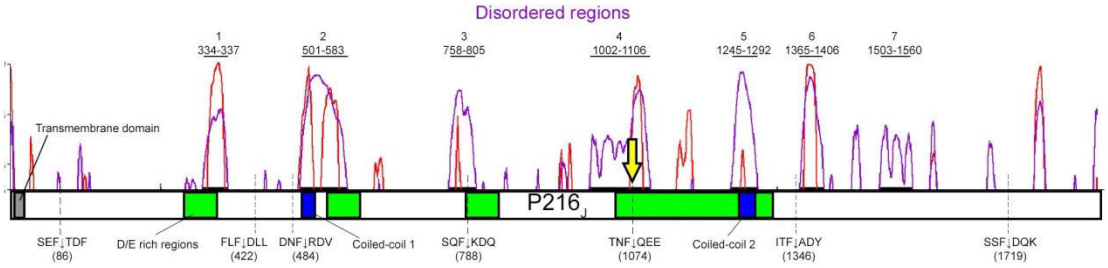


Figure 5.1. Schematic depiction of the modular nature of $P216_J$.
Disordered regions are shown as purple (PONDR VSL2) or red (PONDR VL-XT) peaks above the molecule, with major domains as indicated. Coiled-coils and transmembrane domains as predicted by COILS and TMpred respectively are also indicated. Vertical dotted lines show location of putative S/T-X-F↓-X-D/E-like cleavage motifs. The yellow arrow indicates the dominant S/T-X-F↓-X-D/E cleavage site that generates P120 and P85. D/E rich acidic domains are indicated by green boxes.

Previously, we determined that P216 is target to a dominant cleavage event that generates an N-terminal 120 kDa protein (P120) and a C-terminal 85 kDa protein (P85) (Wilton *et al.*, 2009). The cleavage site was precisely mapped to the sequence ^{1072}TNF↓QEE^{1077} by identification of an unusual tryptic fragment with sequence ^{1075}QEEADLDQDGQDDSK^{1089} (Bogema *et al.*, 2011). The cleavage site resides within a disordered region spanning 105 amino acids and conforms strictly to the motif S/T-X-F↓-X-D/E (Bogema *et al.*, 2012). Endoproteolytic processing is a feature of all members of the P97 and P102 adhesin families (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Wilton *et al.*, 2009) and of P159, encoded by the *mhj_0494* gene which lies 59 bases upstream of the P216 gene (*mhj_0493*) in the *M. hyopneumoniae* genome (Raymond *et al.*, 2013). Dominant cleavage events occur at

S/T-X-F↓X-D/E sites in most of the P97 and P102 adhesin paralogs and in P159. The association of this cleavage motif with regions of protein disorder is hypothesised to influence the efficiency of cleavage (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2012; Raymond *et al.*, 2013). Apart from the dominant cleavage site ^{1072}T-N-F↓Q-E-E^{1077} described above, PATTINProt identified S/T-X-F↓X-D/E motifs at amino acid positions ^{84}S-E-F↓T-D-F^{89}, ^{786}S-Q-F↓K-D-Q^{792} and 29 related motifs with either S/T or D/E residues mismatched. Four of these similar motifs at positions ^{420}F-L-F↓D-D-L^{425}, ^{484}D-N-F↓R-D-V^{489}, ^{1344}I-T-F↓A-D-Y^{1349} and ^{1717}S-S-F↓D-Q-K^{1722} represent plausible cleavage sites in P216$_J$, but only ^{786}S-Q-F↓K-D-Q^{792} resides within one of the disordered regions spanning more than 40 amino acids (Figure 5.1). We characterised cleavage fragments of P216 to determine if these putative S/T-X-F↓X-D/E sites are prone to endoproteolytic cleavage (see below). The more recently identified cleavage motif L-X-V↓X-V/A-X (Raymond *et al.*, 2013) was not found within the P216 sequence.

5.4.2. Immunoblotting studies with anti-F1$_{P216}$ – F3$_{P216}$ sera

In order to identify protein fragments from P216$_J$, whole cell lysates of *M. hyopneumoniae* were separated by 1D SDS-PAGE and blotted to PVDF membrane. Blots were probed with antisera raised against recombinant P216 fragments F1 – F3$_{P216}$ generated previously (Wilton *et al.*, 2009). Western blotting did not identify any protein at the mass of the intact 216 kDa pre-protein, but confirmed the presence of the dominant cleavage fragments P120 and P85. The staining pattern also indicated that P216 is a target of further processing events, with significant reactivity detected from proteins with masses of 20 – 80 kDa (Figure 5.2). In addition, as the recombinant fragments from which the antisera were generated only span amino acids 35 – 1444 of P216 (Wilton *et al.*, 2009), C-terminal-specific cleavage fragments may not be detected by blotting. These data were useful for identifying further cleavage fragments of P216$_J$ (see below).

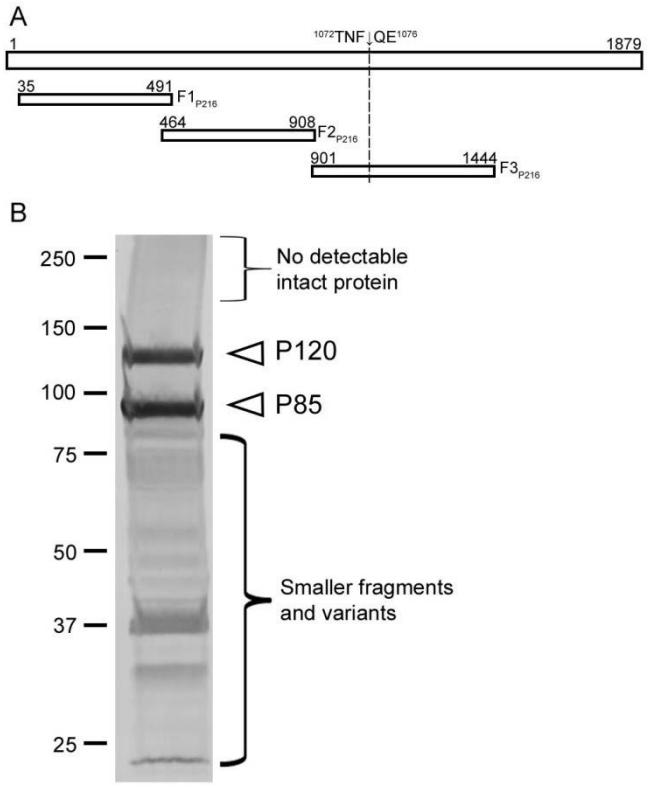


Figure 5.2. Detection of P216 cleavage fragments by immunoblotting with antisera generated to recombinant protein fragments.

A. Recombinant fragments of $P216_{232}$ used to generate antisera. B. 1D SDS-PAGE of *M. hyopneumoniae* whole cell lysates probed with a pool of F1 – $F3_{P216}$ antisera (1:200 dilution).

5.4.3. Identification of cleavage sites in P216

LC-MS/MS analysis of protein spots cut from pH 4 – 7 and 6 – 11 2D gels showed that the vast majority of cleavage fragments localised in the p*I* range from 6 – 11 with a notable exception (discussed later). These analyses were performed on whole cell lysates prepared separately from strains J and 232. Data from 2D gels from strain 232 are presented here, although our results suggest that processing events in strains J and 232 are similar. We identified 40 spots on an 11 cm pH 6 – 11 gel that generated tryptic peptides that mapped to the $P216_{232}$ sequence, with the major spot trains highlighted in Figure 5.3. Some spots with poor peptide coverage could not be accurately defined as fragments and were not included. On pH 6 – 11 2D gels, six spot trains (labelled 1 – 6) were identified containing peptides mapping to P216 as well as two single spots containing P216 fragments (spots 7 and 8) and a single unique spot resolved on a 4 – 7

2D gel (spot 9) (Figure 5.3). All spots within a given spot train were identified to contain similar or identical peptide information mapping to one or more fragments of the P216 protein. Both dominant cleavage fragments, the N-terminal P120 (113 kDa on gel) and C-terminal P85 (99 kDa on gel) were identified at their previously reported masses with peptide coverage to within 50 (20 for P85) amino acids of the dominant cleavage motif ^{1072}T-N-F↓Q-E^{1076}. Smaller cleavage products (spot trains 3 – 8, Figure 5.3) were identified at masses of approximately 85, 78, 62, 51, 33 and 21 kDa in 2D pH 6 – 11 gels. Despite the apparent simplicity of the cleavage pattern on 2D gels, suggesting 8 cleavage fragments, peptide coverage reveals some spot trains contain multiple, overlapping, co-migrating fragments. Careful analysis of peptide coverage, p*I* and molecular mass data from these 2D gels identified 16 possible cleavage fragments (and possibly more minor variants due to aminopeptidase activity, discussed later) derived from the P216 precursor protein (Figure 5.9).

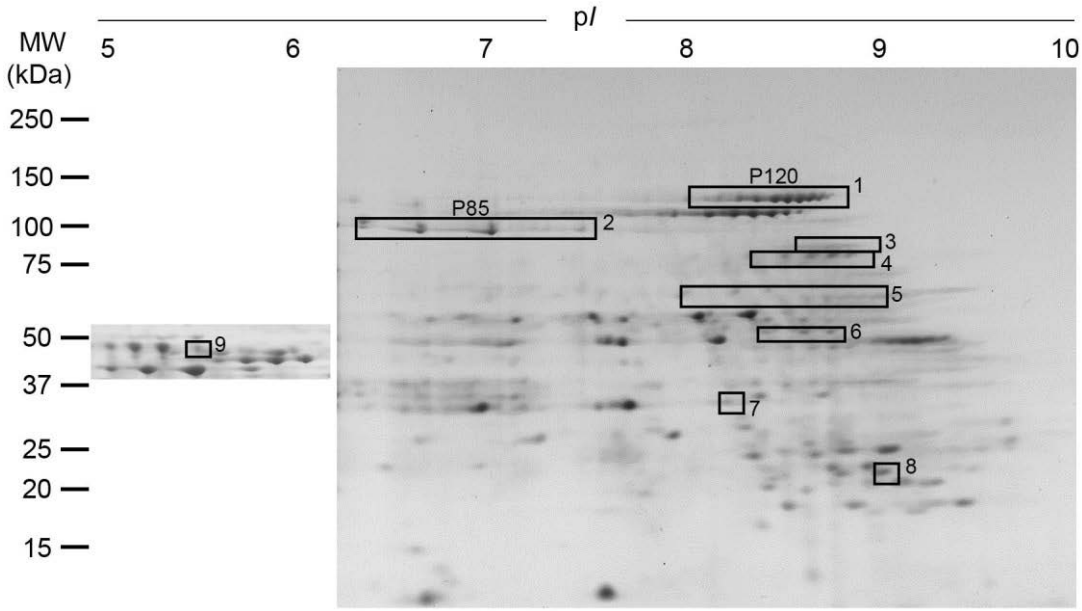


Figure 5.3. Identification of protein fragments of P216$_{232}$ by 2D PAGE.
Gel image depicting the location of spot trains cut from a 6 – 11 gel of *M. hyopneumoniae* whole cell lysate and one unique spot from a 4 – 7 gel, found to contain peptides mapping to P216. Schematic representation of peptide coverage obtained by LC-MS/MS can be seen in Figure 5.9.

5.4.4. Surface exposed cleavage products of P216

Many P216 protein fragments were identified to be surface exposed. While it has been previously determined that P216 resides at the cell surface, cell surface biotinylation of intact *M. hyopneumoniae* strain J followed by purification using avidin affinity chromatography and subsequent 2D gel electrophoresis has allowed us to identify specific products of proteolytic processing that are also surface exposed. 33 spots from six spot trains were identified to contain peptides mapping to $P216_J$ on pH 3 – 10 2D gels. Again, all spots within a given spot train were identified to contain similar peptide information mapping to one or more fragments of P216, with cleavage patterns complicated by overlapping and co-migrating fragments. Although the process of biotinylation may alter the migration of protein cleavage products along the p*I* gradient, a number of fragments were identified by peptide mapping to correspond to fragments characterised from unlabelled gels. Dominant fragments P120 and P85 correspond to spots in trains labelled A and B. Minor cleavage fragments C and D1 correspond approximately to fragments 3.2 and 4.2 from unlabelled gels (Figure 5.4, see also Figure 5.9).

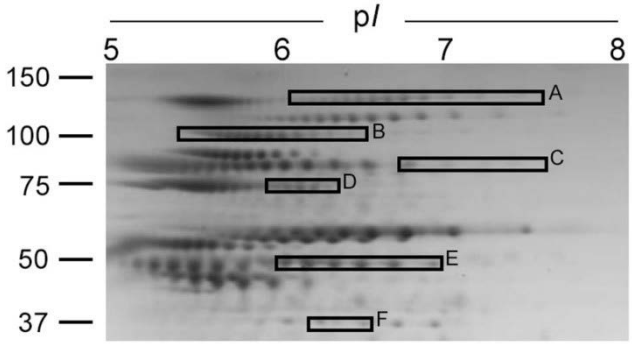


Figure 5.4. Identification of biotinylated (surface) fragments of $P216_J$ by 2D SDS-PAGE and LC-MS/MS.

A. p*I* 3 – 10 2D gel of biotinylated proteins of *M. hyopneumoniae* showing spot trains containing peptides mapping to $P216_J$ (labelled A – F). Schematic representation of surface exposed cleavage fragments of P216 can be seen in Figure 5.9.

5.4.5. Identification of intact P216

Purified biotinylated surface proteins from the insoluble fraction of a Triton X-114 preparation of *M. hyopneumoniae* strain J were separated by 1D SDS-PAGE, in-gel

trypsin digested and analysed by LC-MS/MS (Figure 5.5). In a high molecular mass slice ~200 – 250 kDa, 46 peptides were identified mapping to the full length of $P216_J$ with a Mascot protein score of 2096 and 31% coverage of the molecule (Figure S5.3). Peptides spanning the dominant cleavage site were not identified to confirm the ^{1072}S-T-F↓Q-E^{1076} site was intact. This is to be expected, as the site resides in a stretch of 42 amino acids lacking lysine or arginine residues to create tryptic peptides detectable by the LC-MS/MS used here. In addition, a semi-tryptic peptide ^{13}AAAIIGSTVFGTVVGLASK^{31} was identified at the N-terminus of the protein (see later). This is the first time the P216 pre-protein has been identified, and underscores the efficiency of the dominant cleavage site, particularly since antisera failed to detect intact P216 in blots of whole cell lysates, as noted in Figure 5.2.

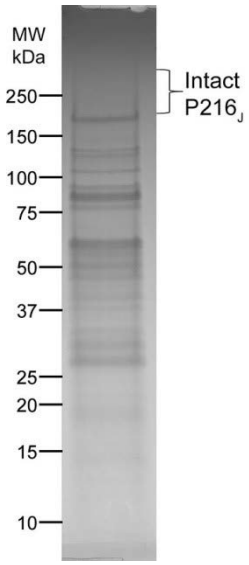


Figure 5.5. Identification of intact $P216_J$ from a 1D gel of purified biotinylated surface proteins from a Triton X-114 insoluble fraction.
Peptides mapping to the entire length of the protein were identified by LC-MS/MS from a band excised at ~250 kDa as indicated (Figure S5.3).

5.4.6. Identification of true N-termini and cleavage sites

LC-MS/MS identification of semi-tryptic peptides has previously been used to aid in the determination of proteolytic cleavage sites, usually in conjunction with supporting Edman degradation sequence information or mass context from 2D gels (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Raymond *et al.*, 2013). Atypical tryptic peptides identified

by LC-MS/MS, where one terminal residue is inconsistent with the action of trypsin cleavage C-terminal to K/R (semi-tryptic peptides) allowed us to identify proteolytic sites which are consistent with the sizes of fragments observed on gels. Semi-tryptic peptides were identified at multiple positions within P120 (Figure 5.6 and Table 5.1). The dominant cleavage site that generates P120 and P85 was also confirmed by identifying the semi-tryptic N-terminal peptide sequence, ^{1074}QEEADLDQEGQDDSK^{1088}, by LC-MS/MS (Table 5.1). Within P85, four semi-tryptic peptides delineated putative cleavage sites (Table 5.1). In addition to these, peptide 11 was identified to possess a semi-tryptic C-terminus at ^{1338}RLMNTPITF^{1346}↓ADY. This site, I-T-F↓A-D-Y resembles a dominant cleavage motif and may be the cleavage site generating fragment 5.3. In P97 a TNT↓NTN motif was identified as a site of proteolytic cleavage (Djordjevic *et al.*, 2004; Jenkins *et al.*, 2006). In P216, fragment mass and p*I* data strongly suggests a cleavage event at a site similar to this motif; ^{1417}NNTN^{1420} to generate C-terminal fragment 6.3 and acidic fragment 9, although no semi-tryptic peptides were identified in this region.

Table 5.1. Semi-tryptic peptides identified which align to putative cleavage sites.

Peptide	N-terminal residues	C-terminal residues	Matching fragments
1	PKS	^{83}SEFTDFVSK	Most N-term peptide
2	KDE	^{297}TFLSSIDLK	F2
3	KPN	^{398}SIKDLVNATLAR	4.2 and D1
4	KAF	^{496}GLLYPGVNEELEQAR	6.1 C-term
5	LKA	^{540}INNQEGLEEDDNITER	Separating 5.1 and 5.2
6	TNF	^{1074}QEEADLDQEGQDDSK	Separating P120 and P85
7	DKW	^{1228}LASIPLVIHQQMLR	4.2
8	LPE	^{1562}NYLNLVNQPWK	E2 C-term
9	VFF	^{1685}GNWENSSMNSQAQTPTWEK	8
10	TRT	^{1726}FVLTTNAPLPLWK	C-term D2
11	RLMNTPITF^{1346}	ADY	5.3

Underlined residues represent the N-terminus of other semi-tryptic peptides identified by LC-MS/MS, which may indicate clipping by a surface accessible aminopeptidase. Supporting mass spectra in Figure S5.1.

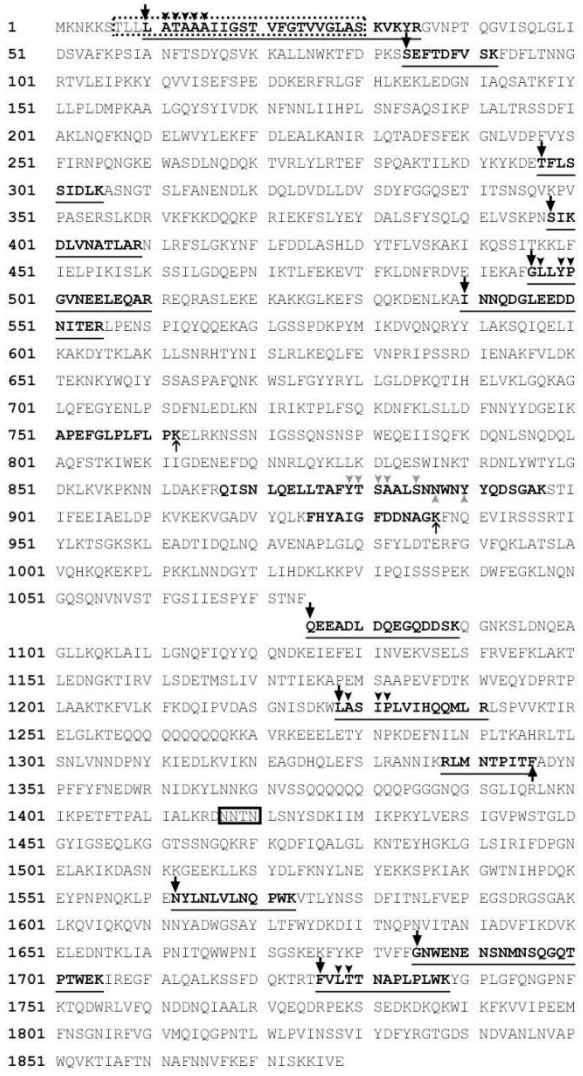


Figure 5.6. Semi-tryptic peptides identified from P216$_J$ indicating putative cleavage sites.

Filled downward-pointing arrows indicate N-terminal semi-tryptic peptides and an upward-pointing arrow shows a C-terminal semi-tryptic peptide. Upward-pointing line arrows show putative C-terminal tryptic peptides that match peptide coverage from unlabelled or biotinylated 2D gel spots. Smaller arrowheads represent N-terminal peptides that are likely to result from aminopeptidase activity following an endoproteolytic event. Grey arrows show semi-tryptic peptides indicative of peptidase activity not matching to any identified cleavage fragments. Residues in dashed box indicate the predicted transmembrane domain and the solid box indicates the NTNT-like ^{1417}NNTN^{1420} motif.

5.4.7. N-terminal dimethyl labelling

To further validate the presence of cleavage motifs, an N-terminal dimethyl labelling technique (N-TAILS) was used. N-terminal dimethyl labelling of *M. hyopneumoniae* whole cell lysates identified N-terminal peptides mapping within $P216_J$. Six overlapping peptides spanning amino acids 10-35 were labelled. These peptides differ by the sequential loss of one amino acid indicative of aminopeptidase activity (Figure 5.7).

Figure 5.7. Identification of overlapping dimethyl-labelled peptides in the putative transmembrane domain of $P216_J$ by N-TAILS.

Amino acid positions are shown above residues, * indicates N-terminal dimethyl label. Mascot scores and expectation values of each peptide are indicated. Supporting mass spectra in Figure S5.2.

5.4.8. Regions of P216 that bind heparin

Heparin-agarose affinity chromatography using whole cell lysates of *M. hyopneumoniae* strains J and 232 was used to identify cleavage fragments of P216 that bind heparin. Proteins displaying high affinity for heparin (elutions in >0.75 M NaCl) were separated by 1D SDS-PAGE followed by in-gel trypsin digestion and LC-MS/MS. Heparin-binding fragments of $P216_J$ displayed masses from 92 - 135 (slice 5), 73 - 92 (slice 6), 32 - 37 (slice 10) and 20 - 26 kDa (slice 12). Three tryptic peptides were also detected in a fragment at approximately 250 kDa that mapped only to the N-terminal P120 fragment of $P216_J$ (Figure 5.8). This may be evidence of the intact P216 pre-protein, or the presence of low abundance P120 dimers. Fragments identified in slices 5 and 6 map to the known fragments P120 and P85 respectively. Smaller fragments identified in slices 10 and 12 mapped to C-terminal fragments of P85, with slice 12 approximately matching fragment 8 as identified on 2D gels (Figure 5.8).

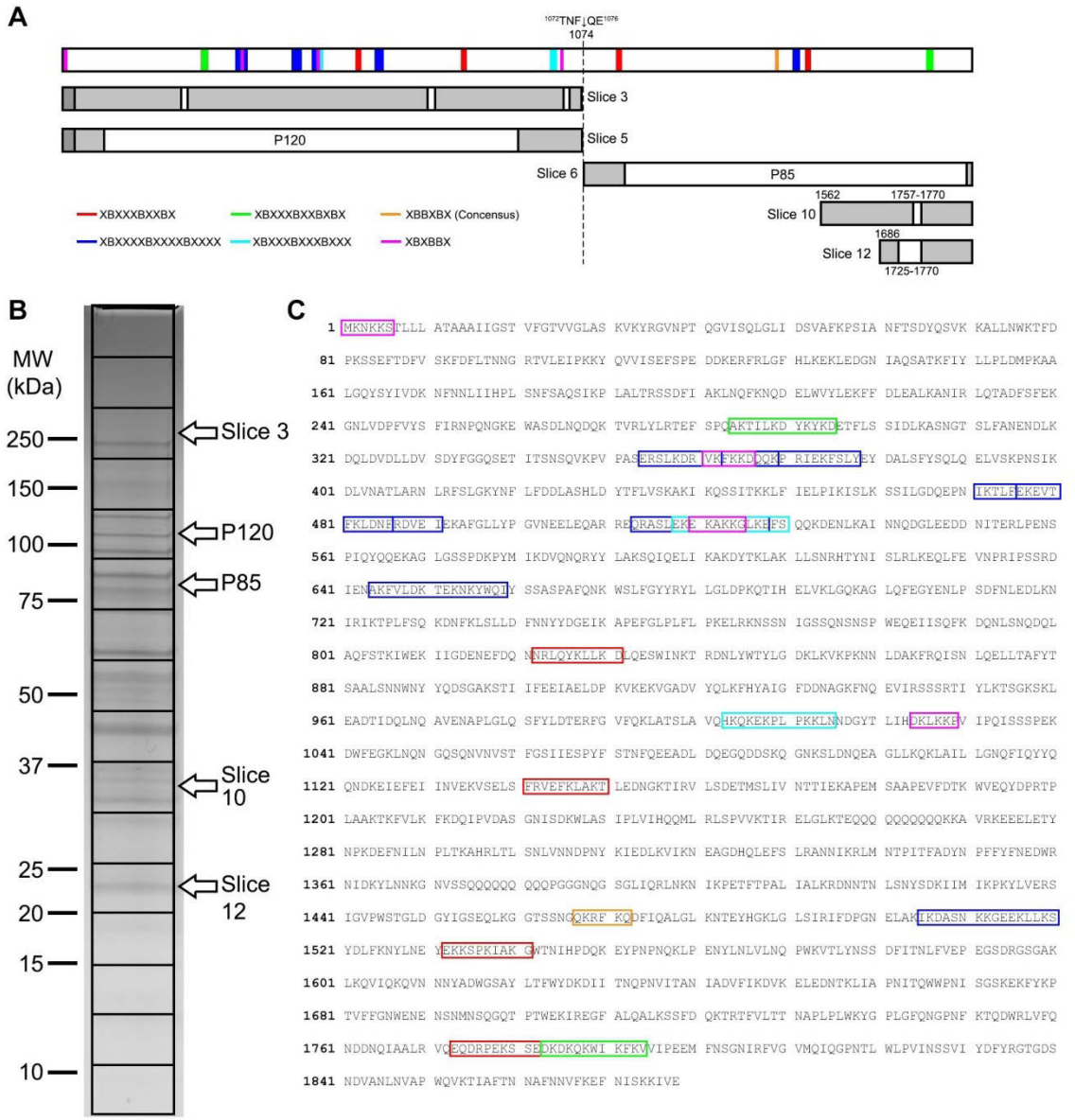


Figure 5.8. Heparin binding fragments of P216$_J$.

A. Schematic representation of cleavage fragments identified from the gel in B. The positions of putative heparin-binding motifs throughout the protein are indicated above. Grey bars represent putative fragment extension where no peptides were identified, with areas of peptide coverage indicated by white bars. Dark grey bars represent N-terminal regions of variable aminopeptidase activity as shown in Figure 5.7. B. 1D gel of high affinity heparin-binding protein elutions (>0.75 M NaCl) of whole cell lysates of *M. hyopneumoniae.* Arrows indicate slices from which peptides mapping to P216$_J$ fragments were identified. C. Putative heparin binding motifs identified by PATTINProt are boxed in corresponding colours and show distribution throughout the entire protein.

Strain 232 showed high-affinity heparin-binding fragments of P216 at masses ranging from 50 – 250 kDa. Peptides mapping only to P120 were again detectable at approximately 250 kDa further supporting evidence that P216 can be found as an intact pre-protein only upon enrichment, or that P120 dimers are formed in *M. hyopneumoniae*, albeit in low abundance. P120 and P85 were also identified at their correct molecular masses in the gel. Cleavage fragments mapping to the N- and C-terminus of P120 were also identified between 50 – 65 kDa corresponding approximately to fragments 5.1, 5.2, 6.1 and 6.2 from unlabelled 2D gels, as depicted in Figure 5.9. Unlike J strain, no lower mass cleavage fragments were identified. This data suggests that the degree of processing occurring in $P216_{232}$ is less compared to $P216_{J}$.

Putative heparin-binding regions conforming to motifs previously identified in P159 (Raymond *et al.*, 2013) represented as single or tandem repeats of X-B-X(1,4)-B-X(1,4)-B-X (where B is a basic residue, K or R) were identified at 20 sites within $P216_{J}$ (Figure 5.8). A single consensus heparin-binding motif (XBBXBX) (Cardin and Weintraub, 1989) was identified in P85 at ^{1467}QKRFKQ^{1472}. This distribution of putative heparin binding sites provides an explanation for all identified fragments of P216 that bind heparin and suggest that all low abundance fragments of P216 (see Figure 5.9) retain the ability to bind heparin.

5.4.9. P216 binding capacity to PK15 cell surface proteins

Previously, latex bead adherence assays demonstrated that recombinant fragments F1-$F3_{P216}$ had the capacity to bind PK15 cells, and microtitre plate-based cilium adherence assays showed F2 and $F3_{P216}$, but not $F1_{P216}$ reproducibly bound cilia (Wilton *et al.*, 2009). Here, a systems wide strategy (Raymond *et al.*, 2013) was applied to identify *M. hyopneumoniae* proteins that interact with proteins that are displayed on the cell surface of porcine kidney epithelial-like cell monolayers (PK15 cells). PK15 cell surface proteins were biotinylated and immobilised onto avidin agarose. Native *M. hyopneumoniae* proteins were incubated with this mixture in order to identify interacting proteins. Following elution, proteins were separated by 1D SDS-PAGE, trypsin digested and analysed by LC-MS/MS. The two dominant cleavage fragments, P120 and P85, were identified to bind PK15 cell surface proteins using this method.

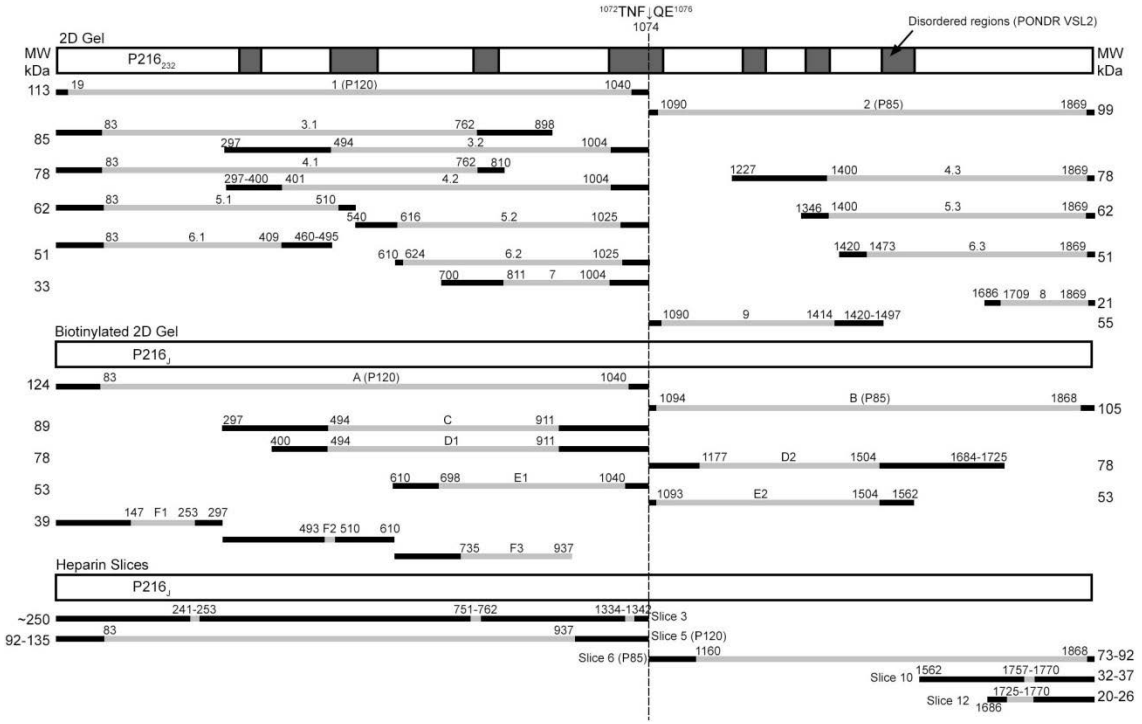


Figure 5.9. Cleavage fragments of P216.

Fragments were identified from 2D PAGE, cell surface biotinylation, enrichment and 2D PAGE and heparin chromatography followed by 1D SDS-PAGE as indicated. Fragments were characterised from indicated apparent molecular mass and p*I* information (where applicable) and peptide matches from LC-MS/MS of in-gel trypsin digestions, with areas of peptide coverage indicated by grey bars. Black bars represent putative fragment extension where no peptides were identified. Residue numbers are indicated.

5.5. Discussion

Whole-cell proteomic analyses of *Mycoplasma hyopneumoniae* identified P216 as one of the three most frequently identified proteins, suggesting it is abundantly expressed (Wilton *et al.*, 2009) (See Chapter 3). $P216_J$ is a modular protein containing four discrete acidic domains (p*I*s ranging from 3.76 – 5.03) 50 – 80 residues in length that are interspersed between amino acids 295 to 1320 in the 1878 amino acid protein. The remaining regions of the molecule are highly basic. In addition, $P216_J$ has seven putative disordered regions each spanning more than 40 consecutive amino acids interspersed throughout the molecule. Adhesins belonging to the P97/P102 paralog families are often cleaved at S/T-X-F↓-X-D/E-like motifs that reside within or abut regions of disorder (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2012; Raymond *et al.*, 2013). The efficiency of cleavage at these sites is perceived to be influenced by the flexible nature of disordered regions which enables access to molecules (e.g. proteases) that interact at these sites during interactions with other proteins (Uversky and Dunker, 2010). $P216_J$ undergoes a dominant cleavage event at the C-terminal side of a phenylalanine residue at position 1074 in the motif ^{1072}T-N-F↓Q-E^{1076} generating N- and C-terminal fragments of 120 kDa (P120) and 85 kDa (P85) (Bogema *et al.*, 2011; Wilton *et al.*, 2009). This dominant cleavage event appears to be conserved in geographically-diverse strains of *M. hyopneumoniae* (Wilton *et al.*, 2009) underscoring its importance to the function of this molecule. Consistent with our hypothesis, the dominant cleavage site ^{1072}T-N-F↓Q-E^{1076} in P216 lies within a predicted disordered region of 104 amino acids (Figure 5.1) using the PONDR predictors VSL2 and VL-XT.

In $P216_J$, we identified five sequences that strictly conform to the S/T-X-F↓-X-D/E-like motif and numerous others that shared sequence identity to the motif but lacked either the S/T or D/E residues. Here we identify a semi-tryptic peptide with sequence ^{1338}RLMNTPITF^{1346} which provides evidence of cleavage at a site C-terminal to a phenylalanine residue at position 1346 in the S/T-X-F↓-X-D/E-like motif ^{1344}I-T-F↓A-D-Y^{1349}. Of the numerous S/T-X-F↓-X-D/E-like motifs identified in P216, only ^{1344}I-T-F↓A-D-Y^{1349} abuts a region of protein disorder. As this cleavage event is consistent with protein fragment data from 2D gels (Figure 5.3) we conclude this is a legitimate albeit inefficient cleavage site generating fragment 5.3. Of the sites within P216 that match the

dominant cleavage motif, ^{786}S-Q-F↓K-D-Q^{792} also resides within a region of protein disorder of 48 amino acids as predicted by the PONDR predictor VSL2. However, the VL-XT algorithm is not in agreement with VSL2, and there is no fragment data to confirm cleavage at this site.

While a previous study indicated that P120 and P85 represented the major cleavage fragments of P216, more sophisticated analyses presented here indicated that processing in P216 is more extensive than previously thought. We identified 18-20 putative cleavage fragments ranging from 20 – 120 kDa spanning different regions within the 216 kDa pre-protein (Figure 5.9) by LC-MS/MS analysis of *M. hyopneumoniae* strain J and 232 proteins resolved by a combination of 2D-PAGE and enrichment of biotinylated surface proteins by avidin chromatography. Cleavage fragments of $P216_J$ were also inferred by immunoblotting using antisera raised against recombinant polyhistidine fusion proteins spanning regions of $P216_{232}$ (Wilton *et al.*, 2009). Consistent with the identification of cleavage fragments, 11 semi-tryptic peptides (Table 5.1) where one end was not consistent with cleavage by trypsin were characterised. The location of these unusual tryptic peptides within $P216_J$ is consistent with the mass and p*I* of P216 cleavage fragments observed by 2D-PAGE. With the exception of semi-tryptic peptide 11 ^{1338}RLMNTPITF^{1346}, none of these cleavage sites bear sequence identity with the S/T-X-F↓-X-D/E-like motif nor are they strictly related to the more recently identified cleavage motif L-X-V↓X-V/A-X (Deutscher *et al.*, 2012; Raymond *et al.*, 2013) identified in the P97 paralog Mhp385. However, we noted that a K/R residue resides within three amino acids N-terminal to the cleavage site in 7 of 11 semi-tryptic peptides shown in Table 5.1. We hypothesise that cleavage is performed by a trypsin-like protease and following cleavage, surface accessible aminopeptidases target neo-N-termini, removing several amino acids (Figure 5.6). Consistent with this hypothesis i) we recently identified cleavage fragments of P159 with tryptic N-terminal fragments indicating that surface accessible trypsin-like activity occurs in *M. hyopneumoniae* (Raymond *et al.*, 2013); ii) *M. hyopneumoniae* possesses several putative serine endoproteases, with MHJ_0525 and MHJ_0568 expressed during culture of *M. hyopneumoniae in vitro* (See Chapter 3); and iii) trypsin-like activity has been demonstrated by *M. hyopneumoniae* (Moitinho-Silva *et al.*, 2013). While these events cloud interpretation of the precise sites of endoproteolytic processing they provide the

first evidence that *M. hyopneumoniae* expresses aminopeptidases that target neo-N-termini generated by endoproteases that process adhesin families.

These data provide evidence that the information content of P216 and other members of the P97 and P102 paralog families may be significantly underestimated. We present evidence here that P216 can be cleaved into smaller fragments, the identities of which can be confirmed by LC-MS/MS. However, we also present evidence in Figure 5.6 and Table 5.1 that a considerable number of cleavage events may be generating numerous low-abundance fragments with masses ranging from 1 – 20 kDa. Given the large number of putative heparin-binding motifs distributed throughout the P216 sequence, and the ability of the dominant cleavage fragments to bind directly to epithelial cell surface proteins as indicated by PK15 binding experiments, many of these fragments may interact with surface receptors and extracellular matrix components on the surface of porcine cells in the respiratory tract and in tissue sites distant from the lung. These events are reminiscent of the widely documented process of ectodomain shedding in eukaryotes, a proteolytic mechanism that releases biologically-functional proteins and peptides into the extracellular environment (Horiuchi, 2013). While we have no data here to indicate that these smaller peptides are generated or that they have further biological activity, studies are underway to examine this hypothesis.

The genome of *M. hyopneumoniae* possesses ten genes that are annotated to encode putative proteases; seven of which are putative peptidases (Moitinho-Silva *et al.*, 2013; Robinson *et al.*, 2013). Our analyses of P216 cleavage fragments by LC-MS/MS identified a series of semi-tryptic peptides whose patterns of formation are consistent with aminopeptidase activity on the surface of *M. hyopneumoniae* (Figure 5.6). As part of a larger proteogenomic study of *M. hyopneumoniae*, we used dimethyl labelling to map *bona fide* start sites in proteins (Berry, Djordjevic, unpublished data). Using this approach, we identified a series of six overlapping, dimethyl-labelled semi-tryptic peptides that differed from each other by the serial loss of an N-terminal amino acid. The most N-terminal dimethyl-labelled peptide commenced at amino acid position 10 in the centre of the only transmembrane domain predicted by TMpred in the P216 sequence. As the peptide ^{10}LATAAAIIGSTVFGTVVGLASKVKYR^{35} is the most N-terminal tryptic peptide identified in P216, our data indicates that an endopeptidase

targeting hydrophobic regions in secreted molecules is functional in *M. hyopneumoniae*. The site of cleavage is depicted by the sequence ^{7}T-L-L↓L-A-T^{12} which bears some similarity to the L-X-V↓X-V/A-X motif first described by us (Deutscher *et al.*, 2012; Raymond *et al.*, 2013). It can be noted that this motif has consistently been identified in hydrophobic regions, whereby X largely represents hydrophobic amino acids (typically L,A,V,L,F or I). Dimethyl-labelled peptides show evidence of sequential removal of amino acid leucine at position 10, alanine at positions 11, 13, 14 and 15 and threonine at position 12 (Figure 5.7). The identification of the series of semi-tryptic peptides shown in Figure 5.7 suggests that cleavage at ^{7}T-L-L↓L-A-T^{12} occurs with a degree of efficiency that exceeds those events that generate the majority of the low abundant cleavage fragments shown in Figure 5.9. This trend of clipping of hydrophobic residues can also be seen in semi-tryptic peptides series 4, ^{496}GLLYPGVNEELEQAR, and semi-tryptic peptide series 10, ^{1726}FVLTTNAPLPLWK (Table 5.1). While it is tempting to speculate that genome-reduced Mollicute species may rely on the concerted activity of an endoprotease(s) and aminopeptidases to recover the hydrophobic amino acids that comprise the transmembrane domain of secreted proteins after they have been transported to the cell surface, further analysis is needed. For this hypothesis to be valid, both the endoprotease and the aminopeptidases would need to be surface accessible.

Here we have identified the intact 216 kDa pre-protein for the first time, despite not identifying it by Western blotting of whole cell lysates. It was identified in biotinylated *M. hyopneumoniae* strain J cells from a Triton X-114 insoluble fraction that had been purified by avidin chromatography. The presence of the intact pre-protein only in this enriched biotinylated fraction is consistent with secretion of the full-length molecule and suggests all processing of this molecule occurs at the cell surface of *M. hyopneumoniae*, which is consistent with our hypothesis of an ectodomain shedding-like process occurring in this bacterium. These data highlight the versatile manner in which the P97 and P102 adhesins can be presented on the cell surface simply by regulating protease activity.

The combined evidence presented here and in cleavage patterns reported in P159 (Raymond *et al.*, 2013) indicate that *M. hyopneumoniae* displays aminopeptidase activity on its cell surface that target neo-N-termini created by endoproteolytic

processing. We recently showed that MHJ_0125 functions as a glutamyl aminopeptidase and moonlights at the cell surface as an adhesin that binds glycosaminoglycans and plasminogen (Robinson *et al.*, 2013). MHJ_0125 shows preference for glutamic acid (E), alanine (A) and leucine (L) residues on N-termini but not arginine (R), proline (P), valine (V) or phenylalanine (F) residues (Robinson *et al.*, 2013). Recently we characterised MHJ_0461 as a surface-accessible leucyl aminopeptidase (Jarocki, Raymond & Djordjevic, unpublished data). Given their amino acid preferences, these aminopeptidases are likely to play a role in the sequential removal of amino acids of neo-N-termini generated by endoproteolytic processing including the semi-tryptic peptide series 4, 7 and 10 in Table 5.1 and in Figure 5.7.

Heparin-agarose chromatography of cell lysates of *M. hyopneumoniae* identified P120, P85 and smaller fragments spanning the C-terminus of P85. These data reconfirmed previous studies showing the recombinant fragments spanning the N-terminal 1444 amino acids of P216 expressed in *Escherichia coli* bind heparin and extend these studies by defining heparin-binding domains in the C-terminus of P85 between amino acids 1444-1878. The ability of *M. hyopneumoniae* to bind heparin underpins its ability to bind to porcine cilia (Zhang *et al.*, 1994) and to recognise receptors on the surface of porcine kidney epithelial-like cells (PK15 cells) (Burnett *et al.*, 2006; Deutscher *et al.*, 2010). Consistent with this view, P120 and P85 were recovered from affinity columns using biotinylated PK15 surface proteins as bait. Ligand blotting studies suggest that *M. hyopneumoniae* expresses many proteins that bind heparin (Burnett *et al.*, 2006; Wilton *et al.*, 2009) underscoring the high degree of redundancy in the ability to adhere to and colonise target cells in pigs. Like the other highly expressed adhesins of *M. hyopneumoniae* that are extensively processed (Bogema *et al.*, 2012; Djordjevic *et al.*, 2004; Raymond *et al.*, 2013), these data highlight the importance of endoproteolysis as a means of generating surface diversity and ectodomain shedding in this species.

Chapter 6. Discussion and concluding comments

The genome-reduced swine respiratory pathogen *Mycoplasma hyopneumoniae* presents a significant problem to the agricultural industry that is currently inadequately controlled. It is globally widespread and leaves infected pigs susceptible to porcine enzootic pneumonia and at increased risk of secondary infections. Current control strategies include commercial bacterin vaccines which do not prevent colonization, and antibiotics to control Mycoplasma infections and also combat secondary infections, however heavy antibiotic use can be a driver in the development of antibiotic resistance (Haesebrouck *et al.*, 2004; Jordan *et al.*, 2009; Kolodziejczyk and Pejsak, 2004). An improved understanding of this minimal bacterial pathogen will assist in the development of targeted vaccines and therapeutics.

As *M. hyopneumoniae* is a minimal organism, its genome encodes for fewer than 700 predicted open reading frames. It has lost genes required for cell wall synthesis and many metabolic pathways through a process of degenerative evolution (Razin and Hayflick, 2010; Razin *et al.*, 1998). These features make it an attractive organism for systems-wide analysis of the proteome; without a cell wall to interfere in sample preparation, and a bioinformatically-predicted minimal proteome to catalogue, making data analysis far more accessible. With bioinformatics and high-throughput analysis software becoming increasingly more powerful and accessible, large-scale experiments, even using complex eukaryotic samples, can be rapidly processed (Mann *et al.*, 2013). By examining a minimal genome organism, we can go beyond cataloguing of proteins and move into more detailed characterisation of processes, in a systems-oriented manner. This systems-biology approach is expected to shed light on overall processes underlying interactions of the bacteria with the host, which is particularly relevant in an organism with limited coding capacity.

M. hyopneumoniae is largely incapable of synthesising many nutrients and biomolecules such as amino acids and cholesterols and relies heavily on acquisition of these components from the host (Razin and Hayflick, 2010; Razin *et al.*, 1998). The ability to adhere and bind to the swine respiratory tract is thus essential to bacterial survival and propagation. It is known that *M. hyopneumoniae* surface proteins play a

key role in host cell adhesion, with members of the P97 and P102 paralogous families of adhesins becoming increasingly well-studied in the past decade (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Jenkins *et al.*, 2006; Raymond *et al.*, 2014; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Wilton *et al.*, 2009; Wilton *et al.*, 1998). The paralog families encode upwards of ten large open reading frames, the majority of which occur in two-gene operons (Minion *et al.*, 2004) and are abundant in proteome analyses. It is remarkable that a bacterium with a reduced genome should assign such a large proportion of its coding capacity to encoding functionally redundant proteins; however it is likely to be an evolutionarily conserved trait as adherence is such an essential component to its survival. A key feature of these important proteins is that they are post-translationally processed by proteases, resulting in the generation of multiple, multifunctional fragments (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Jenkins *et al.*, 2006; Raymond *et al.*, 2014; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Wilton *et al.*, 2009; Wilton *et al.*, 1998). These protein fragments have also been shown to bind to host molecules such as extracellular matrix components and interact with the immune system. Such complex regulation of protein expression by proteolytic processing cannot be determined by genomic or transcriptomic approaches. The recent trend of high-throughput proteomics techniques which utilise high-resolution mass spectrometry to analyse complex samples in a peptide-centric manner also fail to detect this kind of regulation. In light of the extensive post-translational processing of the adhesin proteins, multiple approaches, as presented in this work, are required to comprehensively investigate the proteome of *M. hyopneumoniae*.

A major aim of the work presented here was to investigate the expressed, global proteome of *M. hyopneumoniae* to gain better insight into how this genome-reduced organism is able to successfully colonise and thrive within the host. This approach utilised peptide-centric methods such as 2DLC-MS/MS in combination with protein-centric methods such as 1D GeLC-MS/MS and 2D gel analysis to not only catalogue the proteome, but also identify post-translational processing events. These have previously been determined to play an important role in modifying surface proteins

which contribute to pathogenesis (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Burnett *et al.*, 2006; Deutscher *et al.*, 2010; Deutscher *et al.*, 2012; Djordjevic *et al.*, 2004; Jenkins *et al.*, 2006; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2010; Seymour *et al.*, 2011; Seymour *et al.*, 2012; Tacchi *et al.*, 2014; Wilton *et al.*, 2009). The data presented in this thesis breaks new ground in determining the extent of proteolytic processing used by *M. hyopneumoniae* to modify proteins to increase diversity without resorting to genome expansion. It has shown that at least 30 proteins, including but not limited to the P97/P102 adhesin families, are processed and displayed on the cell surface of *M. hyopneumoniae* indicating that proteases have a role in influencing surface protein architecture and potentially contributing to virulence. This work also identified aminopeptidase modifications creating variants of many proteins. Proteolysis, particularly endoproteolysis occurring within a polypeptide chain, may disrupt protein structure and therefore function. If for example, the catalytic or binding domain of a protein is disrupted it can no longer effect that function, however this process may also generate new functional subunits of a protein (such as for the adhesin proteins) which are capable of carrying out another function independently of the pre-protein. Thus, post-translational processing events serve to create multifunctional proteins and increase protein diversity. This is particularly relevant if these proteins or their fragments are presented at the cell surface, where they may interact with a slew of host molecules. While the P97/P102 adhesin proteins are abundant and important in pathogenesis, they do not represent the full extent of interactions of *M. hyopneumoniae* proteins with the host.

The surface proteome analysis revealed that *M. hyopneumoniae* uses another strategy, in addition to proteolytic processing, to increase protein diversity at the cell surface. Proteins with well-characterised functions in the cytoplasm were identified at the cell surface, where they potentially carry out alternative functions. These proteins, if not modified post-translationally, are termed moonlighting proteins. Initially, serendipitous observations led to the identification of moonlighting proteins (Jeffery, 1999), however in the past decade they have gained notoriety, and while they not necessarily easier to identify than in the past, their moonlighting functions are now more readily accepted and the body of literature citing moonlighting proteins is growing (Henderson and Martin, 2014). Many of the moonlighting proteins identified in *M. hyopneumoniae*

presented here are glycolytic enzymes and their putative moonlighting functions include adherence to surface proteins of host epithelial cells, glycosaminoglycans, plasminogen, fibronectin or actin. The ability of bacterial proteins to bind this array of host molecules is associated with colonisation, cell invasion and evasion of the host immune system (Pancholi and Chhatwal, 2003). A number of these putative moonlighting proteins were also shown to be immunogenic. Immunogenicity is a good indicator of expression, secretion and surface localisation *in vivo*, as these proteins are likely to be exposed in order to be recognised by the host immune system (Cacciotto *et al.*, 2010). Lactate dehydrogenase, pyruvate dehydrogenase complex components, elongation factor Tu, chaperone protein DnaK, triosephosphate isomerase and glyceraldehyde phosphate dehydrogenase are notable examples of multifunctional or moonlighting proteins which have been identified here, as well as in other bacteria, including other *Mycoplasma* species (Dallo *et al.*, 2002), *Bacillus* species (Commichau *et al*., 2009; Matta *et al*., 2010), *Streptococcus* (Mohan *et al.*, 2014), *Leptospira* (Wolff *et al.*, 2013), and also in parasites (Gomez-Arreaza *et al.*, 2014; Karkowska-Kuleta and Kozik, 2014), yeasts and fungi (Gancedo and Flores, 2008; Ikeda and Ichikawa, 2014), and mammalian cells (Petit *et al.*, 2014; Sasikumar *et al*., 2012). In addition, these proteins were also identified to be surface-exposed in multiple strains of *M. hyopneumoniae*, suggesting their moonlighting activities may also be conserved between strains. As these proteins are well-conserved amongst bacterial species, these may represent attractive targets for vaccine or therapeutic development that could be applicable to a range of bacterial pathogens. Notably, this approach is novel in that it is the first to describe the presence of multiple putative moonlighting proteins in the one pathogen underscoring the potential importance of moonlighting and protein multifunctionality in microbial pathogenesis.

Some of the identified proteins with well-characterised functions in the cytoplasm were also identified to be proteolytically processed. While we cannot exclude the possibility that processing occurs in the cell cytosol, this is unlikely because cleavage is expected to impact protein folding and destroy enzymatic function. As such it is hypothesised that cleavage is confined to the cell surface where cleavage fragments undertake novel roles in pathogenesis and immune evasion. At this stage it remains unclear if these protein fragments are expressed exclusively at the cell surface or also within the cytosol,

however many fragments display the capacity to bind to host components such as glycosaminoglycans, fibronectin, plasminogen, actin and epithelial cell surface proteins, which is a strong indication of novel functional surface exposure. Cleavage patterns of Elongation factor Tu were examined in Chapter 3, where three internal cleavage sites were identified by N-terminal labelling or identification of semi-tryptic peptides. This data is consistent with EfTu being a target of several cleavage events on the cell surface of *M. hyopneumoniae* and suggests that EfTu functions as an adhesin with discrete binding domains for a range of host molecules. This further emphasises the importance of initial adherence events which drives *M. hyopneumoniae* to display a diverse array of proteins on the cell surface.

Amongst the most commonly identified proteins in the surfaceome and global proteome analyses was a P97 adhesin paralog, P216. Undertaking a detailed analysis of this molecule has allowed for the elucidation of the full gamut of endoproteolytically-derived minor cleavage fragments of the P216 pre-protein. The data presented in Chapter 5 provides evidence that the information content of P216 is significantly underestimated. In light of the combined evidence provided throughout the thesis, this is likely true of all the P97 and P102 paralogous families of adhesins, and also of many other proteins of *M. hyopneumoniae*. Considering its limited coding capacity, a significant proportion of the *M. hyopneumoniae* genome is dedicated to encoding large ORFS, with 54 (8%) encoding ORFs >800 amino acids in length (~90-100 kDa). With cleavage such a prominent feature of the proteome, it is reasonable to speculate that proteolytic processing may in fact be more energetically favourable than genome expansion to increase coding capacity. The analysis of P216 revealed that processing was more extensive than previously identified, with variable efficiency of cleavage at multiple motifs throughout the protein. The previously identified dominant cleavage motif, S/T-X-F↓X-D/E was identified in five locations within the protein; however only one of these sites was determined to be a *bone fide* cleavage site. The distinguishing feature of this site was its presence within a region of intrinsic protein disorder, which is likely to contribute to the flexibility of this region and the overall accessibility of the motif to the protease responsible for cleavage. This motif was also determined to be a site of efficient cleavage, which is consistent with previous findings in other adhesins and adhesin-related proteins (Bogema *et al.*, 2012; Bogema *et al.*, 2011; Deutscher *et*

al., 2012; Raymond *et al.*, 2014; Raymond *et al.*, 2013; Seymour *et al.*, 2012; Tacchi *et al.*, 2014). In addition to the dominant cleavage motif, several other sites of less efficient cleavage were identified, suggesting that multiple proteases are involved in this processing. Furthermore, aminopeptidase activity was detected using N-terminal labelling and through the detection of semi-tryptic peptides. This is an important discovery as it suggests that fragments of P216 may be variably displayed at the cell surface through the regulation of proteolysis. This is reminiscent of the effect, although not the mechanism, of variable surface antigen expression or antigenic variation seen in other bacteria such as *Neisseria*, *Streptococcus*, *Mycobacterium* (Banu *et al*., 2002) and other Mycoplasmas (Beier *et al*., 1998; Hermeyer *et al*., 2012), and more commonly in parasites (Giha *et al*., 2000; Pearson *et al*., 1981; Thai and Forney, 2000), which also share common features of persistent chronic infections. By targeting and inhibiting proteolysis, we may be able to disrupt these pathogenic mechanisms to reduce the adhesive capacity, and possibly the infective capacity of these bacteria. Unpublished studies performed in the Djordjevic laboratory have determined that *M. hyopneumoniae* is not viable when grown under normal laboratory culture conditions in the presence of broad-spectrum protease inhibitors. This finding is not surprising as proteases play important roles in the maintenance of normal cell homeostasis, however the ability to identify which specific protease(s) are essential have until very recently been limited to characterisation of individual proteins by traditional biochemical methods (Jarocki *et al.*, 2014).

The use of N-terminal labelling was intended to provide a method of high-throughput screening for true and neo N-termini, resulting from *in vivo* cleavage by proteases. This information would elucidate cleavage motifs which in turn could assist in the future determination of protease identities by biochemical characterisation of recombinant protease activity. Unfortunately, the high-throughput mapping of cleavage motifs using N-terminal dimethylation in *Mycoplasma* did not provide complete, definitive information as the labelling strategy was ineffective compared to model mammalian counterparts (Kleifeld *et al.*, 2011). This is because the process of dimethyl labelling requires a blocking step, whereby primary amines on lysine residues throughout the proteins are blocked by dimethylation prior to digestion with trypsin to create new, reactive N-termini. The internal tryptic peptides are then removed from the sample by reaction with an amine-reactive polymer and unlabelled peptides with N-termini

generated *in vivo* are enriched. *Mycoplasma hyopneumoniae* has a lysine-rich proteome (~11% of the predicted proteome), with comparably fewer arginine residues (3% of the predicted proteome). Following the blocking step, trypsin can only cleave at arginine residues, as lysines are blocked, creating large, multiply-labelled tryptic peptides, which are not readily detectable by LC-MS/MS analyses such as those performed here. In some cases labelled peptides can exceed 150 amino acids in length. As such, not all of the cleavage sites identified here were validated by labelling methods and were supplemented by the identification of semi-tryptic peptides. Future studies will be improved by refining this enrichment and detection process to obtain a more definitive and reliable identification of cleavage sites. The use of alternative enzymes that do not cleave at lysine or arginine residues may provide a simple solution to the generation of labelled N-terminal peptides with more reliable sizes and ionisation capacity. The most obvious candidate would be endoprotease GluC which cleaves peptide bonds C-terminal to glutamic acid and is a common enzyme used to generate peptides in proteomics (Drapeau *et al*., 1972). The drawback to using this enzyme however is that glutamic acid often forms part of the dominant cleavage motif in the adhesins, and this would result in the proximal digestion of the N-terminal sequence to less than 5 amino acids, and make identification of the labelled N-terminal impossible by the methods used here. The use of several complementary enzymes may still be a viable solution to this issue for future studies, as well as potentially providing better coverage of the global proteome (Meyer *et al*., 2011; Schilling and Overall, 2008). Ultimately, with the technologies currently available, a combination of N-terminal labelling or enrichment techniques, protein-centric approaches to retain mass context of protein fragments and the critical examination of semi-tryptic peptides will provide the most comprehensive information regarding proteolytic processing.

Throughout the chapters presented in this thesis, it was found that *Mycoplasma hyopneumoniae* uses a diverse range of tactics to compensate for a reduced coding capacity, without resorting to genome expansion. Since the completion of the Human Genome Project, it has been known that genome size does not directly correlate with phenotypic complexity of an organism. Diverse organisms such as worm, fly, chicken and human have comparable genome sizes, yet display significant discrepancy in apparent phenotypic complexity. Post-transcriptional regulation and gene splicing may

account for some of this; however post-translational modifications, including proteolytic processing may introduce another level of complexity that cannot be predicted by the genome or transcriptome (Lange and Overall, 2013). Serendipitous discovery of processing in key proteins of some pathogenic bacteria have previously been described, however these kinds of post-translational modifications are not well-understood or predictable and it is not yet clear if these processes occur in other organisms. With renewed focus on genome and RNA analysis due to the increasing accessibility of next-generation sequencing, these processes may be continue to be overlooked. As such, this needs to be taken into account when undertaking global proteomic investigations.

A striking recurring theme throughout this thesis is the need for protein-centric methods to enable comprehensive characterisation of the proteome that can detect post-translational modifications including proteolytic cleavage. The experimental design used here had to take into account the previous reports of biologically-significant adhesin proteins undergoing post-translational processing. Thus, we relied heavily on protein-centric, GeLC-MS/MS approaches to identify protein cleavage fragments. However, if the sole aim of this project was maximum coverage of the proteome, more high-throughput methodologies would likely have been used. This is also largely the case in other organisms, where no evidence of processing has been previously detected, as there are few apparent benefits, and far greater apparent drawbacks to using gel-based proteomics in a global analysis, although reports on the limitations of 2D gels vary greatly (Bodzon-Kulakowska *et al*., 2007; Gorg *et al*., 2009; Meyer *et al.*, 2011; Oliveira *et al*., 2014). The rapid developments in mass spectrometry, increasing mass accuracy, sensitivity and throughput have seen peptide-centric, "bottom-up" approaches using 2D LC-MS techniques such as MudPIT, become the methods of choice for large-scale high throughput analysis of proteomes (Porteus *et al*., 2011). If a complete proteome were to be analysed by a high-throughput bottom-up approach, only one representative protein for any expressed gene would be identified in most cases (Meyer *et al.*, 2011). Unless a peptide which carries a distinguishing modification, known as a "proteotypic" peptide is identified, or in the case of proteolytic cleavage, if a peptide displays atypical digestion signature (eg. semi-tryptic peptides) then the modified protein will not be identified by these peptide-centric methods. Unfortunately, gel-based

proteomics, particularly 2D gels are perceived to be more technically challenging, associated with greater protein losses and solubility issues than peptide-centric approaches, however this technology is still being utilised by those with an interest in molecular mechanisms and disease processes (Oliveira *et al.*, 2014).

The ultimate form of protein-centric analyses would have to be considered "top-down" mass spectrometry, which provides mass spectrometric characterisation of intact proteins (Kelleher, 2004). Ideally, this technique allows for 100% sequence coverage and full characterization of proteoforms, the specific molecular form of the protein resulting from combinations of genetic variation, alternative splicing, and post-translational modifications (Catherman *et al.,* 2014). However, the technical difficulty of proteome-wide analysis at the intact protein level has caused top down proteomics to lag behind peptide-centric approaches in terms of proteome coverage, sensitivity, and throughput (Catherman *et al.,* 2014). These methods are currently limited by methods of protein solubilisation which are incompatible with electrospray ionisation required for mass spectrometry. Recent developments in sample purification post-solubilisation however, have increased the practicality of such analyses (Kim *et al.,* 2015) and these methods are now promising for use in the characterisation of the full extent of cleavage products produced by *M. hyopneumoniae*.

It is becoming more apparent that comprehensive characterisation of any proteome requires at least some element of protein-centric analysis to be considered "comprehensive" and not simply "cataloguing" (Hoehenwarter *et al.,* 2006). The time invested in deeper analysis of the proteome by protein-centric methods will ultimately provide richer information that can be later used in targeted quantitative proteomics (Oliveira *et al.*, 2014). Further to this, subproteome and functional analyses are also required for the comprehensive characterisation of the proteome, in order to classify potential moonlighting proteins which cannot be determined by identification of the protein from the global proteome alone (Hoehenwarter *et al.*, 2006; Meyer *et al.*, 2011). Functional analyses can also provide clues as to which proteins are essential for various processes, including (as described here) various features of pathogenesis.

Large scale functional analysis at the genome level has generally been determined through the generation of random insertional mutants by transposon mutagenesis

however, Mycoplasma species are relatively refractory to genetic manipulation. This inability to genetically manipulate *M. hyopneumoniae* has been a major obstacle to progress in identifying which genes are essential for survival and pathogenesis of the organism and also genes with functions in cellular invasion and biofilm formation. As such, the ability to generate random transposon mutants of *M. hyopneumoniae* and screen them in a high-throughput manner using PCR has only recently been acquired following extensive optimisation (Maglennon *et al.,* 2013). However this has already allowed for the compilation of a pool of over 10 000 mutants (Maglennon *et al.*, 2013), which will enable the evaluation of the effect of inactivated genes on bacterial processes, and the identification of attenuated mutants in *in vivo* screens (Raymond, unpublished data). Recently published work by Mobegi *et al.* (2014) has provided a proof-of-concept study, using high-density transposon mutagenesis, high-throughput sequencing and integrative genomics to discover potential drug targets at the genome level in *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Moraxella catarrhalis*. Through identification of essential genes in bacterial pathogens, and comparison to the host genome and a genes from commensal host microbiota, potential drug targets could be selected that would be unlikely to produce off target effects in the host or host native microflora (Mobegi *et al.*, 2014). The work presented in this thesis will greatly assist in guiding this future research in narrowing down vaccine and/or therapeutic candidates and better elucidating the functions of essential and non-essential genes.

Collectively, the observations presented in this thesis highlight the complex nature of the proteome in a genome-reduced organism. The findings presented here have indicated several avenues for future research, building towards the development of novel vaccines and therapeutics for the prevention and treatment of *M. hyopneumoniae* infection. It is clear that other, more complex, microbial pathogens employ similar strategies such as moonlighting (which has been identified in all three kingdoms of life) and proteolytic processing to increase protein diversity, particularly at the cell surface for the purposes of host-binding, invasion or immune evasion. Importantly, these findings have only been elucidated through non-targeted, systems-wide approaches that were function-focused and unbiased towards *in silico* predictions. If these phenomena are widespread in nature, the ability to identify and characterise them using a minimal

model organism may greatly improve our understanding, not just of these processes in *Mycoplasma*, but of biology as a whole.

Bibliography

1. Adams, C., Pitzer, J., and Minion, F.C. (2005). In vivo expression analysis of the P97 and P102 paralog families of *Mycoplasma hyopneumoniae*. *Infect Immun* 73, 7784-7787.
2. Agarwal, S., Kulshreshtha, P., Bambah Mukku, D., and Bhatnagar, R. (2008). alpha-Enolase binds to human plasminogen on the surface of *Bacillus anthracis*. *Biochim Biophys Acta* 1784, 986-994.
3. Alderete, J.F., Millsap, K.W., Lehker, M.W., and Benchimol, M. (2001). Enzymes on microbial pathogens and *Trichomonas vaginalis*: molecular mimicry and functional diversity. *Cell Microbiol* 3, 359-370.
4. Ansong, C., Tolic, N., Purvine, S.O., Porwollik, S., Jones, M., Yoon, H., Payne, S.H., Martin, J.L., Burnet, M.C., Monroe, M.E.*, et al.* (2011). Experimental annotation of post-translational features and translated coding regions in the pathogen Salmonella Typhimurium. *BMC Genomics* 12, 433.
5. *Quantity of antimicrobial products sold for veterinary use in Australia 1999/2000-2001/2002*. (2005) Australian Pesticides & Veterinary Medicines Authority (APVMA). Australia.
6. Archambaud, C., Gouin, E., Pizarro-Cerda, J., Cossart, P., and Dussurget, O. (2005). Translation elongation factor EF-Tu is a target for Stp, a serine-threonine phosphatase involved in virulence of *Listeria monocytogenes. Mol Microbiol* 56, 383-396.
7. auf dem Keller, U., and Schilling, O. (2010). Proteomic techniques and activity-based probes for the system-wide study of proteolysis. *Biochimie* 92, 1705-1714.
8. Babady, N.E., Pang, Y.P., Elpeleg, O., and Isaya, G. (2007). Cryptic proteolytic activity of dihydrolipoamide dehydrogenase. *Proc Natl Acad Sci U S A* 104, 6158-6163.
9. Balasubramanian, S., Kannan, T.R., and Baseman, J.B. (2008). The surface-exposed carboxyl region of *Mycoplasma pneumoniae* elongation factor Tu interacts with fibronectin. *Infect Immun* 76, 3116-3123.
10. Balish, M.F. (2014). *Mycoplasma pneumoniae*, an underutilized model for bacterial cell biology. *J Bacteriol* 196, 3675-3682.
11. Banu, S., Honore, N., Saint-Joanis, B., Philpott, D., Prevost, M.C., and Cole, S.T. (2002). Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol Microbiol* 44, 9-19.
12. Barel, M., Hovanessian, A.G., Meibom, K., Briand, J.P., Dupuis, M., and Charbit, A. (2008). A novel receptor - ligand pathway for entry of *Francisella tularensis* in monocyte-like THP-1 cells: interaction between surface nucleolin and bacterial elongation factor Tu. *BMC Microbiol* 8, 145.
13. Barton, M.D. (2014). Impact of antibiotic use in the swine industry. *Curr Opin Microbiol* 19, 9-15.
14. Beier, T., Hotzel, H., Lysnyansky, I., Grajetzki, C., Heller, M., Rabeling, B., Yogev, D., and Sachse, K. (1998). Intraspecies polymorphism of vsp genes and expression profiles of variable surface protein antigens (Vsps) in field isolates of *Mycoplasma bovis. Vet Microbiol* 63, 189-203.
15. Bendtsen, J.D., Kiemer, L., Fausboll, A., and Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiol* 5, 58.
16. Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340, 783-795.
17. Bentley, S., Sebaihia, M., and Crossman, L. (2005). Enemies within. *Nat Rev Micro* 3, 8-9.

18. Bereiter, M., Young, T.F., Joo, H.S., and Ross, R.F. (1990). Evaluation of the ELISA and comparison to the complement fixation test and radial immunodiffusion enzyme assay for detection of antibodies against *Mycoplasma hyopneumoniae* in swine serum. *Vet Microbiol* 25, 177-192.
19. Bergmann, S., and Hammerschmidt, S. (2007). Fibrinolysis and host response in bacterial infections. *Thromb Haemost* 98, 512-520.
20. Bergmann, S., Rohde, M., Chhatwal, G.S., and Hammerschmidt, S. (2001). alpha-Enolase of *Streptococcus pneumoniae* is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. *Mol Microbiol* 40, 1273-1287.
21. Bergmann, S., Rohde, M., and Hammerschmidt, S. (2004). Glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus pneumoniae* is a surface-displayed plasminogen-binding protein. *Infect Immun* 72, 2416-2419.
22. Bergmann, S., Schoenen, H., and Hammerschmidt, S. (2013). The interaction between bacterial enolase and plasminogen promotes adherence of *Streptococcus pneumoniae* to epithelial and endothelial cells. *Int J Med Microbiol* 303, 452-462.
23. Berven, F.S., Karlsen, O.A., Straume, A.H., Flikka, K., Murrell, J.C., Fjellbirkeland, A., Lillehaug, J.R., Eidhammer, I., and Jensen, H.B. (2006). Analysing the outer membrane subproteome of *Methylococcus capsulatus* (Bath) using proteomics and novel biocomputing tools. *Arch Microbiol* 184, 362-377.
24. Bledi, Y., Inberg, A., and Linial, M. (2003). PROCEED: A proteomic method for analysing plasma membrane proteins in living mammalian cells. *Brief Funct Genomic Proteomic* 2, 254-265.
25. Bodzon-Kulakowska, A., Bierczynska-Krzysik, A., Dylag, T., Drabik, A., Suder, P., Noga, M., Jarzebinska, J., and Silberring, J. (2007). Methods for samples preparation in proteomic research. *J Chromatogr B Analyt Technol Biomed Life Sci* 849, 1-31.
26. Boel, G., Jin, H., and Pancholi, V. (2005). Inhibition of cell surface export of group A streptococcal anchorless surface dehydrogenase affects bacterial adherence and antiphagocytic properties. *Infect Immun* 73, 6237-6248.
27. Bogema, D.R., Deutscher, A.T., Woolley, L.K., Seymour, L.M., Raymond, B.B., Tacchi, J.L., Padula, M.P., Dixon, N.E., Minion, F.C., Jenkins, C*., et al.* (2012). Characterization of cleavage events in the multifunctional cilium adhesin Mhp684 (P146) reveals a mechanism by which *Mycoplasma hyopneumoniae* regulates surface topography. *mBio* 3.
28. Bogema, D.R., Scott, N.E., Padula, M.P., Tacchi, J.L., Raymond, B.B., Jenkins, C., Cordwell, S.J., Minion, F.C., Walker, M.J., and Djordjevic, S.P. (2011). Sequence TTKF↓QE defines the site of proteolytic cleavage in Mhp683 protein, a novel glycosaminoglycan and cilium adhesin of *Mycoplasma hyopneumoniae*. *J Biol Chem* 286, 41217-41229.
29. Boone, T.J., Burnham, C.A., and Tyrrell, G.J. (2011). Binding of group B streptococcal phosphoglycerate kinase to plasminogen and actin. *Microb Pathog* 51, 255-261.
30. Bordier, C. (1981). Phase separation of integral membrane proteins in Triton X-114 solution. *J Biol Chem* 256, 1604-1607.
31. Browning, G.F., Marenda, M.S., Noormohammadi, A.H., and Markham, P.F. (2011). The central role of lipoproteins in the pathogenesis of mycoplasmoses. *Vet Microbiol* 153, 44-50.
32. Brune, D., Denslow, N.D., Kobayashi, R., Lane, W.S., Leone, J.W., Madden, B.J., Neveu, J.M., and Pohl, J. (2006). ABRF ESRG 2005 study: identification of seven modified amino acids by Edman sequencing. *J Biomol Tech* 17, 308-326.

33. Burnett, T.A., Dinkla, K., Rohde, M., Chhatwal, G.S., Uphoff, C., Srivastava, M., Cordwell, S.J., Geary, S., Liao, X., Minion, F.C.*, et al.* (2006). P159 is a proteolytically processed, surface adhesin of *Mycoplasma hyopneumoniae*: defined domains of P159 bind heparin and promote adherence to eukaryote cells. *Mol Microbiol* 60, 669-686.
34. Cacciotto, C., Addis, M.F., Pagnozzi, D., Chessa, B., Coradduzza, E., Carcangiu, L., Uzzau, S., Alberti, A., and Pittau, M. (2010). The liposoluble proteome of *Mycoplasma agalactiae*: an insight into the minimal protein complement of a bacterial membrane. *BMC Microbiol* 10, 225.
35. Candela, M., Fiori, J., Dipalo, S., and Brigidi, P. (2010). Development of a high-performance affinity chromatography-based method to study the biological interaction between whole micro-organisms and target proteins. *Lett Appl Microbiol* 51, 678-682.
36. Cardin, A.D., and Weintraub, H.J. (1989). Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* 9, 21-32.
37. Carmo, A.A., Costa, B.R., Vago, J.P., de Oliveira, L.C., Tavares, L.P., Nogueira, C.R., Ribeiro, A.L., Garcia, C.C., Barbosa, A.S., Brasil, B.S.*, et al.* (2014). Plasmin Induces In Vivo Monocyte Recruitment through Protease-Activated Receptor-1-, MEK/ERK-, and CCR2-Mediated Signaling. *J Immunol* 193, 3654-3663.
38. Carsons, S.E. (1989). *Fibronectin in Health and Disease* (Florida: CRC Press, Inc.).
39. Catherman, A.D., Skinner, O.S., and Kelleher, N.L. (2014). Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun* 445, 683-693.
40. Chen, H., Yu, S., Shen, X., Chen, D., Qiu, X., Song, C., and Ding, C. (2011). The *Mycoplasma gallisepticum* alpha-enolase is cell surface-exposed and mediates adherence by binding to chicken plasminogen. *Microb Pathog*.
41. Chhatwal, G.S. (2002). Anchorless adhesins and invasins of Gram-positive bacteria: a new class of virulence factors. *Trends Microbiol* 10, 205-208.
42. Clark, L.K., Armstrong, C.H., Freeman, M.J., Scheidt, A.B., Sands-Freeman, L., and Knox, K. (1991). Investigating the transmission of *Mycoplasma hyopneumoniae* in a swine herd with enzootic pneumonia. *Vet Med* 86, 543-550.
43. Cole, J.N., Ramirez, R.D., Currie, B.J., Cordwell, S.J., Djordjevic, S.P., and Walker, M.J. (2005). Surface analyses and immune reactivities of major cell wall-associated proteins of group a streptococcus. *Infect Immun* 73, 3137-3146.
44. Cole, S.R., Ashman, L.K., and Ey, P.L. (1987). Biotinylation: an alternative to radioiodination for the identification of cell surface antigens in immunoprecipitates. *Mol Immunol* 24, 699-705.
45. Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem Sci* 25, 147-150.
46. Commichau, F.M., Rothe, F.M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., Hammer, E., Volker, U., and Stulke, J. (2009). Novel Activities of Glycolytic Enzymes in Bacillus subtilis. *Mol Cell Proteomics* 8, 1350-1360.
47. Cordwell, S.J. (2006). Technologies for bacterial surface proteomics. *Curr Opin Microbiol* 9, 320-329.
48. Cordwell, S.J., and Thingholm, T.E. (2010). Technologies for plasma membrane proteomics. *Proteomics* 10, 611-627.
49. Coutte, L., Alonso, S., Reveneau, N., Willery, E., Quatannens, B., Locht, C., and Jacob-Dubuisson, F. (2003). Role of adhesin release for mucosal colonization by a bacterial pathogen. *J Exp Med* 197, 735-742.
50. Coutte, L., Antoine, R., Drobecq, H., Locht, C., and Jacob-Dubuisson, F. (2001). Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. *EMBO J* 20, 5040-5048.

51. Cromwell, G.L. (2002). Why and how antibiotics are used in swine production. *Animal Biotech* 13, 7-27.
52. Dallo, S.F., Kannan, T.R., Blaylock, M.W., and Baseman, J.B. (2002). Elongation factor Tu and E1 beta subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in *Mycoplasma pneumoniae*. *Mol Microbiol* 46, 1041-1051.
53. Davis, K.L., and Wise, K.S. (2002). Site-specific proteolysis of the MALP-404 lipoprotein determines the release of a soluble selective lipoprotein-associated motif-containing fragment and alteration of the surface phenotype of *Mycoplasma fermentans*. *Infect Immun* 70, 1129-1135.
54. de Castro, L.A., Rodrigues Pedroso, T., Kuchiishi, S.S., Ramenzoni, M., Kich, J.D., Zaha, A., Henning Vainstein, M., and Bunselmeyer Ferreira, H. (2006). Variable number of tandem amino acid repeats in adhesion-related CDS products in *Mycoplasma hyopneumoniae* strains. *Vet Microbiol* 116, 258-269.
55. de Miguel, N., Lustig, G., Twu, O., Chattopadhyay, A., Wohlschlegel, J.A., and Johnson, P.J. (2010). Proteome analysis of the surface of *Trichomonas vaginalis* reveals novel proteins and strain-dependent differential expression. *Mol Cell Proteomics* 9, 1554-1566.
56. De Petro, G., Tavian, D., Marchina, E., and Barlati, S. (2002). Induction of fibronectin mRNA by urokinase- and tissue-type plasminogen activator in human skin fibroblasts: differential role of u-PA and t-PA at the fibronectin protein level. *Biol Chem* 383, 177-187.
57. DeBey, M.C., and Ross, R.F. (1994). Ciliostasis and loss of cilia induced by *Mycoplasma hyopneumoniae* in porcine tracheal organ cultures. *Infect Immun* 62, 5312-5318.
58. Demina, I.A., Serebryakova, M.V., Ladygina, V.G., Rogova, M.A., Zgoda, V.G., Korzhenevskyi, D.A., and Govorun, V.M. (2009). Proteome of the bacterium *Mycoplasma gallisepticum*. *Biochem-Moscow* 74, 165-174.
59. Dempewolf, C., Morris, J., Chopra, M., Jayanthi, S., Kumar, T.K.S., Li, W.N., and Ieee (2013). *Identification of Consensus Glycosaminoglycan Binding Strings in Proteins*. In *2013 International Conference on Information Science and Applications* (New York: Ieee).
60. Deutscher, A.T., Jenkins, C., Minion, F.C., Seymour, L.M., Padula, M.P., Dixon, N.E., Walker, M.J., and Djordjevic, S.P. (2010). Repeat regions R1 and R2 in the P97 paralogue Mhp271 of *Mycoplasma hyopneumoniae* bind heparin, fibronectin and porcine cilia. *Mol Microbiol* 78, 444-458.
61. Deutscher, A.T., Tacchi, J.L., Minion, F.C., Padula, M.P., Crossett, B., Bogema, D.R., Jenkins, C., Kuit, T.A., Walker, M.J., and Djordjevic, S.P. (2012). *Mycoplasma hyopneumoniae* Surface proteins Mhp385 and Mhp384 bind host cilia and glycosaminoglycans and are endoproteolytically processed by proteases that recognize different cleavage motifs. *J Proteome Res* 11, 1924-1936.
62. Djordjevic, S.P., Cordwell, S.J., Djordjevic, M.A., Wilton, J., and Minion, F.C. (2004). Proteolytic processing of the *Mycoplasma hyopneumoniae* cilium adhesin. *Infect Immun* 72, 2791-2802.
63. Djordjevic, S.P., Eamens, G.J., Romalis, L.F., Nicholls, P.J., Taylor, V., and Chin, J. (1997). Serum and mucosal antibody responses and protection in pigs vaccinated against *Mycoplasma hyopneumoniae* with vaccines containing a denatured membrane antigen pool and adjuvant. *Aust Vet J* 75, 504-511.
64. Djordjevic, S.P., Eamens, G.J., Romalis, L.F., and Saunders, M.M. (1994). An improved enzyme-linked-immunosorbent-assay (ELISA) for the detection of porcine serum antibodies against *Mycoplasma hyopneumoniae*. *Vet Microbiol* 39, 261-273.

65. Djordjevic, S.P., and Tacchi, J.L. (2014). *Posttranslational Modifcation of Proteins in the Mollicutes*. In *Mollicutes: Molecular Biology and Pathogenesis*, G.F. Browning, and C. Citti, eds. (UK: Caister Academic Press), pp. 91-106.
66. Doucet, A., Butler, G.S., Rodriguez, D., Prudova, A., and Overall, C.M. (2008). Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. *Mol Cell Proteomics* 7, 1925-1951.
67. Doucet, A., and Overall, C.M. (2008). Protease proteomics: revealing protease in vivo functions using systems biology approaches. *Mol Aspects Med* 29, 339-358.
68. Doucet, A., and Overall, C.M. (2011). Broad coverage identification of multiple proteolytic cleavage site sequences in complex high molecular weight proteins using quantitative proteomics as a complement to edman sequencing. *Mol Cell Proteomics* 10, M110 003533.
69. Drapeau, G.R., Boily, Y., and Houmard, J. (1972). Purification and properties of an extracellular protease of *Staphylococcus aureus*. *J Biol Chem* 247, 6720-6726.
70. Dreisbach, A., Hempel, K., Buist, G., Hecker, M., Becher, D., and van Dijl, J.M. (2010). Profiling the surfacome of *Staphylococcus aureus*. *Proteomics* 10, 3082-3096.
71. Dreisbach, A., van Dijl, J.M., and Buist, G. (2011). The cell surface proteome of *Staphylococcus aureus*. *Proteomics* 11, 3154-3168.
72. Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M., and Uversky, V.N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* 272, 5129-5148.
73. Dybvig, K., and Voelker, L.L. (1996). Molecular biology of mycoplasmas. *Annu Rev Microbiol* 50, 25-57.
74. Edman, M., Jarhede, T., Sjostrom, M., and Wieslander, A. (1999). Different sequence patterns in signal peptides from mycoplasmas, other gram-positive bacteria, and *Escherichia coli*: a multivariate data analysis. *Proteins* 35, 195-205.
75. Edwards, J.A., Groathouse, N.A., and Boitano, S. (2005). *Bordetella bronchiseptica* adherence to cilia is mediated by multiple adhesin factors and blocked by surfactant protein A. *Infect Immun* 73, 3618-3626.
76. Elia, G. (2008). Biotinylation reagents for the study of cell surface proteins. *Proteomics* 8, 4012-4024.
77. Erlinger, R. (1995). Glycosaminoglycans in porcine lung: an ultrastructural study using cupromeronic blue. *Cell Tissue Res* 281, 473-483.
78. Escobar, J., Van Alstine, W.G., Baker, D.H., and Johnson, R.W. (2002). Growth performance and whole-body composition of pigs experimentally infected with *Mycoplasma hyopneumoniae*. *J Anim Sci* 80, 384-391.
79. Eshel, G., Shepon, A., Makov, T., and Milo, R. (2014). Land, irrigation water, greenhouse gas, and reactive nitrogen burdens of meat, eggs, and dairy production in the United States. *Proc Natl Acad Sci U S A* 111, 11996-12001.
80. Fagan, P.K., Djordjevic, S.P., Chin, J., Eamens, G.J., and Walker, M.J. (1997). Oral immunization of mice with attenuated *Salmonella typhimurium* aroA expressing a recombinant *Mycoplasma hyopneumoniae* antigen (NrdF). *Infect Immun* 65, 2502-2507.
81. Fagan, P.K., Djordjevic, S.P., Eamens, G.J., Chin, J., and Walker, M.J. (1996). Molecular characterization of a ribonucleotide reductase (nrdF) gene fragment of *Mycoplasma hyopneumoniae* and assessment of the recombinant product as an experimental vaccine for enzootic pneumonia. *Infect Immun* 64, 1060-1064.
82. Fagan, P.K., Walker, M.J., Chin, J., Eamens, G.J., and Djordjevic, S.P. (2001). Oral immunization of swine with attenuated *Salmonella typhimurium* aroA SL3261 expressing a recombinant antigen of *Mycoplasma hyopneumoniae* (NrdF) primes the

immune system for a NrdF specific secretory IgA response in the lungs. *Microb Pathog* 30, 101-110.
83. Faldynova, M., Videnska, P., Havlickova, H., Sisak, F., Juricova, H., Babak, V., Steinhauser, L., and Rychlik, I. (2013). Prevalence of antibiotic resistance genes in faecal samples from cattle, pigs and poultry. *Vet Med* 58, 298-304.
84. Ferrer-Navarro, M., Gomez, A., Yanes, O., Planell, R., Aviles, F.X., Pinol, J., Pons, J.A.P., and Querol, E. (2006). Proteome of the bacterium *Mycoplasma penetrans*. *J Proteome Res* 5, 688-694.
85. Ferron, F., Longhi, S., Canard, B., and Karlin, D. (2006). A practical overview of protein disorder prediction methods. *Proteins* 65, 1-14.
86. Fisunov, G.Y., Alexeev, D.G., Bazaleev, N.A., Ladygina, V.G., Galyamina, M.A., Kondratov, I.G., Zhukova, N.A., Serebryakova, M.V., Demina, I.A., and Govorun, V.M. (2011). Core proteome of the minimal cell: comparative proteomics of three mollicute species. *PLoS ONE* 6, e21964.
87. Flintoft, L. (2009). Systems Biology: Small genome, complex regulation. *Nat Rev Genet* 11, 3-3.
88. Foster, T.J., Geoghegan, J.A., Ganesh, V.K., and Hook, M. (2014). Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nat Rev Microbiol* 12, 49-62.
89. Foulston, L., Elsholz, A.K., DeFrancesco, A.S., and Losick, R. (2014). The extracellular matrix of *Staphylococcus aureus* biofilms comprises cytoplasmic proteins that associate with the cell surface in response to decreasing pH. *mBio* 5, e01667-01614.
90. Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M.*, et al.* (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397-403.
91. Friis, N.F. (1975). Some recommendations concerning primary isolation of Mycoplasma suipneumoniae and *Mycoplasma flocculare* a survey. *Nord Vet Med* 27, 337-339.
92. Fulde, M., Bernardo-Garcia, N., Rohde, M., Nachtigall, N., Frank, R., Preissner, K.T., Klett, J., Morreale, A., Chhatwal, G.S., Hermoso, J.A.*, et al.* (2014). Pneumococcal phosphoglycerate kinase interacts with plasminogen and its tissue activator. *Thromb Haemost* 111, 401-416.
93. Furuya, H., and Ikeda, R. (2011). Interaction of triosephosphate isomerase from *Staphylococcus aureus* with plasminogen. *Microbiol Immunol* 55, 855-862.
94. Gancedo, C., and Flores, C.L. (2008). Moonlighting proteins in Yeasts. *Microbiol Mol Biol Rev* 72, 197.
95. Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O., and Hutchison, C.A., 3rd (2008). One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A* 105, 20404-20409.
96. Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M.*, et al.* (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52-56.
97. Giha, H.A., Staalsoe, T., Dodoo, D., Roper, C., Satti, G.M., Arnot, D.E., Hviid, L., and Theander, T.G. (2000). Antibodies to variable *Plasmodium falciparum*-infected erythrocyte surface antigens are associated with protection from novel malaria infections. *Immunol Lett* 71, 117-126.
98. Gillings, M.R., Gaze, W.H., Pruden, A., Smalla, K., Tiedje, J.M., and Zhu, Y.G. (2014). Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J*.

99. Glowalla, E., Tosetti, B., Kronke, M., and Krut, O. (2009). Proteomics-Based Identification of Anchorless Cell Wall Proteins as Vaccine Candidates against *Staphylococcus aureus*. *Infect Immun* 77, 2719-2729.
100. Gomez-Arreaza, A., Acosta, H., Quinones, W., Concepcion, J.L., Michels, P.A., and Avilan, L. (2014). Extracellular functions of glycolytic enzymes of parasites: unpredicted use of ancient proteins. *Molecular and biochemical parasitology* 193, 75-81.
101. Goodwin, R.F.W., Pomeroy, A.P., and Whittles.P (1965). Production of enzootic pneumonia in pigs with a Mycoplasma. *Vet Rec* 77, 1247-&.
102. Gorg, A., Drews, O., Luck, C., Weiland, F., and Weiss, W. (2009). 2-DE with IPGs. *Electrophoresis* 30 Suppl 1, S122-132.
103. Gorg, A., Weiss, W., and Dunn, M.J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 4, 3665-3685.
104. Gruber, C.C., and Sperandio, V. (2014). Posttranscriptional control of microbe-induced rearrangement of host cell actin. *mBio* 5, e01025-01013.
105. Guimaraes, A.M.S., Santos, A.P., SanMiguel, P., Walter, T., Timenetsky, J., and Messick, J.B. (2011). Complete Genome Sequence of *Mycoplasma suis* and Insights into Its Biology and Adaption to an Erythrocyte Niche. *PLoS ONE* 6.
106. Gunawardana, Y., and Niranjan, M. (2013). Bridging the gap between transcriptome and proteome measurements identifies post-translationally regulated genes. *Bioinformatics* 29, 3060-3066.
107. Guryca, V., Lamerz, J., Ducret, A., and Cutler, P. (2012). Qualitative improvement and quantitative assessment of N-terminomics. *Proteomics* 12, 1207-1216.
108. Haesebrouck, F., Pasmans, F., Chiers, K., Maes, D., Ducatelle, R., and Decostere, A. (2004). Efficacy of vaccines against bacterial diseases in swine: what can we expect? *Vet Microbiol* 100, 255-268.
109. Han, D.K., Eng, J., Zhou, H., and Aebersold, R. (2001). Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* 19, 946-951.
110. Hashimoto, M., Tawaratsumida, K., Kariya, H., Aoyama, K., Tamura, T., and Suda, Y. (2006). Lipoprotein is a predominant Toll-like receptor 2 ligand in *Staphylococcus aureus* cell wall components. *Int Immunol* 18, 355-362.
111. Hempel, K., Herbst, F.A., Moche, M., Hecker, M., and Becher, D. (2011). Quantitative proteomic view on secreted, cell surface-associated, and cytoplasmic proteins of the methicillin-resistant human pathogen *Staphylococcus aureus* under iron-limited conditions. *J Proteome Res* 10, 1657-1666.
112. Hempel, K., Pane-Farre, J., Otto, A., Sievers, S., Hecker, M., and Becher, D. (2010). Quantitative Cell Surface Proteome Profiling for SigB-Dependent Protein Expression in the Human Pathogen *Staphylococcus aureus* via Biotinylation Approach. *J Proteome Res* 9, 1579-1590.
113. Henderson, B. (2014). An overview of protein moonlighting in bacterial infection. *Biochem Soc Trans* 42, 1720-1727.
114. Henderson, B., and Martin, A. (2011). Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect Immun* 79, 3476-3491.
115. Henderson, B., and Martin, A.C. (2014). Protein moonlighting: a new factor in biology and medicine. *Biochem Soc Trans* 42, 1671-1678.
116. Henderson, B., Nair, S., Pallas, J., and Williams, M.A. (2011). Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin-binding proteins. *FEMS Microbiol Rev* 35, 147-200.

117. Henderson, B., Wilson, M., and Hyams, J. (1998). Cellular microbiology: cycling into the millennium. *Trends Cell Biol* 8, 384-387.
118. Hermeyer, K., Buchenau, I., Thomasmeyer, A., Baum, B., Spergser, J., Rosengarten, R., and Hewicker-Trautwein, M. (2012). Chronic pneumonia in calves after experimental infection with *Mycoplasma bovis* strain 1067: characterization of lung pathology, persistence of variable surface protein antigens and local immune response. *Acta Vet Scand* 54, 9.
119. Hernandez, S., Ferragut, G., Amela, I., Perez-Pons, J., Pinol, J., Mozo-Villarias, A., Cedano, J., and Querol, E. (2014). MultitaskProtDB: a database of multitasking proteins. *Nucleic Acids Res* 42, D517-D520.
120. Hicks, S.W., and Galan, J.E. (2013). Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. *Nat Rev Microbiol* 11, 316-326.
121. Hoehenwarter, W., Ackermann, R., Zimny-Arndt, U., Kumar, N.M., and Jungblut, P.R. (2006). The necessity of functional proteomics: protein species and molecular function elucidation exemplified by in vivo alpha A crystallin N-terminal truncation. *Amino Acids* 31, 317-323.
122. Hofmann, K., and Stoffel, W. (1993). TMbase - A database of membrane spanning proteins segments. *Biol Chem Hoppe-Seyler 374,166*.
123. Horiuchi, K. (2013). A brief history of tumor necrosis factor alpha--converting enzyme: an overview of ectodomain shedding. *Keio J Med* 62, 29-36.
124. Hsu, T., Artiushin, S., and Minion, F.C. (1997). Cloning and functional analysis of the P97 swine cilium adhesin gene of *Mycoplasma hyopneumoniae*. *J Bacteriol* 179, 1317-1323.
125. Hsu, T., and Minion, F.C. (1998a). Identification of the cilium binding epitope of the *Mycoplasma hyopneumoniae* P97 adhesin. *Infect Immun* 66, 4762-4766.
126. Hsu, T., and Minion, F.C. (1998b). Molecular analysis of the P97 cilium adhesin operon of *Mycoplasma hyopneumoniae*. *Gene* 214, 13-23.
127. Hynes, R.O. (2009). The extracellular matrix: not just pretty fibrils. *Science* 326, 1216-1219.
128. Ikeda, R., and Ichikawa, T. (2014). Interaction of surface molecules on *Cryptococcus neoformans* with plasminogen. *FEMS Yeast Res* 14, 445-450.
129. Impens, F., Colaert, N., Helsens, K., Ghesquiere, B., Timmerman, E., De Bock, P.J., Chain, B.M., Vandekerckhove, J., and Gevaert, K. (2010a). A quantitative proteomics design for systematic identification of protease cleavage events. *Mol Cell Proteomics* 9, 2327-2333.
130. Impens, F., Colaert, N., Helsens, K., Plasman, K., Van Damme, P., Vandekerckhove, J., and Gevaert, K. (2010b). MS-driven protease substrate degradomics. *Proteomics* 10, 1284-1296.
131. Inamine, J.M., Ho, K.C., Loechel, S., and Hu, P.C. (1990). Evidence that UGA is read as a tryptophan codon rather than as a stop codon by *Mycoplasma pneumoniae*, Mycoplasma genitalium, and Mycoplasma gallisepticum. *J Bacteriol* 172, 504-506.
132. Jacques, M., Blanchard, B., Foiry, B., Girard, C., and Kobisch, M. (1992). In vitro colonization of porcine trachea by *Mycoplasma hyopneumoniae*. *Ann Rech Vet* 23, 239-247.
133. Jaffe, J.D., Berg, H.C., and Church, G.M. (2004a). Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* 4, 59-77.
134. Jaffe, J.D., Stange-Thomann, N., Smith, C., DeCaprio, D., Fisher, S., Butler, J., Calvo, S., Elkins, T., FitzGerald, M.G., Hafez, N.*, et al.* (2004b). The complete genome and proteome of *Mycoplasma mobile*. *Genome Res* 14, 1447-1461.

135. Jarocki, V.M., Santos, J., Tacchi, J.L., Raymond, B.B., Deutscher, A.T., Jenkins, C., Padula, M.P., and Djordjevic, S.P. (2015). MHJ_0461 is a multifunctional leucine aminopeptidase on the surface of *Mycoplasma hyopneumoniae*. *Open Biol* 5.
136. Jarocki, V.M., Tacchi, J.L., and Djordjevic, S.P. (2014). Non-proteolytic functions of microbial proteases increases pathological complexity. *Proteomics*.
137. Jechalke, S., Heuer, H., Siemens, J., Amelung, W., and Smalla, K. (2014). Fate and effects of veterinary antibiotics in soil. *Trends Microbiol*.
138. Jeffery, C.J. (1999). Moonlighting proteins. *Trends Biochem Sci* 24, 8-11.
139. Jeffery, C.J. (2003). Moonlighting proteins: old proteins learning new tricks. *Trends Genet* 19, 415-417.
140. Jenkins, C., Samudrala, R., Geary, S.J., and Djordjevic, S.P. (2008). Structural and functional characterization of an organic hydroperoxide resistance protein from *Mycoplasma gallisepticum*. *J Bacteriol* 190, 2206-2216.
141. Jenkins, C., Wilton, J.L., Minion, F.C., Falconer, L., Walker, M.J., and Djordjevic, S.P. (2006). Two domains within the *Mycoplasma hyopneumoniae* cilium adhesin bind heparin. *Infect Immun* 74, 481-487.
142. Joki, S., and Saano, V. (1994). Ciliary beat frequency at six levels of the respiratory tract in cow, dog, guinea-pig, pig, rabbit and rat. *Clin Exp Pharmacol Physiol* 21, 427-434.
143. Jordan, D., Chin, J.J., Fahy, V.A., Barton, M.D., Smith, M.G., and Trott, D.J. (2009). Antimicrobial use in the Australian pig industry: results of a national survey. *Aust Vet J* 87, 222-229.
144. Jordan, F.T., Forrester, C.A., Ripley, P.H., and Burch, D.G. (1998). *In vitro* and *in vivo* comparisons of valnemulin, tiamulin, tylosin, enrofloxacin, and lincomycin/spectinomycin against *Mycoplasma gallisepticum*. *Avian Diss* 42, 738-745.
145. Jores, J., Meens, J., Buettner, F.F., Linz, B., Naessens, J., and Gerlach, G.F. (2009). Analysis of the immunoproteome of *Mycoplasma mycoides subsp. mycoides* small colony type reveals immunogenic homologues to other known virulence traits in related Mycoplasma species. *Vet Immunol Immunopathol* 131, 238-245.
146. Juncker, A.S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* 12, 1652-1662.
147. Jung, D.W., Kim, W.H., and Williams, D.R. (2014). Chemical genetics and its application to moonlighting in glycolytic enzymes. *Biochem Soc Trans* 42, 1756-1761.
148. Karkowska-Kuleta, J., and Kozik, A. (2014). Moonlighting proteins as virulence factors of pathogenic fungi, parasitic protozoa and multicellular parasites. *Mol Oral Microbiol* 29, 270-283.
149. Karr, J.R., Sanghvi, J.C., Macklin, D.N., Arora, A., and Covert, M.W. (2013). WholeCellKB: model organism databases for comprehensive whole-cell models. *Nucleic Acids Res* 41, D787-792.
150. Kelleher, N.L. (2004). Top-down proteomics. *Anal Chem* 76, 197A-203A.
151. Kim, K.H., Compton, P.D., Tran, J.C., and Kelleher, N.L. (2015). Online matrix removal platform for coupling gel-based separations to whole protein electrospray ionization mass spectrometry. *J Proteome Res* 14, 2199-2206.
152. Kim, M.F., Heidari, M.B., Stull, S.J., McIntosh, M.A., and Wise, K.S. (1990). Identification and mapping of an immunogenic region of *Mycoplasma hyopneumoniae* p65 surface lipoprotein expressed in *Escherichia coli* from a cloned genomic fragment. *Infect Immun* 58, 2637-2643.

153. Kim, S.H., Turnbull, J., and Guimond, S. (2011). Extracellular matrix and cell signalling: the dynamic cooperation of integrin, proteoglycan and growth factor receptor. *Journal Endocrinol* 209, 139-151.
154. Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J.N., and Overall, C.M. (2011). Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc* 6, 1578-1611.
155. Kolodziejczyk, P., and Pejsak, Z. (2004). Biological properties of *Mycoplasma hyopneumoniae*, agent of mycoplasmal pneumonia in swine. *Med Weter* 60, 341-344.
156. Kucharewicz, I., Kowal, K., Buczko, W., and Bodzenta-Lukaszyk, A. (2003). The plasmin system in airway remodeling. *Thrombosis Res* 112, 1-7.
157. Kuhner, S., van Noort, V., Betts, M.J., Leo-Macias, A., Batisse, C., Rode, M., Yamada, T., Maier, T., Bader, S., Beltran-Alvarez, P*., et al.* (2009). Proteome Organization in a Genome-Reduced Bacterium. *Science* 326, 1235-1240.
158. Kunert, A., Losse, J., Gruszin, C., Huhn, M., Kaendler, K., Mikkat, S., Volke, D., Hoffmann, R., Jokiranta, T.S., Seeberger, H*., et al.* (2007). Immune evasion of the human pathogen *Pseudomonas aeruginosa*: elongation factor Tuf is a factor H and plasminogen binding protein. *J Immunol* 179, 2979-2988.
159. Kyhse-Andersen, J. (1984). Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. *J Biochem Biophys Methods* 10, 203-209.
160. La Celle, P., Blumenstock, F.A., and Saba, T.M. (1991). Blood-borne fragments of fibronectin after thermal injury. *Blood* 77, 2037-2041.
161. Lahteenmaki, K., Edelman, S., and Korhonen, T.K. (2005). Bacterial metastasis: the host plasminogen system in bacterial invasion. *Trends Microbiol* 13, 79-85.
162. Lahteenmaki, K., Kuusela, P., and Korhonen, T.K. (2001). Bacterial plasminogen activators and receptors. *FEMS Microbiol Rev* 25, 531-552.
163. Lange, P.F., and Overall, C.M. (2013). Protein TAILS: when termini tell tales of proteolysis and function. *Curr Opin Chem Biol* 17, 73-82.
164. Layh-Schmitt, G., Podtelejnikov, A., and Mann, M. (2000). Proteins complexed to the P1 adhesin of *Mycoplasma pneumoniae*. *Microbiology* 146 (Pt 3), 741-747.
165. Lazarev, V.N., Levitskii, S.A., Basovskii, Y.I., Chukin, M.M., Akopian, T.A., Vereshchagin, V.V., Kostrjukova, E.S., Kovaleva, G.Y., Kazanov, M.D., Malko, D.B*., et al.* (2011). Complete genome and proteome of *Acholeplasma laidlawii*. *J Bacteriol* 193, 4943-4953.
166. Le Carrou, J., Laurentie, M., Kobisch, M., and Gautier-Bouchardon, A.V. (2006). Persistence of *Mycoplasma hyopneumoniae* in experimentally infected pigs after marbofloxacin treatment and detection of mutations in the parC gene. *Antimicrob Agents Chemother* 50, 1959-1966.
167. Li, X., Romero, P., Rani, M., Dunker, A.K., and Obradovic, Z. (1999). Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome informatics Workshop on Genome Informatics* 10, 30-40.
168. Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure* 11, 1453-1459.
169. Liu, W., Feng, Z., Fang, L., Zhou, Z., Li, Q., Li, S., Luo, R., Wang, L., Chen, H., Shao, G*., et al.* (2011). Complete genome sequence of *Mycoplasma hyopneumoniae* strain 168. *J Bacteriol* 193, 1016-1017.
170. Lopez-Otin, C., and Overall, C.M. (2002). Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol* 3, 509-519.

171. Lopez, J.L. (2007). Two-dimensional electrophoresis in proteome expression analysis. *J Chromatogr B* 849, 190-202.
172. Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* 252, 1162-1164.
173. Ly, L., and Wasinger, V.C. (2011). Protein and peptide fractionation, enrichment and depletion: tools for the complex proteome. *Proteomics* 11, 513-534.
174. Madsen, M.L., Nettleton, D., Thacker, E.L., Edwards, R., and Minion, F.C. (2006a). Transcriptional profiling of *Mycoplasma hyopneumoniae* during heat shock using microarrays. *Infect Immun* 74, 160-166.
175. Madsen, M.L., Nettleton, D., Thacker, E.L., and Minion, F.C. (2006b). Transcriptional profiling of *Mycoplasma hyopneumoniae* during iron depletion using microarrays. *Microbiology* 152, 937-944.
176. Madsen, M.L., Oneal, M.J., Gardner, S.W., Strait, E.L., Nettleton, D., Thacker, E.L., and Minion, F.C. (2007). Array-based genomic comparative hybridization analysis of field strains of *Mycoplasma hyopneumoniae*. *J Bacteriol* 189, 7977-7982.
177. Madsen, M.L., Puttamreddy, S., Thacker, E.L., Carruthers, M.D., and Minion, F.C. (2008). Transcriptome changes in *Mycoplasma hyopneumoniae* during infection. *Infect Immun* 76, 658-663.
178. Maglennon, G.A., Cook, B.S., Deeney, A.S., Bosse, J.T., Peters, S.E., Langford, P.R., Maskell, D.J., Tucker, A.W., Wren, B.W., Rycroft, A.N.*, et al.* (2013). Transposon mutagenesis in *Mycoplasma hyopneumoniae* using a novel mariner-based system for generating random mutations. *Vet Res* 44, 124.
179. Maier, T., Marcos, J., Wodke, J.A., Paetzold, B., Liebeke, M., Gutierrez-Gallego, R., and Serrano, L. (2013). Large-scale metabolome analysis and quantitative integration with genomics and proteomics data in *Mycoplasma pneumoniae*. *Mol Biosyst* 9, 1743-1755.
180. Maier, T., Schmidt, A., Guell, M., Kuhner, S., Gavin, A.C., Aebersold, R., and Serrano, L. (2011). Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol Syst Biol* 7, 511.
181. Maillet, I., Berndt, P., Malo, C., Rodriguez, S., Brunisholz, R.A., Pragai, Z., Arnold, S., Langen, H., and Wyss, M. (2007). From the genome sequence to the proteome and back: evaluation of *E. coli* genome annotation with a 2-D gel-based proteomics approach. *Proteomics* 7, 1097-1106.
182. Mann, M., Kulak, N.A., Nagaraj, N., and Cox, J. (2013). The coming age of complete, accurate, and ubiquitous proteomes. *Mol Cell* 49, 583-590.
183. Marois, C., Le Carrou, J., Kobisch, M., and Gautier-Bouchardon, A.V. (2007). Isolation of *Mycoplasma hyopneumoniae* from different sampling sites in experimentally infected and contact SPF piglets. *Vet Microbiol* 120, 96-104.
184. Matta, S.K., Agarwal, S., and Bhatnagar, R. (2010). Surface localized and extracellular Glyceraldehyde-3-phosphate dehydrogenase of *Bacillus anthracis* is a plasminogen binding protein. *Biochim Biophys Acta* 1804, 2111-2120.
185. McAuliffe, L., Ellis, R.J., Miles, K., Ayling, R.D., and Nicholas, R.A. (2006). Biofilm formation by mycoplasma species and its role in environmental persistence and survival. *Microbiology* 152, 913-922.
186. McDonnell, A.V., Jiang, T., Keating, A.E., and Berger, B. (2006). Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics* 22, 356-358.
187. Menegatti, A.C., Tavares, C.P., Vernal, J., Klein, C.S., Huergo, L., and Terenzi, H. (2010). First partial proteome of the poultry pathogen *Mycoplasma synoviae*. *Vet Microbiol* 145, 134-141.

188. Meyer, B., Papasotiriou, D.G., and Karas, M. (2011). 100% protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* 41, 291-310.
189. Minion, F.C., Adams, C., and Hsu, T. (2000). R1 region of P97 mediates adherence of *Mycoplasma hyopneumoniae* to swine cilia. *Infect Immun* 68, 3056-3060.
190. Minion, F.C., Lefkowitz, E.J., Madsen, M.L., Cleary, B.J., Swartzell, S.M., and Mahairas, G.G. (2004). The genome sequence of *Mycoplasma hyopneumoniae* strain 232, the agent of swine mycoplasmosis. *J Bacteriol* 186, 7123-7133.
191. Minion, F.C., Menon, S.A., Mahairas, G.G., and Wannemuehler, M.J. (2003). Enhanced murine antigen-specific gamma interferon and immunoglobulin G2a responses by using mycobacterial ESAT-6 sequences in DNA vaccines. *Infect Immun* 71, 2239-2243.
192. Mobegi, F.M., van Hijum, S.A., Burghout, P., Bootsma, H.J., de Vries, S.P., van der Gaast-de Jongh, C.E., Simonetti, E., Langereis, J.D., Hermans, P.W., de Jonge, M.I.*, et al.* (2014). From microbial gene essentiality to novel antimicrobial drug targets. *BMC Genomics* 15, 958.
193. Mohan, S., Hertweck, C., Dudda, A., Hammerschmidt, S., Skerka, C., Hallstrom, T., and Zipfel, P.F. (2014). Tuf of *Streptococcus pneumoniae* is a surface displayed human complement regulator binding protein. *Mol Immunol* 62, 249-264.
194. Moitinho-Silva, L., Kondo, M.Y., Oliveira, L.C., Okamoto, D.N., Paes, J.A., Machado, M.F., Veronez, C.L., Motta, G., Andrade, S.S., Juliano, M.A.*, et al.* (2013). *Mycoplasma hyopneumoniae* in vitro peptidase activities: identification and cleavage of kallikrein-kinin system-like substrates. *Vet Microbiol* 163, 264-273.
195. Muhlradt, P.F., Kiess, M., Meyer, H., Sussmuth, R., and Jung, G. (1997). Isolation, structure elucidation, and synthesis of a macrophage stimulatory lipopeptide from *Mycoplasma fermentans* acting at picomolar concentration. *J Exp Med* 185, 1951-1958.
196. Nieves, W., Heang, J., Asakrah, S., Honer zu Bentrup, K., Roy, C.J., and Morici, L.A. (2010). Immunospecific responses to bacterial elongation factor Tu during Burkholderia infection and immunization. *PLoS One* 5, e14361.
197. Nobbs, A.H., Lamont, R.J., and Jenkinson, H.F. (2009). Streptococcus adherence and colonization. *Microbiol Mol Biol Rev* 73, 407-450.
198. Nunomura, K., Nagano, K., Itagaki, C., Taoka, M., Okamura, N., Yamauchi, Y., Sugano, S., Takahashi, N., Izumi, T., and Isobe, T. (2005). Cell surface labeling and mass spectrometry reveal diversity of cell surface markers and signaling molecules expressed in undifferentiated mouse embryonic stem cells. *Mol Cell Proteomics* 4, 1968-1976.
199. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., and Dunker, A.K. (2005). Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* 61 Suppl 7, 176-182.
200. Okamba, F.R., Moreau, E., Bouh, K.C.S., Gagnon, C.A., Massie, B., and Arella, M. (2007). Immune responses induced by replication-defective adenovirus expressing the C-terminal portion of the *Mycolasma hyopneumoniae* p97 adhesin. *Clin Vaccine Immunol* 14, 767-774.
201. Olivares-Hernandez, R., Bordel, S., and Nielsen, J. (2011). Codon usage variability determines the correlation between proteome and transcriptome fold changes. *BMC Syst Biol* 5, 33.
202. Oliveira, B.M., Coorssen, J.R., and Martins-de-Souza, D. (2014). 2DE: the phoenix of proteomics. *J Proteomics* 104, 140-150.
203. Oneal, M.J., Schafer, E.R., Madsen, M.L., and Minion, F.C. (2008). Global transcriptional analysis of *Mycoplasma hyopneumoniae* following exposure to norepinephrine. *Microbiology* 154, 2581-2588.

204. Osawa, S., Jukes, T.H., Watanabe, K., and Muto, A. (1992). Recent evidence for evolution of the genetic code. *Microbiol Rev* 56, 229-264.
205. Otake, S., Dee, S., Corzo, C., Oliveira, S., and Deen, J. (2010). Long-distance airborne transport of infectious PRRSV and *Mycoplasma hyopneumoniae* from a swine population infected with multiple viral variants. *Vet Microbiol* 145, 198-208.
206. Overall, C.M. (2002). Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. *Mol Biotechnol* 22, 51-86.
207. Ozinsky, A., Underhill, D.M., Fontenot, J.D., Hajjar, A.M., Smith, K.D., Wilson, C.B., Schroeder, L., and Aderem, A. (2000). The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. *Proc Natl Acad Sci U S A* 97, 13766-13771.
208. Pancholi, V. (2001). Multifunctional alpha-enolase: its role in diseases. *Cell Mol Life Sci* 58, 902-920.
209. Pancholi, V., and Chhatwal, G.S. (2003). Housekeeping enzymes as virulence factors for pathogens. *Int J Med Microbiol* 293, 391-401.
210. Parraga-Nino, N., Colome-Calls, N., Canals, F., Querol, E., and Ferrer-Navarro, M. (2012). A comprehensive proteome of *Mycoplasma genitalium*. *J Proteome Res* 11, 3305-3316.
211. Pearson, T.W., Kar, S.K., McGuire, T.C., and Lundin, L.B. (1981). Trypanosome variable surface antigens: studies using two-dimensional gel electrophoresis and monoclonal antibodies. *J Immunol* 126, 823-828.
212. Pendarvis, K., Padula, M.P., Tacchi, J.L., Petersen, A.C., Djordjevic, S.P., Burgess, S.C., and Minion, F.C. (2014). Proteogenomic mapping of *Mycoplasma hyopneumoniae* virulent strain 232. *BMC Genomics* 15, 576.
213. Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8, 785-786.
214. Petit, F.M., Serres, C., and Auer, J. (2014). Moonlighting proteins in sperm-egg interactions. *Biochem Soc Trans* 42, 1740-1743.
215. Petit, F.M., Serres, C., Bourgeon, F., Pineau, C., and Auer, J. (2013). Identification of sperm head proteins involved in zona pellucida binding. *Hum Reprod* 28, 852-865.
216. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comp Chem* 25, 1605-1612.
217. Pinto, P., Klein, C., Zaha, A., and Ferreira, H. (2009). Comparative proteomic analysis of pathogenic and non-pathogenic strains from the swine pathogen *Mycoplasma hyopneumoniae*. *Proteome Sci* 7, 45.
218. Pinto, P.M., Chemale, G., de Castro, L.A., Costa, A.P.M., Kich, J.D., Vainstein, M.H., Zaha, A., and Ferreira, H.B. (2007). Proteomic survey of the pathogenic *Mycoplasma hyopneumoniae* strain 7448 and identification of novel post-translationally modified and antigenic proteins. *Vet Microbiol* 121, 83-93.
219. Porteus, B., Kocharunchitt, C., Nilsson, R.E., Ross, T., and Bowman, J.P. (2011). Utility of gel-free, label-free shotgun proteomics approaches to investigate microorganisms. *Appl Microbiol Biotechnol* 90, 407-416.
220. Quigley, J.P., Gold, L.I., Schwimmer, R., and Sullivan, L.M. (1987). Limited cleavage of cellular fibronectin by plasminogen activator purified from transformed cells. *Proc Natl Acad Sci U S A* 84, 2776-2780.
221. Rabenstein, D.L. (2002). Heparin and heparan sulfate: structure and function. *Nat Prod Rep* 19, 312-331.

222. Rabijns, A., De Bondt, H.L., and De Ranter, C. (1997). Three-dimensional structure of staphylokinase, a plasminogen activator with therapeutic potential. *Nature Struct Biol* 4, 357-360.
223. Rahman, O., Cummings, S., Harrington, D., and Sutcliffe, I. (2008). Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. *World J Microbiol Biotechnol* 24, 2377-2382.
224. Raymond, B.B., and Djordjevic, S. (2015). Exploitation of plasmin(ogen) by bacterial pathogens of veterinary significance. *Vet Microbiol* 178, 1-13.
225. Raymond, B.B., Jenkins, C., Seymour, L.M., Tacchi, J.L., Widjaja, M., Jarocki, V.M., Deutscher, A.T., Turnbull, L., Whitchurch, C.B., Padula, M.P.*, et al.* (2014). Proteolytic processing of the cilium adhesin MHJ_0194 (P123) in *Mycoplasma hyopneumoniae* generates a functionally diverse array of cleavage fragments that bind multiple host molecules. *Cell Microbiol*.
226. Raymond, B.B., Tacchi, J.L., Jarocki, V.M., Minion, F.C., Padula, M.P., and Djordjevic, S.P. (2013). P159 from *Mycoplasma hyopneumoniae* binds porcine cilia and heparin and is cleaved in a manner akin to ectodomain shedding. *J Proteome Res* 12, 5891-5903.
227. Razin, S., and Hayflick, L. (2010). Highlights of mycoplasma research--an historical perspective. *Biologicals* 38, 183-190.
228. Razin, S., Yogev, D., and Naot, Y. (1998). Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev* 62, 1094-1156.
229. Reinders, J., Zahedi, R.P., Pfanner, N., Meisinger, C., and Sickmann, A. (2006). Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. *J Proteome Res* 5, 1543-1554.
230. Reolon, L.A., Martello, C.L., Schrank, I.S., and Ferreira, H.B. (2014). Survey of Surface Proteins from the Pathogenic *Mycoplasma hyopneumoniae* Strain 7448 Using a Biotin Cell Surface Labeling Approach. *PLoS One* 9, e112596.
231. Robinson, M.W., Buchtmann, K.A., Jenkins, C., Tacchi, J.L., Raymond, B.B., To, J., Roy Chowdhury, P., Woolley, L.K., Labbate, M., Turnbull, L.*, et al.* (2013). MHJ_0125 is an M42 glutamyl aminopeptidase that moonlights as a multifunctional adhesin on the surface of *Mycoplasma hyopneumoniae*. *Open Biol* 3, 130017.
232. Rodriguez-Ortega, M.J., Norais, N., Bensi, G., Liberatori, S., Capo, S., Mora, M., Scarselli, M., Doro, F., Ferrari, G., Garaguso, I.*, et al.* (2006). Characterization and identification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome. *Nat Biotechnol* 24, 191-197.
233. Romero, Obradovic, and Dunker, K. (1997). Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome informatics Workshop on Genome Informatics* 8, 110-124.
234. Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. *Proteins* 42, 38-48.
235. Rubin, M.S., and Tzagoloff, A. (1973). Assembly of the mitochondrial membrane system. IX. Purification, characterization, and subunit structure of yeast and beef cytochrome oxidase. *J Biol Chem* 248, 4269-4274.
236. Sasikumar, A.N., Perez, W.B., and Kinzy, T.G. (2012). The many roles of the eukaryotic elongation factor 1 complex. *Wiley interdisciplinary reviews RNA* 3, 543-555.
237. Scarman, A.L., Chin, J.C., Eamens, G.J., Delaney, S.F., and Djordjevic, S.P. (1997). Identification of novel species-specific antigens of *Mycoplasma hyopneumoniae* by preparative SDS-PAGE ELISA profiling. *Microbiology* 143 (Pt 2), 663-673.

238. Schafer, E.R., Oneal, M.J., Madsen, M.L., and Minion, F.C. (2007). Global transcriptional analysis of *Mycoplasma hyopneumoniae* following exposure to hydrogen peroxide. *Microbiology* 153, 3785-3790.
239. Schilling, O., and Overall, C.M. (2007). Proteomic discovery of protease substrates. *Curr Opin Chem Biol* 11, 36-45.
240. Schilling, O., and Overall, C.M. (2008). Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* 26, 685-694.
241. Schmidt, J.A., Browning, G.F., and Markham, P.F. (2004). *Mycoplasma hyopneumoniae* p65 surface lipoprotein is a lipolytic enzyme with a preference for shorter-chain fatty acids. *J Bacteriol* 186, 5790-5798.
242. Schmidt, J.A., Browning, G.F., and Markham, P.F. (2007). *Mycoplasma hyopneumoniae* mhp379 is a Ca2+-dependent, sugar-nonspecific exonuclease exposed on the cell surface. *J Bacteriol* 189, 3414-3424.
243. Scott, N.E., Marzook, N.B., Deutscher, A., Falconer, L., Crossett, B., Djordjevic, S.P., and Cordwell, S.J. (2010). Mass spectrometric characterization of the *Campylobacter jejuni* adherence factor CadF reveals post-translational processing that removes immunogenicity while retaining fibronectin binding. *Proteomics* 10, 277-288.
244. Severin, A., Nickbarg, E., Wooters, J., Quazi, S.A., Matsuka, Y.V., Murphy, E., Moutsatsos, I.K., Zagursky, R.J., and Olmsted, S.B. (2007). Proteomic analysis and identification of *Streptococcus pyogenes* surface-associated proteins. *J Bacteriol* 189, 1514-1522.
245. Seymour, L.M., Deutscher, A.T., Jenkins, C., Kuit, T.A., Falconer, L., Minion, F.C., Crossett, B., Padula, M., Dixon, N.E., Djordjevic, S.P*., et al.* (2010). A processed multidomain *Mycoplasma hyopneumoniae* adhesin binds fibronectin, plasminogen, and swine respiratory cilia. *J Biol Chem* 285, 33971-33978.
246. Seymour, L.M., Falconer, L., Deutscher, A.T., Minion, F.C., Padula, M.P., Dixon, N.E., Djordjevic, S.P., and Walker, M.J. (2011). Mhp107 is a member of the multifunctional adhesin family of *Mycoplasma hyopneumoniae*. *J Biol Chem* 286, 10097-10104.
247. Seymour, L.M., Jenkins, C., Deutscher, A.T., Raymond, B.B., Padula, M.P., Tacchi, J.L., Bogema, D.R., Eamens, G.J., Woolley, L.K., Dixon, N.E*., et al.* (2012). Mhp182 (P102) binds fibronectin and contributes to the recruitment of plasmin(ogen) to the *Mycoplasma hyopneumoniae* cell surface. *Cell Microbiol* 14, 81-94.
248. Shahinian, H., Tholen, S., and Schilling, O. (2013). Proteomic identification of protease cleavage sites: cell-biological and biomedical applications. *Expert Rev Proteomics* 10, 421-433.
249. Siqueira, F.M., Gerber, A.L., Guedes, R.L., Almeida, L.G., Schrank, I.S., Vasconcelos, A.T., and Zaha, A. (2014). Unravelling the transcriptome profile of the Swine respiratory tract mycoplasmas. *PLoS One* 9, e110327.
250. Sirover, M.A. (1999). New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. *Biochim Biophys Acta* 1432, 159-184.
251. Smith, L.M., Kelleher, N.L., and Consortium for Top Down, P. (2013). Proteoform: a single term describing protein complexity. *Nat Methods* 10, 186-187.
252. Sodeinde, O.A., Sample, A.K., Brubaker, R.R., and Goguen, J.D. (1988). Plasminogen activator/coagulase gene of *Yersinia pestis* is responsible for degradation of plasmid-encoded outer membrane proteins. *Infect Immun* 56, 2749-2752.
253. Solis, N., and Cordwell, S.J. (2011). Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. *Proteomics* 11, 3169-3189.

254. Solis, N., Larsen, M.R., and Cordwell, S.J. (2010). Improved accuracy of cell surface shaving proteomics in *Staphylococcus aureus* using a false-positive control. *Proteomics* 10, 2037-2049.
255. Stevens, J.M., Galyov, E.E., and Stevens, M.P. (2006). Actin-dependent movement of bacterial pathogens. *Nat Rev Microbiol* 4, 91-101.
256. Sueyoshi, N., Nimura, T., Onouchi, T., Baba, H., Takenaka, S., Ishida, A., and Kameshita, I. (2012). Functional processing of nuclear Ca2+/calmodulin-dependent protein kinase phosphatase (CaMKP-N): evidence for a critical role of proteolytic processing in the regulation of its catalytic activity, subcellular localization and substrate targeting in vivo. *Arch Biochem Biophys* 517, 43-52.
257. Syrovets, T., and Simmet, T. (2004). Novel aspects and new roles for the serine protease plasmin. *Cell Mol Life Sci* 61, 873-885.
258. Szczepanek, S.M., Frasca, S., Jr., Schumacher, V.L., Liao, X., Padula, M., Djordjevic, S.P., and Geary, S.J. (2010). Identification of lipoprotein MslA as a neoteric virulence factor of *Mycoplasma gallisepticum*. *Infect Immun* 78, 3475-3483.
259. Tacchi, J.L., Raymond, B.B., Jarocki, V.M., Berry, I.J., Padula, M.P., and Djordjevic, S.P. (2014). Cilium adhesin P216 (MHJ_0493) is a target of ectodomain shedding and aminopeptidase activity on the surface of *Mycoplasma hyopneumoniae*. *J Proteome Res* 13, 2920-2930.
260. Takeuchi, O., Kawai, T., Muhlradt, P.F., Morr, M., Radolf, J.D., Zychlinsky, A., Takeda, K., and Akira, S. (2001). Discrimination of bacterial lipoproteins by Toll-like receptor 6. *Int Immunol* 13, 933-940.
261. Talman, A.M., Chong, R., Chia, J., Svitkina, T., and Agaisse, H. (2014). Actin network disassembly powers dissemination of *Listeria monocytogenes*. *J Cell Sci* 127, 240-249.
262. Tan, S., Tan, H.T., and Chung, M.C.M. (2008). Membrane proteins and membrane proteomics. *Proteomics* 8, 3924-3932.
263. Thacker, E.L. (2006). *Mycoplasmal diseases*. In *Diseases of Swine*, B.E. Straw, J.J. Zimmerman, S. D'Allaire, and D.J. Taylor, eds. (Oxford, UK: Blackwell Publishing).
264. Thacker, E.L., Halbur, P.G., Paul, P.S., and Thacker, B.J. (1998a). Detection of intracellular porcine reproductive and respiratory syndrome virus nucleocapsid protein in porcine macrophages by flow cytometry. *Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc* 10, 308-311.
265. Thacker, E.L., Thacker, B.J., Boettcher, T.B., and Jayappa, H. (1998b). Comparison of antibody production, lymphocyte stimulation, and protection induced by four commercial *Mycoplasma hyopneumoniae* bacterins. *Swine Health and Production* 6, 107-112.
266. Thai, K.Y., and Forney, J.D. (2000). Analysis of the conserved cysteine periodicity of Paramecium variable surface antigens. *J Euk Microbiol* 47, 242-248.
267. Tjalsma, H., Antelmann, H., Jongbloed, J.D., Braun, P.G., Darmon, E., Dorenbos, R., Dubois, J.Y., Westers, H., Zanen, G., Quax, W.J.*, et al.* (2004). Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. *Microbiol Mol Biol Rev* 68, 207-233.
268. Tjalsma, H., Bolhuis, A., Jongbloed, J.D., Bron, S., and van Dijl, J.M. (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* 64, 515-547.
269. Tjalsma, H., Lambooy, L., Hermans, P.W., and Swinkels, D.W. (2008). Shedding & shaving: disclosure of proteomic expressions on a bacterial face. *Proteomics* 8, 1415-1428.

270. Toledo, A., Coleman, J.L., Kuhlow, C.J., Crowley, J.T., and Benach, J.L. (2012). The Enolase of *Borrelia burgdorferi* Is a Plasminogen Receptor Released in Outer Membrane Vesicles. *Infect Immun* 80, 359-368.
271. Tonello, F., Morante, S., Rossetto, O., Schiavo, G., and Montecucco, C. (1996). Tetanus and botulism neurotoxins: a novel group of zinc-endopeptidases. *Adv Exp Med Biol* 389, 251-260.
272. Travis, J., and Potempa, J. (2000). Bacterial proteinases as targets for the development of second-generation antibiotics. *Biochim Biophys Acta* 1477, 35-50.
273. Truong, D., Copeland, J.W., and Brumell, J.H. (2014). Bacterial subversion of host cytoskeletal machinery: hijacking formins and the Arp2/3 complex. *Bioessays* 36, 687-696.
274. Tunio, S.A., Oldfield, N.J., Berry, A., Ala'Aldeen, D.A., Wooldridge, K.G., and Turner, D.P. (2010). The moonlighting protein fructose-1, 6-bisphosphate aldolase of *Neisseria meningitidis*: surface localization and role in host cell adhesion. *Mol Microbiol* 76, 605-615.
275. Turk, B. (2006). Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discovery* 5, 785-799.
276. Ueberle, B., Frank, R., and Herrmann, R. (2002). The proteome of the bacterium *Mycoplasma pneumoniae*: comparing predicted open reading frames to identified gene products. *Proteomics* 2, 754-764.
277. UniProt, C. (2015). UniProt: a hub for protein information. *Nucleic Acids Res* 43, D204-212.
278. Uversky, V.N., and Dunker, A.K. (2010). Understanding protein non-folding. *Biochim Biophys Acta* 1804, 1231-1264.
279. van Noort, V., Seebacher, J., Bader, S., Mohammed, S., Vonkova, I., Betts, M.J., Kuhner, S., Kumar, R., Maier, T., O'Flaherty, M.*, et al.* (2012). Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol Syst Biol* 8, 571.
280. Vasconcelos, A.T., Ferreira, H.B., Bizarro, C.V., Bonatto, S.L., Carvalho, M.O., Pinto, P.M., Almeida, D.F., Almeida, L.G., Almeida, R., Alves-Filho, L.*, et al.* (2005). Swine and poultry pathogens: the complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. *J Bacteriol* 187, 5568-5577.
281. Ventura, C.L., Malachowa, N., Hammer, C.H., Nardone, G.A., Robinson, M.A., Kobayashi, S.D., and DeLeo, F.R. (2010). Identification of a Novel *Staphylococcus aureus* Two-Component Leukotoxin Using Cell Surface Proteomics. *PLoS One* 5.
282. Vicca, J., Maes, D., Stakenborg, T., Butaye, P., Minion, F., Peeters, J., de Kruif, A., Decostere, A., and Haesebrouck, F. (2007). Resistance mechanism against fluoroquinolones in *Mycoplasma hyopneumoniae* field isolates. *Microb Drug Resist* 13, 166-170.
283. Vicca, J., Stakenborg, T., Maes, D., Butaye, P., Peeters, J., de Kruif, A., and Haesebrouck, F. (2004). In vitro susceptibilities of *Mycoplasma hyopneumoniae* field isolates. *Antimicrob Agents Chemother* 48, 4470-4472.
284. Wachtfogel, Y.T., Abrams, W., Kucich, U., Weinbaum, G., Schapira, M., and Colman, R.W. (1988). Fibronectin degradation products containing the cytoadhesive tetrapeptide stimulate human neutrophil degranulation. *J Clin Invest* 81, 1310-1316.
285. Wang, G.Q., Xia, Y., Cui, J., Gu, Z.N., Song, Y.D., Chen, Y.Q., Chen, H.Q., Zhang, H., and Chen, W. (2014). The Roles of Moonlighting Proteins in Bacteria. *Curr Issues Mol Biol* 16, 15-22.

286. Wang, H., Lottenberg, R., and Boyle, M.D. (1995). A role for fibrinogen in the streptokinase-dependent acquisition of plasmin(ogen) by group A streptococci. *J Infect Dis* 171, 85-92.
287. White, M.Y., Brown, D.A., Sheng, S., Cole, R.N., O'Rourke, B., and Van Eyk, J.E. (2011). Parallel proteomics to improve coverage and confidence in the partially annotated *Oryctolagus cuniculus* mitochondrial proteome. *Mol Cell Proteomics* 10, M110 004291.
288. Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112, 531-552.
289. Wilton, J., Jenkins, C., Cordwell, S.J., Falconer, L., Minion, F.C., Oneal, D.C., Djordjevic, M.A., Connolly, A., Barchia, I., Walker, M.J.*, et al.* (2009). Mhp493 (P216) is a proteolytically processed, cilium and heparin binding protein of *Mycoplasma hyopneumoniae*. *Mol Microbiol* 71, 566-582.
290. Wilton, J.L., Scarman, A.L., Walker, M.J., and Djordjevic, S.P. (1998). Reiterated repeat region variability in the ciliary adhesin gene of *Mycoplasma hyopneumoniae*. *Microbiology* 144 (Pt 7), 1931-1943.
291. Wise, K.S., Foecking, M.F., Roske, K., Lee, Y.J., Lee, Y.M., Madan, A., and Calcutt, M.J. (2006). Distinctive repertoire of contingency genes conferring mutation-based phase variation and combinatorial expression of surface lipoproteins in *Mycoplasma capricolum subsp. capricolum* of the *Mycoplasma mycoides* phylogenetic cluster. *J Bacteriol* 188, 4926-4941.
292. Wise, K.S., and Kim, M.F. (1987a). Identification of intrinsic and extrinsic membrane proteins bearing surface epitopes of *Mycoplasma hyopneumoniae*. *Isr J Med Sci* 23, 469-473.
293. Wise, K.S., and Kim, M.F. (1987b). Major membrane surface proteins of *Mycoplasma hyopneumoniae* selectively modified by covalently bound lipid. *J Bacteriol* 169, 5546-5555.
294. Wodke, J.A., Puchalka, J., Lluch-Senar, M., Marcos, J., Yus, E., Godinho, M., Gutierrez-Gallego, R., dos Santos, V.A., Serrano, L., Klipp, E.*, et al.* (2013). Dissecting the energy metabolism in *Mycoplasma pneumoniae* through genome-scale metabolic modeling. *Mol Syst Biol* 9, 653.
295. Woese, C.R. (1987). Bacterial evolution. *Microbiol Rev* 51, 221-271.
296. Wolff, D.G., Castiblanco-Valencia, M.M., Abe, C.M., Monaris, D., Morais, Z.M., Souza, G.O., Vasconcellos, S.A., Isaac, L., Abreu, P.A., and Barbosa, A.S. (2013). Interaction of Leptospira elongation factor Tu with plasminogen and complement factor H: a metabolic leptospiral protein with moonlighting activities. *PLoS One* 8, e81818.
297. Woolley, L.K., Fell, S.A., Djordjevic, S.P., Eamens, G.J., and Jenkins, C. (2013). Plasmin activity in the porcine airways is enhanced during experimental infection with *Mycoplasma hyopneumoniae*, is positively correlated with proinflammatory cytokine levels and is ameliorated by vaccination. *Vet Microbiol* 164, 60-66.
298. Woolley, L.K., Fell, S.A., Gonsalves, J.R., Raymond, B.B., Collins, D., Kuit, T.A., Walker, M.J., Djordjevic, S.P., Eamens, G.J., and Jenkins, C. (2014). Evaluation of recombinant *Mycoplasma hyopneumoniae* P97/P102 paralogs formulated with selected adjuvants as vaccines against mycoplasmal pneumonia in pigs. *Vaccine* 32, 4333-4341.
299. Wurpel, D.J., Moriel, D.G., Totsika, M., Easton, D.M., and Schembri, M.A. (2014). Comparative analysis of the uropathogenic *Escherichia coli* surface proteome by tandem mass-spectrometry of artificially induced outer membrane vesicles. *J Proteomics*.

300. Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K., and Uversky, V.N. (2010). PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 1804, 996-1010.
301. Yang, C.K., Ewis, H.E., Zhang, X., Lu, C.D., Hu, H.J., Pan, Y., Abdelal, A.T., and Tai, P.C. (2011). Nonclassical protein secretion by *Bacillus subtilis* in the stationary phase is not due to cell lysis. *J Bacteriol* 193, 5607-5615.
302. Young, T.F., Thacker, E.L., Erickson, B.Z., and Ross, R.F. (2000). A tissue culture system to study respiratory ciliary epithelial adherence of selected swine mycoplasmas. *Vet Microbiol* 71, 269-279.
303. Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J.*, et al.* (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608-1615.
304. Yuan, C.L., Yang, X.W., Yang, Z.B., Zhu, N.Y., Zheng, S.B., Hou, P.X., Gu, X.F., Ye, C.R., Yao, C.B., Zhu, J.G.*, et al.* (2010). Proteomic study of *Mycoplasma suis* using the gel-based shotgun strategy. *Vet Microbiol* 142, 303-308.
305. Zaffagnini, M., Fermani, S., Costa, A., Lemaire, S.D., and Trost, P. (2013). Plant cytoplasmic GAPDH: redox post-translational modifications and moonlighting properties. *Front Plant Sci* 4.
306. Zhang, H., Zheng, J., Yi, L., Li, Y., Ma, Z., Fan, H., and Lu, C. (2014). The identification of six novel proteins with fibronectin or collagen type I binding activity from *Streptococcus suis* serotype 2. *J Microbiol* 52, 963-969.
307. Zhang, Q., Young, T.F., and Ross, R.F. (1994). Microtiter plate adherence assay and receptor analogs for *Mycoplasma hyopneumoniae*. *Infect Immun* 62, 1616-1622.
308. Zhang, Q.J., Young, T.F., and Ross, R.F. (1995). Identification and characterization of a *Mycoplasma hyopneumoniae* adhesin. *Infect Immun* 63, 1013-1019.
309. Zhao, B., Lei, L., Vassylyev, D.G., Lin, X., Cane, D.E., Kelly, S.L., Yuan, H., Lamb, D.C., and Waterman, M.R. (2009). Crystal structure of albaflavenone monooxygenase containing a moonlighting terpene synthase active site. *J Biol Chem* 284, 36711-36719.
310. Zielinski, G.C., and Ross, R.F. (1992). Morphological Features and Hydrophobicity of the Cell-Surface of *Mycoplasma hyopneumoniae*. *Am J Vet Res* 53, 1119-1124.
311. Zielinski, G.C., and Ross, R.F. (1993). Adherence of *Mycoplasma hyopneumoniae* to porcine ciliated respiratory tract cells. *Am J Vet Res* 54, 1262-1269.
312. Zielinski, G.C., Young, T., Ross, R.F., and Rosenbusch, R.F. (1990). Adherence of *Mycoplasma hyopneumoniae* to Cell Monolayers. *Am J Vet Res* 51, 339-343.

Appendix I – Supplementary information

Friis Media preparation

Hanks Buffered Salt Solution for Friis media:

Hanks A Components	**Per 500 mL**
Sodium Chloride	80 g
Potassium Chloride	4 g
Magnesium Sulfate heptahydrate	1 g
Magnesium Chloride hexahydrate	1 g
Calcium Chloride	1.4 g
Hanks B Components	**Per 500 mL**
diSodium Hydrogen Orthophosphate	1.5 g
Potassium diHydrogen Orthophosphate	0.6 g

Antibiotic solution (made fresh):

Components	**Per 5.7 mL**
Bacitracin	0.144 g
Ampicillin	0.144 g
MilliQ water	5.7 mL

Yeast extract:

Components	**Per Litre**
Baker's Yeast Type (II)	167 g
MilliQ water	1000 mL

Incubated at 37°C for 30 min, heated to 90-100°C for 5 min and centrifuged to remove cells. Supernatant passed through a 0.22 μm filter and sterilized by autoclaving prior to aliquotting and use.

Friis base media (pH 7.4):

Components	**Amount**
Hanks A	15.2 mL
Hanks B	15.2 mL
Difco Brain Heart Infusion	4.9 g
Difco PPLO w/o CV	5.25 g
0.5% Phenol Red	2.6 mL
MilliQ water	827 mL

Components added to Friis base before use:

Components	**Amount**
Yeast extract	34 mL
Antibiotic solution	5.7 mL
Horse Serum	50 mL
Pig Serum	50 mL

Horse and Pig Serum purchased commercially [Invitrogen].

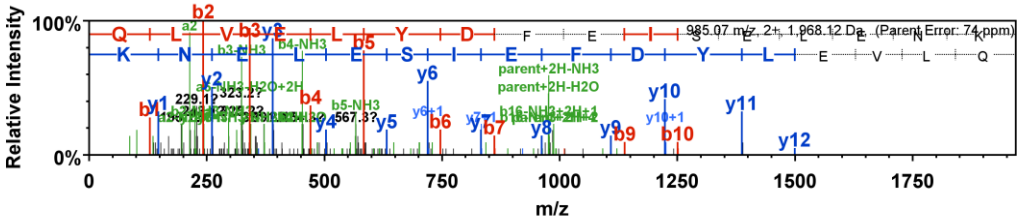
Chapter 3: Supporting information

See also Appendix II.

Supplementary file 3.2. Spectra and fragmentation tables identifying single-peptide protein identifications – "One-hit wonders"

ID #	Accession	m/z	z	Error (ppm)	Peptide	Score
285	Q4A9X7	985.07	2	74	QLVELYDFEISELENK	60
339	Q4AAC5	747.87	2	4.9	VDDENLDSIFSK	70.2
344	Q4AA34	630.37	2	2.1	VVGISSLTTLNR	51.8
345	A4Q7T6	985.07	2	74	QLVELYDFESELENK	68.6
347	Q4A9M3	822.94	2	10	YNLSNFFILISDAK	51.4

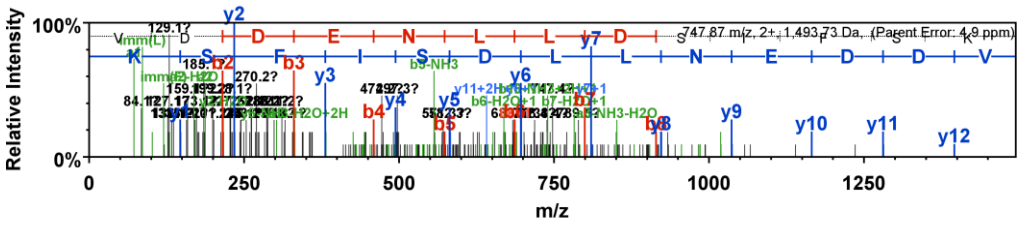
#285 Q4A9X7 MHJ_0356 Putative P37-like ABC transporter substrate-binding lipoprotein – identified from TX-114 1DGE LC-MS/MS (Slice 14)



B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	129.1		112.0		Q	1,969.0	985.0	1,952.0	1,951.0	16
2	242.1		225.1		L	1,840.9	921.0	1,823.9	1,822.9	15
3	341.2		324.2		V	1,727.8	864.4	1,710.8	1,709.8	14
4	470.3		453.2	452.3	E	1,628.8	814.9	1,611.7	1,610.8	13
5	583.3		566.3	565.3	L	1,499.7	750.4	1,482.7	1,481.7	12
6	746.4	373.7	729.4	728.4	Y	1,386.6	693.8	1,369.6	1,368.6	11
7	861.4	431.2	844.4	843.4	D	1,223.6	612.3	1,206.6	1,205.6	10
8	1,008.5	504.8	991.5	990.5	F	1,108.6	554.8	1,091.5	1,090.5	9
9	1,137.5	569.3	1,120.5	1,119.5	E	961.5	481.2	944.5	943.5	8
10	1,250.6	625.8	1,233.6	1,232.6	I	832.4	416.7	815.4	814.4	7
11	1,337.7	669.3	1,320.6	1,319.7	S	719.4	360.2	702.3	701.3	6
12	1,466.7	733.9	1,449.7	1,448.7	E	632.3		615.3	614.3	5
13	1,579.8	790.4	1,562.8	1,561.8	L	503.3		486.3	485.3	4
14	1,708.8	854.9	1,691.8	1,690.8	E	390.2		373.2	372.2	3
15	1,822.9	911.9	1,805.8	1,804.9	N	261.2		244.1		2
16	1,969.0	985.0	1,952.0	1,951.0	K	147.1		130.1		1

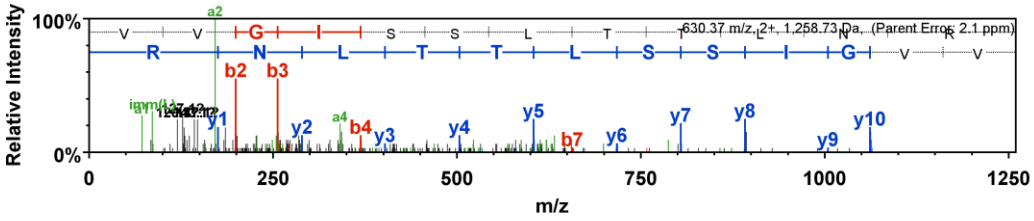
#339 Q4AAC5 MHJ_0205 tRNA modification GTPase MnmE (EC 3.6.-.-) – identified from QTOF 1DGE LC-MS/MS (Slice 7)

Correct mass context for predicted intact protein



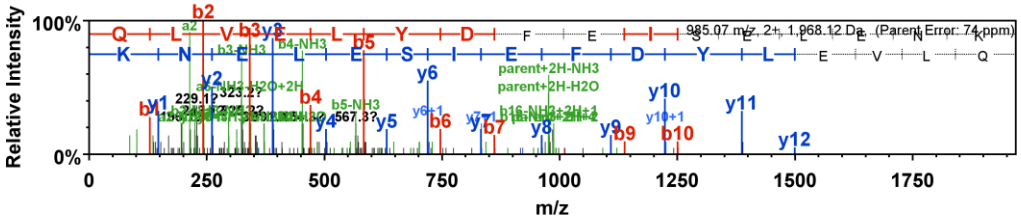
B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	100.1				V	1,494.7	747.9	1,477.7	1,476.7	13
2	215.1			197.1	D	1,395.7	698.3	1,378.6	1,377.7	12
3	330.1			312.1	D	1,280.6	640.8	1,263.6	1,262.6	11
4	459.2			441.2	E	1,165.6	583.3	1,148.6	1,147.6	10
5	573.2		556.2	555.2	N	1,036.6	518.8	1,019.5	1,018.6	9
6	686.3	343.7	669.3	668.3	L	922.5	461.8	905.5	904.5	8
7	799.4	400.2	782.4	781.4	L	809.4	405.2	792.4	791.4	7
8	914.4	457.7	897.4	896.4	D	696.4	348.7	679.3	678.3	6
9	1,001.4	501.2	984.4	983.4	S	581.3		564.3	563.3	5
10	1,114.5	557.8	1,097.5	1,096.5	I	494.3		477.3	476.3	4
11	1,261.6	631.3	1,244.6	1,243.6	F	381.2		364.2	363.2	3
12	1,348.6	674.8	1,331.6	1,330.6	S	234.1		217.1	216.1	2
13	1,494.7	747.9	1,477.7	1,476.7	K	147.1		130.1		1

#344 Q4AA34 MHJ_0297 Putative ABC transporter ATP-binding protein – identified from QTOF overload lane (Slice 1)



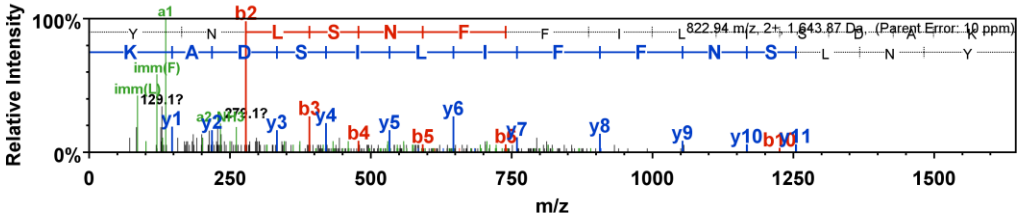
B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	100.1				V	1,259.7	630.4	1,242.7	1,241.7	12
2	199.1				V	1,160.7	580.8	1,143.6	1,142.7	11
3	256.2				G	1,061.6	531.3	1,044.6	1,043.6	10
4	369.2				I	1,004.6	502.8	987.5	986.6	9
5	456.3			438.3	S	891.5	446.2	874.5	873.5	8
6	543.3	272.2		525.3	S	804.5	402.7	787.4	786.4	7
7	656.4	328.7		638.4	L	717.4	359.2	700.4	699.4	6
8	757.4	379.2		739.4	T	604.3		587.3	586.3	5
9	858.5	429.8		840.5	T	503.3		486.3	485.3	4
10	971.6	486.3		953.6	L	402.2		385.2		3
11	1,085.6	543.3	1,068.6	1,067.6	N	289.2		272.1		2
12	1,259.7	630.4	1,242.7	1,241.7	R	175.1		158.1		1

#345 A4Q7T6 MHJ_0696 Uncharacterised protein – identified from TX-114 1DGE LC-MS/MS (Slice 14). Correct mass context for predicted intact protein



B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	129.1		112.0		Q	1,969.0	985.0	1,952.0	1,951.0	16
2	242.1		225.1		L	1,840.9	921.0	1,823.9	1,822.9	15
3	341.2		324.2		V	1,727.8	864.4	1,710.8	1,709.8	14
4	470.3		453.2	452.3	E	1,628.8	814.9	1,611.7	1,610.8	13
5	583.3		566.3	565.3	L	1,499.7	750.4	1,482.7	1,481.7	12
6	746.4	373.7	729.4	728.4	Y	1,386.6	693.8	1,369.6	1,368.6	11
7	861.4	431.2	844.4	843.4	D	1,223.6	612.3	1,206.6	1,205.6	10
8	1,008.5	504.8	991.5	990.5	F	1,108.6	554.8	1,091.5	1,090.5	9
9	1,137.5	569.3	1,120.5	1,119.5	E	961.5	481.2	944.5	943.5	8
10	1,250.6	625.8	1,233.6	1,232.6	I	832.4	416.7	815.4	814.4	7
11	1,337.7	669.3	1,320.6	1,319.7	S	719.4	360.2	702.3	701.3	6
12	1,466.7	733.9	1,449.7	1,448.7	E	632.3		615.3	614.3	5
13	1,579.8	790.4	1,562.8	1,561.8	L	503.3		486.3	485.3	4
14	1,708.8	854.9	1,691.8	1,690.8	E	390.2		373.2	372.2	3
15	1,822.9	911.9	1,805.8	1,804.9	N	261.2		244.1		2
16	1,969.0	985.0	1,952.0	1,951.0	K	147.1		130.1		1

#347 Q4A9M3 MHJ_0462 tRNA (guanine-N(7)-)-methyltransferase (EC 2.1.1.33) (tRNA (guanine(46)-N(7))-methyltransferase) – identified from QTOF overload lane (Slice 25)



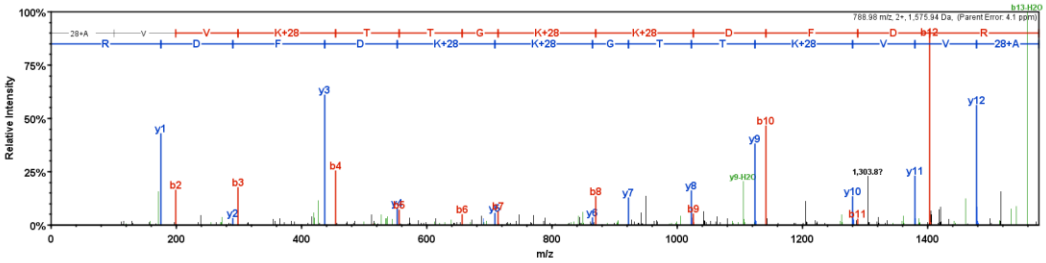
B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	164.1				Y	1,644.9	822.9	1,627.8	1,626.9	14
2	278.1		261.1		N	1,481.8	741.4	1,464.8	1,463.8	13
3	391.2		374.2		L	1,367.8	684.4	1,350.7	1,349.7	12
4	478.2		461.2	460.2	S	1,254.7	627.8	1,237.6	1,236.7	11
5	592.3		575.2	574.3	N	1,167.6	584.3	1,150.6	1,149.6	10
6	739.3	370.2	722.3	721.3	F	1,053.6	527.3	1,036.6	1,035.6	9
7	886.4	443.7	869.4	868.4	F	906.5	453.8	889.5	888.5	8
8	999.5	500.3	982.5	981.5	I	759.5	380.2	742.4	741.5	7
9	1,112.6	556.8	1,095.6	1,094.6	L	646.4	323.7	629.4	628.4	6
10	1,225.7	613.3	1,208.6	1,207.7	I	533.3		516.3	515.3	5
11	1,312.7	656.9	1,295.7	1,294.7	S	420.2		403.2	402.2	4
12	1,427.7	714.4	1,410.7	1,409.7	D	333.2		316.2	315.2	3
13	1,498.8	749.9	1,481.7	1,480.7	A	218.1		201.1		2
14	1,644.9	822.9	1,627.8	1,626.9	K	147.1		130.1		1

Supplementary file S3.5. N-terminal labelling and semi-tryptic peptides of EfTu

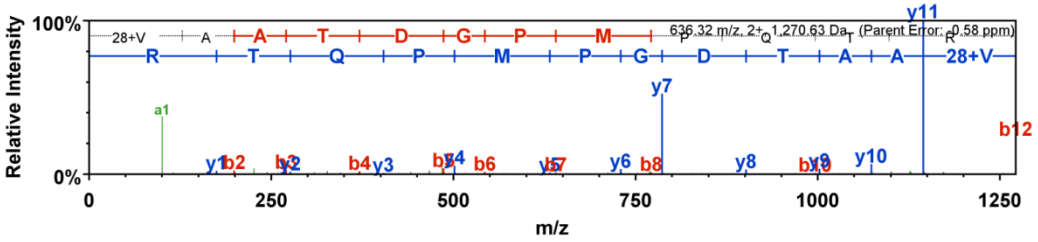
Accession: Q4A9G1

Amino acids	m/z	z	Error (ppm)	Peptide	Score
2-14	788.98	2	4.1	M.AVVKTTGKKDFDR.S + Dimethyl (N-term)	48.8
112-123	636.32	2	0.8	V.VAATDGPMPQTR.E + Dimethyl (N-term)	49.3
215-231	656.35	2	6.4	M.DKPFLMAVEDVFTITGR.G	53.8
316-327	709.93	2	4.4	A.AIYALKKEEGGR.H + Dimethyl (N-term)	50.1

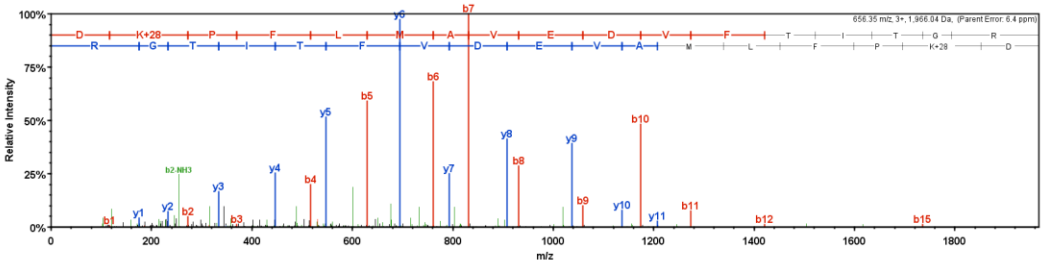
Amino acids 2-14: M.AVVKTTGKKDFDR.S + Dimethyl (N-term)



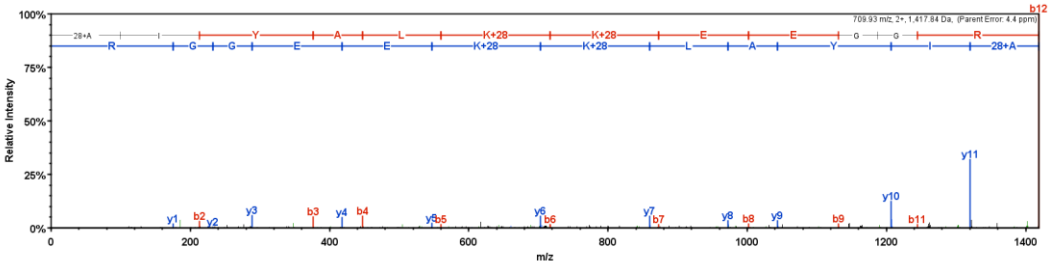
Amino acids 112-123: V.VAATDGPMPQTR.E + Dimethyl (N-term)



Amino acids 215-231: M.DKPFLMAVEDVFTITGR.G



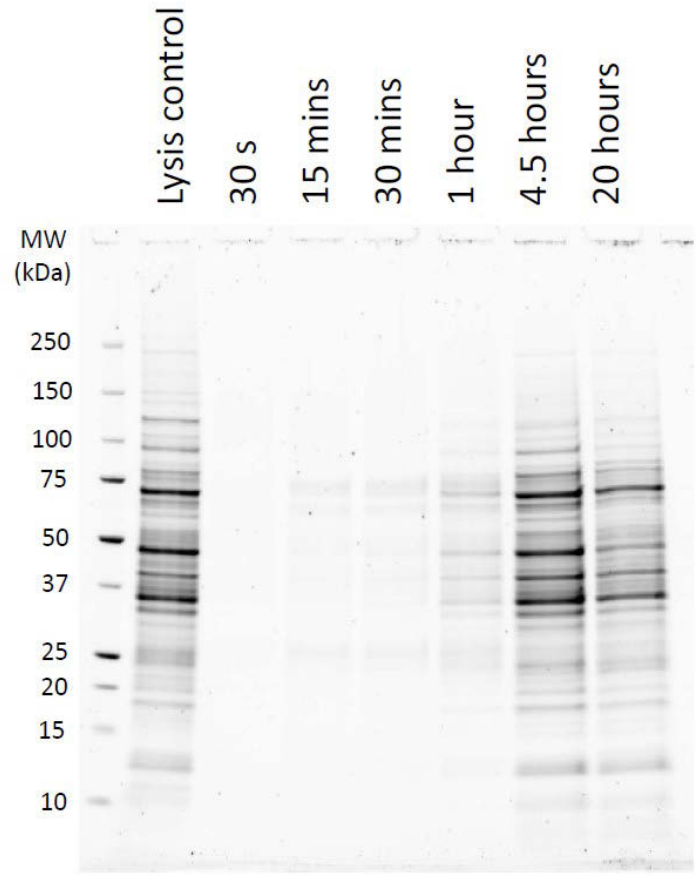
Amino acids 316-327: A.AIYALKKEEGGR.H + Dimethyl (N-term)



Chapter 4: Supporting information

Figure S4.1. Cell lysis controls showing secreted proteins over time.

1D SDS-PAGE gel image showing concentrated equal volumes of secreteomes of cells incubated in PBS at 37°C on a rotating wheel for time periods as indicated.

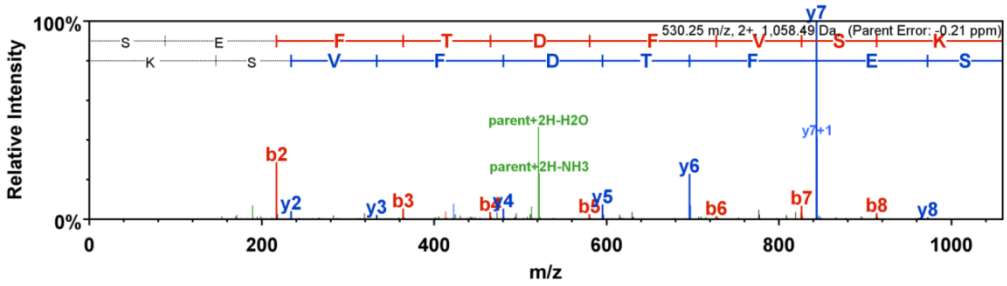


Chapter 5: Supporting information

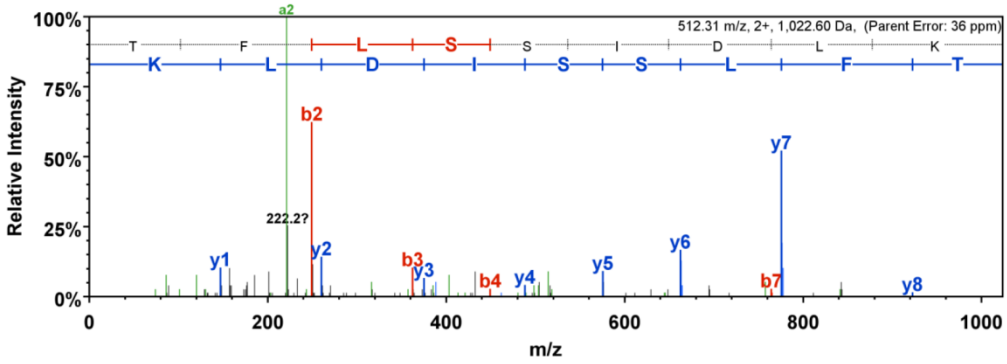
Figure S5.1. Spectra supporting the identification of semi-tryptic peptides presented in table 5.1.

Mascot data files complied in Scaffold and reanalysed by X!Tandem.

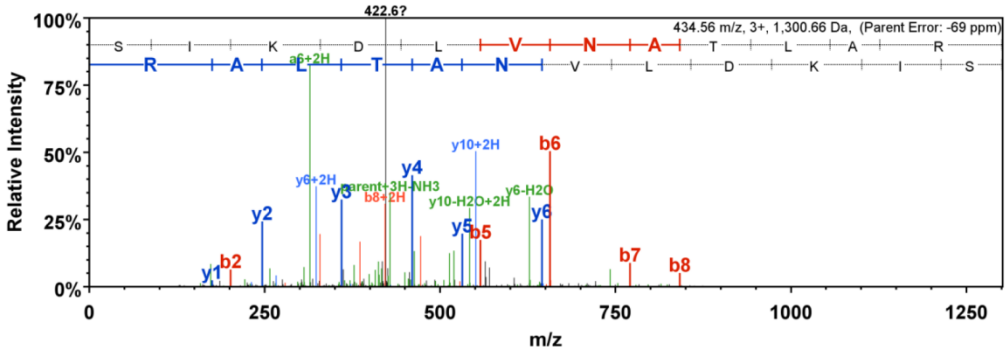
Semi-tryptic peptide 1: ^{83}SEFTDFVSK
Mascot Score: 48.3



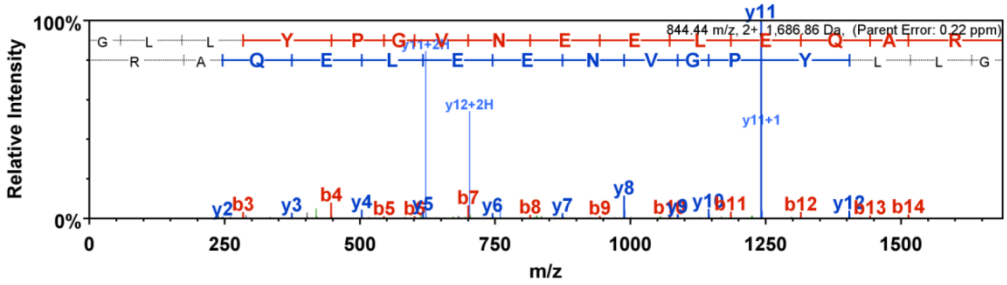
Semi-tryptic peptide 2: ^{297}TFLSSIDLK
Mascot Score: 47.3



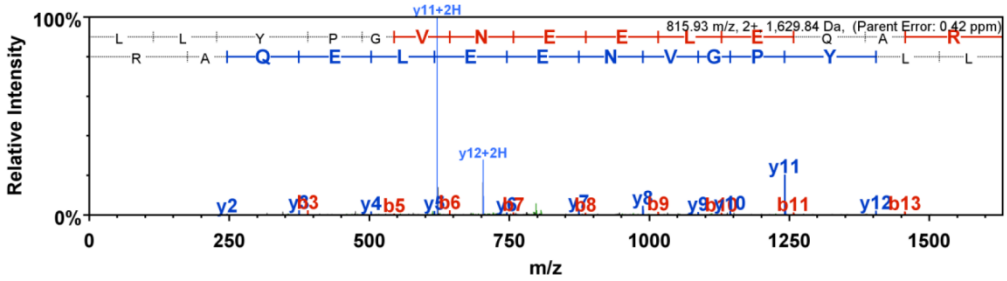
Semi-tryptic peptide 3: ^{398}SIKDLVNATLAR
Mascot Score: 40.1



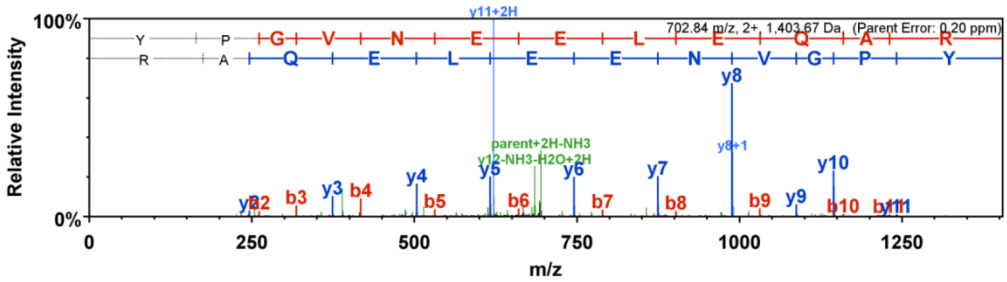
Semi-tryptic peptide 4: [496]GLLYPGVNEELEQAR
Mascot Score: 72.3



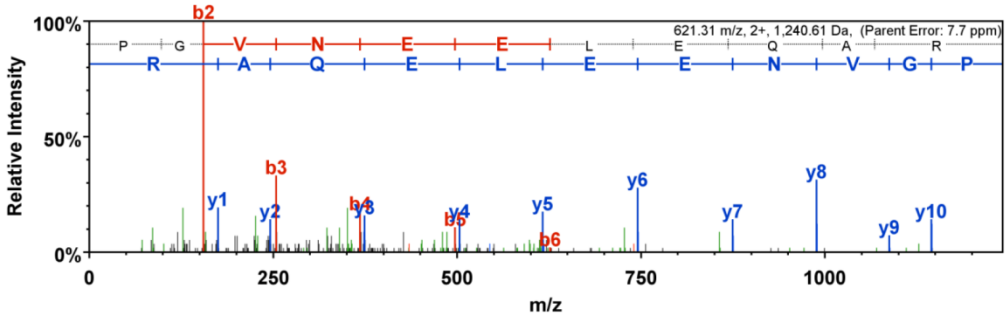
Semi-tryptic peptide 4: [497]LLYPGVNEELEQAR
Mascot Score: 66.4



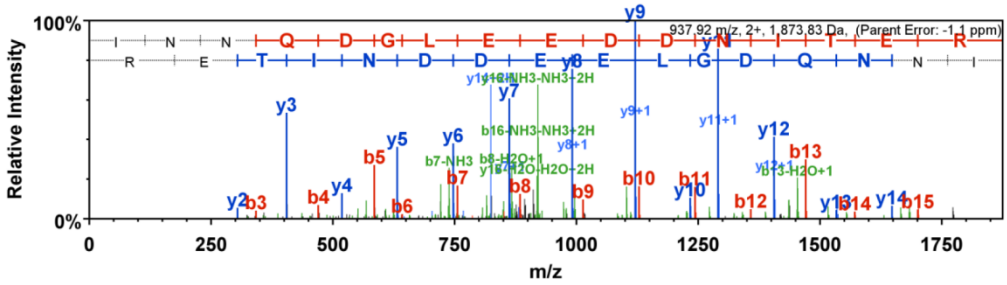
Semi-tryptic peptide 4: [499]YPGVNEELEQAR
Mascot Score: 67.8



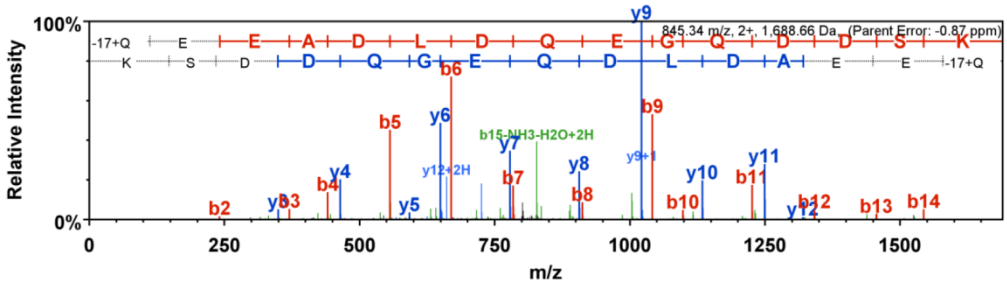
Semi-tryptic peptide 4: [500]PGVNEELEQAR
Mascot Score: 73.3



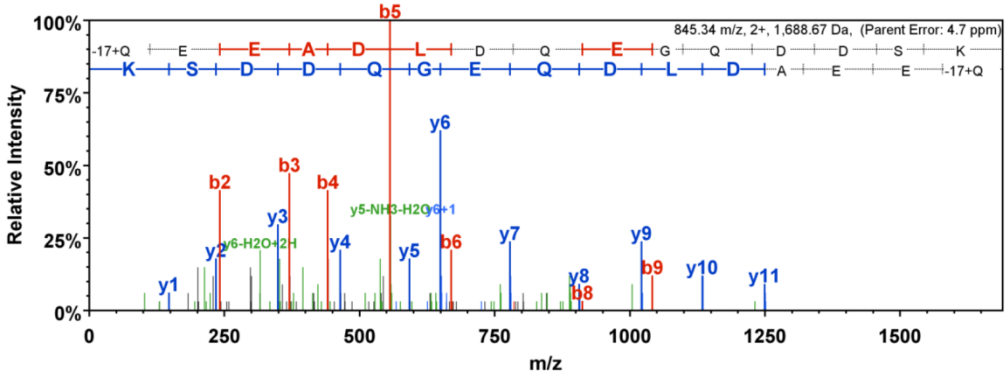
Semi-tryptic peptide 5: ^{540}INNQEGLEEDDNITER
Mascot Score: 93.1



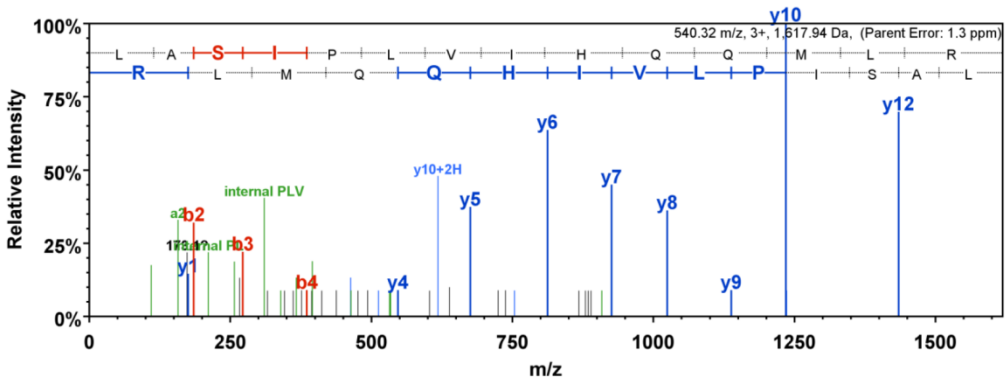
Semi-tryptic peptide 6 (Dominant cleavage site): ^{1074}QEEADLDQEGQDDSK
X!Tandem score: 13.46



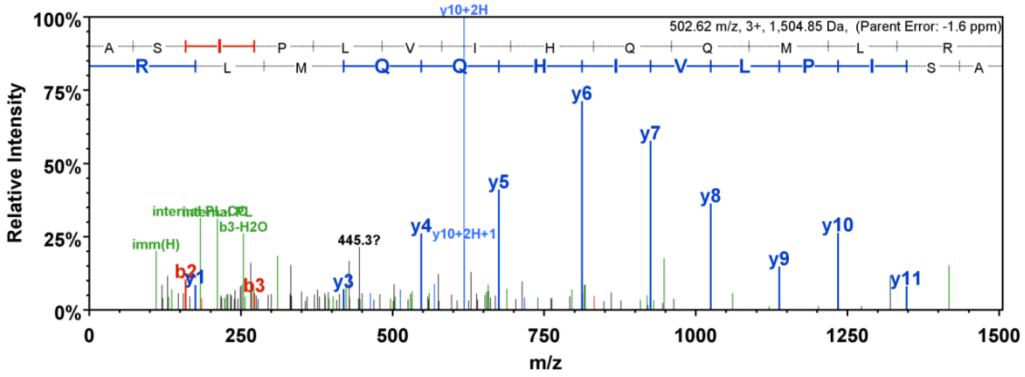
Semi-tryptic peptide 6 (Dominant cleavage site): ^{1074}QEEADLDQEGQDDSK
Mascot Score: 77.4



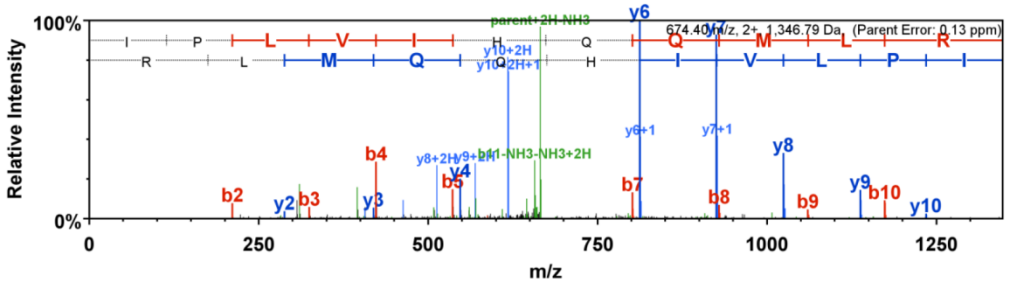
Semi-tryptic peptide 7: ^{1228}LASIPLVIHQQMLR
Mascot Score: 47.1



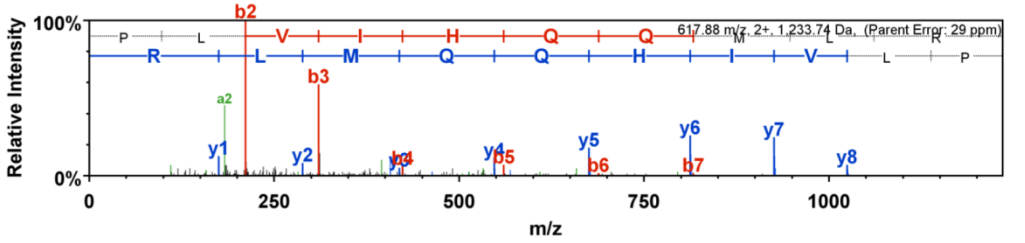
Semi-tryptic peptide 7: ^{1229}ASIPLVIHQQMLR
Mascot Score: 39.0



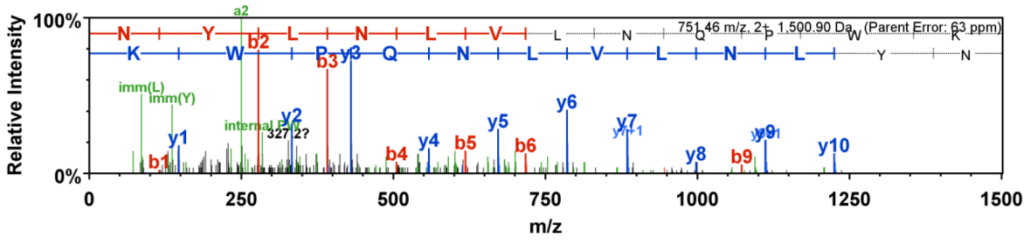
Semi-tryptic peptide 7: ^{1231}IPLVIHQQMLR
Mascot Score: 41.6



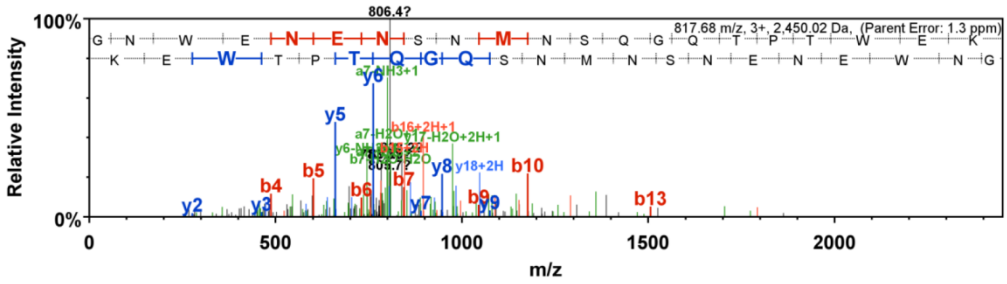
Semi-tryptic peptide 7: ^{1232}PLVIHQQMLR
Mascot Score: 51.0



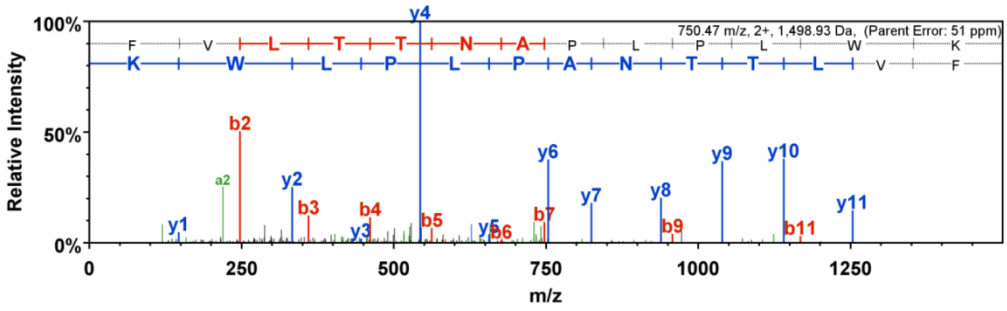
Semi-tryptic peptide 8: ^{1562}NYLNLVNQPWK
Mascot Score: 69.3



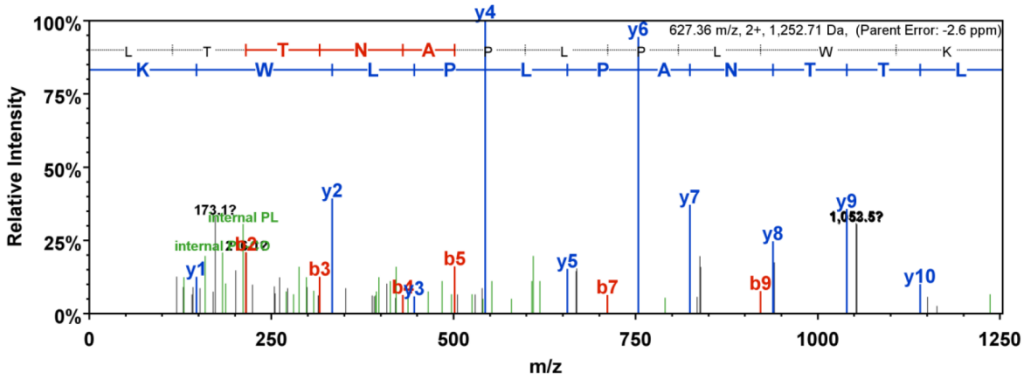
Semi-tryptic peptide 9: ^{1685}GNWENSSMNSQAQTPTWEK
X!Tandem Score: 5.64



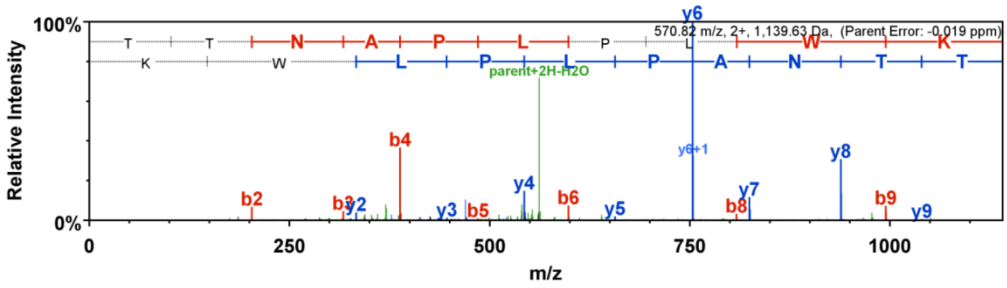
Semi-tryptic peptide 10: ^{1726}FVLTTNAPLPLWK
Mascot Score: 78.1



Semi-tryptic peptide 10: ^{1728}LTTNAPLPLWK
Mascot Score: 54.1



Semi-tryptic peptide 10: ^{1729}TTNAPLPLWK
Mascot Score: 45.0



Semi-tryptic peptide 11: RLMNTPITF[1346]
Mascot Score: 45.2

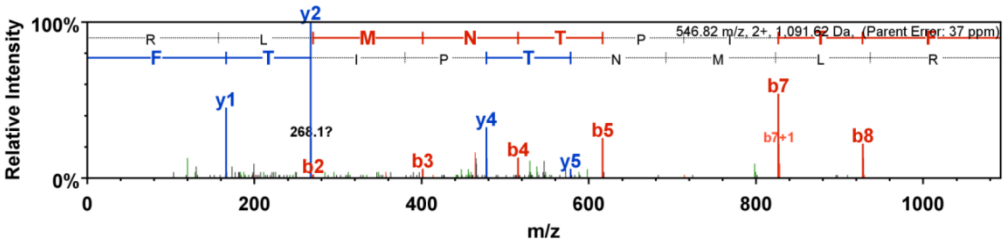
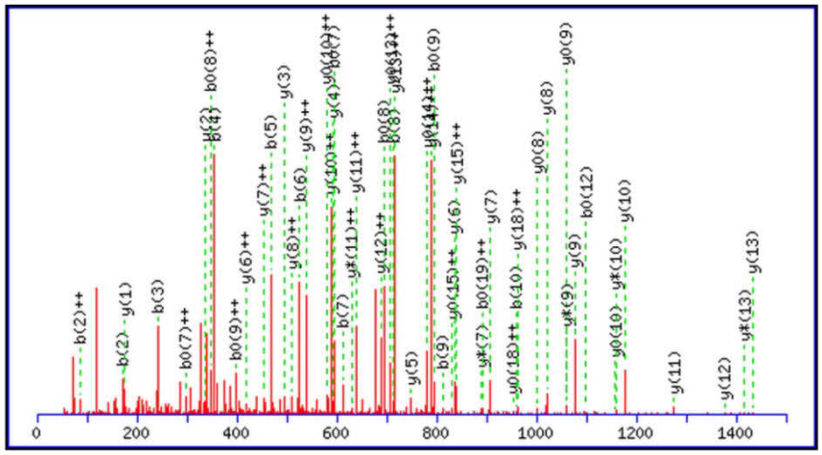


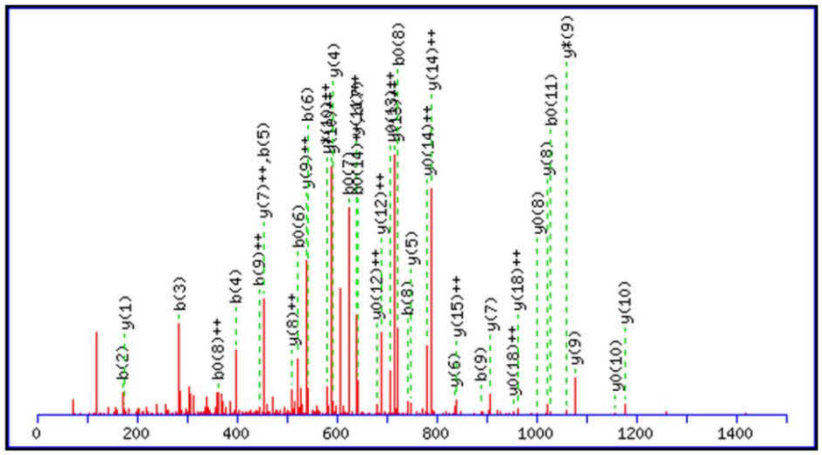
Figure S5.2. Spectra supporting the identification of N-terminal dimethyl-labelled peptides presented in figure 5.6.

^{13}AAAIIGSTVFGTVVGLASKVKYR



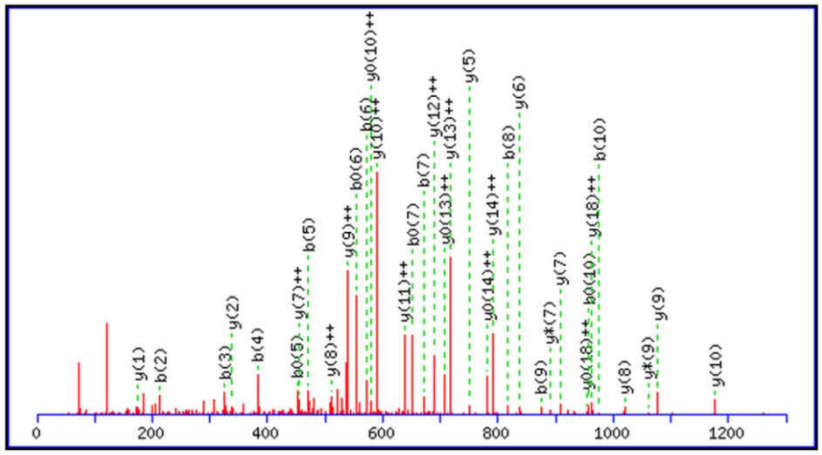
Monoisotopic mass of neutral peptide Mr(calc): 2391.4253

^{14}AAIIGSTVFGTVVGLASKVKYR



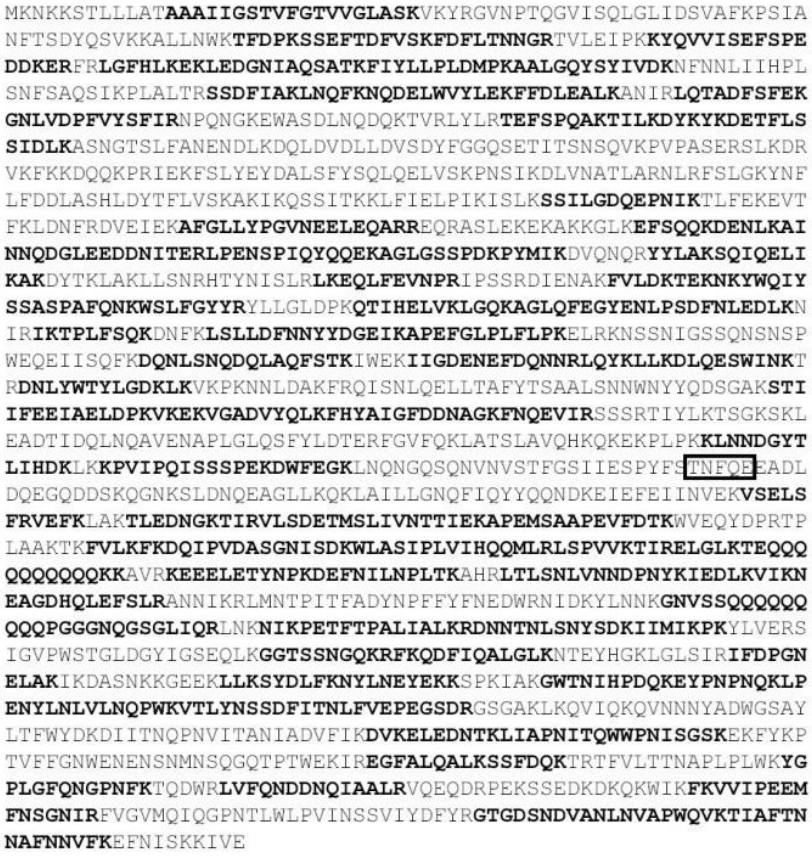
Monoisotopic mass of neutral peptide Mr(calc): 2320.3882

^{15}AIIGSTVFGTVVGLASKVKYR



Monoisotopic mass of neutral peptide Mr(calc): 2249.3511

Figure S5.3. Peptide coverage to intact P216$_J$ identified in figure 5.5.



Identified peptides in bold. Dominant cleavage site boxed.

Appendix II – Electronic supporting information

To improve readability of very large and detailed tables and figures, the following supplementary information is provided in electronic format can be found on the included DVD.

Chapter 3: Electronic supporting information

Supplementary file S3.1. Protein identifications from global proteome analyses.

Supplementary file S3.3. Identifications of gel spots from pH 4-7 2D gels.

Supplementary file S3.4. Identifications of gel spots from pH 6-11 2D gels.

Chapter 4: Electronic supporting information

Table S4.1. Identified surface proteome of M. hyopneumoniae *strain J.*

Appendix II – Electronic supporting information

Contents:

File name: Supp file S3.1.xlsx
Supplementary file S3.1. Protein identifications from global proteome analyses.
Excel workbook file containing protein identifications presented in Chapter 3.
Sheet 1 "Proteins Identified" provides an overview to protein identifications by various methods. Sheet 2 "Peptide report" provides mass spectrometry data for the identifications at the peptide level.

File name: Supp file S3.3.xlsx
Supplementary file S3.3. Identifications of gel spots from pH 4-7 2D gels.
Excel workbook file containing protein identifications from pH 4-7 2D gels.
Sheet 1 "Reference gel" provides an image of the reference gel with excised spot numbers indicated. Subsequent numbered sheets correspond to protein spot numbers and provide protein and peptide identification from mass spectrometry analysis.

File name: Supp file S3.4.xlsx
Supplementary file S3.4. Identifications of gel spots from pH 6-11 2D gels.
Excel workbook file containing protein identifications from pH 6-11 2D gels.
Sheet 1 "Reference gel" provides an image of the reference gel with excised spot numbers indicated. Subsequent numbered sheets correspond to protein spot numbers and provide protein and peptide identification from mass spectrometry analysis.

File name: Table S4.1.xlsx
Table S4.1. Identified surface proteome of M. hyopneumoniae *strain J.*
Excel workbook file containing surface protein identifications presented in Chapter 4.