ithreeinstitute

Expanding the Functional Proteome of *Mycoplasma pneumoniae*

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

2018

Michael Widjaja

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Certificate of Original Authorship and Declaration

I, Michael Widjaja, 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. This research is supported by an Australian Government Research Training Program Scholarship.

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.

For the following, I, Michael Widjaja, declare that more than 50% of the content in the journal article (Chapter 4) and manuscripts (Chapters 5 and 6) included in this thesis:

- Planning of each chapter was conducted by my supervisors and I.
- Experiments performed and analysed by myself (unless stated).
- The initial draft prepared and written by myself. Any subsequent changes in response to co-authors and editor's reviews were performed by me.

Signature:

Production Note:
Signature removed prior to publication.

Date:

20/2/2018

Acknowledgements

First, I would like to thank the markers for taking the time to review this thesis. I would also like to thank the Australian Commonwealth Government Department of Industry for the Australian Postgraduate Award Scholarship I received throughout my PhD candidature.

I would like to thank Prof. Duncan Krause and Assoc. Prof. Daniel Brown for donating Mycoplasma antibodies for my research. Thank you to Dr Marie Trussart, Dr Maria Lluch-Senar, and Prof. Luís Serrano for the opportunity to collaborate with your team to investigate the structure of the *M. pneumoniae* chromosome. It was a pleasure to work with you all and I am glad I had the opportunity to do and the outcome. I would also like to extend this gratitude to Dr Roger Dumke and Dr Anne Gründel, thank you both for the Mycoplasma antibodies and the collaboration on Ef-Tu. I hope we can collaborate again in the future.

A big thank you to all the International Organization for Mycoplasmology (IOM) members for the discussions (science related and not), laughs, and fun times. Especially Art Totten, Assoc. Prof. Meghan May, Prof. Mitch Balish (also for the sialic acid column suggestion), Dr Peter Kuhnert, Prof. Joachim Frey, Dr Isil Tulum, Dr Yuhei Tahara, and Steven Distelhorst. I hope to see you all at the IOM 2018!

Back in Australia I would like to thank Dr Cheryl Jenkins for all her help at EMAI with jabbing bunnies and the antibodies. Thanks to the Microbial Imaging Facility at UTS especially Dr Lynne Turnbull for all the troubleshooting, training, and analysis; in addition to Dr Mike Johnson and Assoc. Prof. Cynthia Whitchurch. Like to thank the tech staff (Harry Simpson, Sarah Osvath, Mercedes Ballesteros, and Luke Beebe), because nothing would happen without any of you!

Thank you to the senior proteomic-ists. Jerran for showing me how to model and telling Steve how good of a student I am and to take me on as an Honours student. Thank you Jess for mentoring me at the start of Honours and PhD and being an inspiration in presentations and communication. On the matter of communication, big thank you to Prof. Liz Harry for your support in helping me build my Science Communication experience, I look forward to pursuing this post-PhD.

I have of course have to thank my supervisors. Thank you Prof. Steven Djordjevic, for taking me on for Honours and especially for a PhD, and for your supervision in the past six years. Huge thank you to Dr. Matt Padula for training and helping me in the lab, you are a wealth of knowledge and an inspiration to all the students in the lab. To my mentor, friend, guy-that-I-stand-next-to-so-I-look-bigger, gym buddy, and now co-supervisor Dr. Benjamin Armando Bernard Raymond, thank you for everything. My students that I have mentored: Kayla, Angus, Harui, and Andy thank you all for giving me the opportunity to mentor and for your commitment to your projects.

Thanks to my PhD buddies. Ka-tey, was good to have someone to share the PhD pain and stresses with on a daily basis so thank you for relating. Krishy, thank you for the great laughs and all the free food! Thanks to the rest of the PhD cohort for the laughs especially Megan, Jacqueline, Greg, Dan, Em, Nat, and Brendan. Big thanks to Samira, my Metabolic buddy, for talking me up and getting me jobs!

Thank you to mum and dad, without you both I wouldn't be where I am today. Also to Om Ajun, I Nil Nil, Maddy, Dede, and Sam for your support! And to my rabbits, Cookie and Diglett, for always being a source of luck! Thanks to all my Munyaks especially Alex, Shaun, Oz, Nick, Ray, TT, and Baba (also with the image design help). Thank you all for your words of motivation and support of how proud you guys are of me.

Lastly the biggest thank you to Isa, my significant other. Thank you for double checking everything of mine, supporting me, and keeping me in check. I cannot think of anyone else that I would want to have spent these past few years with other than you. 1,2, toilet seat!

If I have missed anyone, apologies and thank you.

Publications

1. Berry I.J.[†], Jarocki V.M.[†], Tacchi J.L., Raymond B.B., **Widjaja M.**, Padula M.P., and Djordjevic S.P. (2017) 'N-terminomics identifies widespread endoproteolysis and novel methionine excision in a genome-reduced bacterial pathogen'. *Scientific Reports* (Submitted).

2. (Chapter 5): **Widjaja M.[†]**, Harvey K.L.[†], Hagemann L.[†], Berry I.J., Jarocki V.M., Raymond B.B., Tacchi J.L., Gründel A., Steele J.R., Padula M.P., Charles I.G., Dumke R.[¥], and Djordjevic S.P.[¥] (2017). 'Elongation factor Tu is a multifunctional and processed moonlighting protein'. *Scientific Reports* (Submitted).

3. Trussart M., Yus E., Martinez S., Baù D., Tahara Y.O., Pengo T., **Widjaja M.**, Kretschmer S., Swoger J., Djordjevic S., Turnbull L., Whitchurch C., Miyata M., Marti-Renom M.A., Lluch-Senar M., and Serrano L. (2017). 'Defined chromosome structure in the genome-reduced bacterium *Mycoplasma pneumoniae*'. *Nature Communications*.

4. Tacchi J.L., Raymond B.B., Haynes P.A., Berry I.J., **Widjaja M.**, Bogema D.R., Woolley L.K., Jenkins C., Minion F.C., Padula M.P., and Djordjevic S.P. (2016). 'Post-translational processing targets functionally diverse proteins in *Mycoplasma hyopneumoniae*'. *Open Biology*.

5. (Chapter 4): **Widjaja M.**, Berry I.J., Pont E.J., Padula M.P., and Djordjevic S.P. (2015). 'P40 and P90 from Mpn142 are targets of multiple processing events on the surface of *Mycoplasma pneumoniae*'. *Proteomes*.

6. 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., and Djordjevic S.P. (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'. *Cellular Microbiology*.

[†]shared primary author,[¥]shared principal investigators

Abbreviations

Standard abbreviations

Adenine and thymine	A+T
Community Acquired Respiratory Distress Syndrome	CARDS
Dihydrolipoyl dehydrogenase	Pdh-D
Dihydrolipoyllysine acetyltransferase	Pdh-C
Elongation factor Tu	Ef-Tu
Glyceraldehyde 3-phosphate dehydrogenase	GAPDH
Guanine and cytosine	G+C
High Molecular Weight protein	HMW
Isoelectric Point	pI
Liquid chromatography tandem mass spectrometry	LC-MS/MS
Open Reading Frame	ORF
Phosphate Buffered Saline	PBS
Potential of hydrogen	pH
Pyruvate dehydrogenase E1 α subunit	Pdh-A
Pyruvate dehydrogenase E1 β subunit	Pdh-B
Short Linear Motifs	SLiMs
Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis	SDS-PAGE

Unit abbreviations

Base pair	bp	Molar	M
Dalton	Da	Mole percent	Mol%
Gram	g	Percent	%
Hour	h	Percentage weight per volume	w/v
Litre	L	Percentage weight per weight	v/v
Metre	m	Relative centrifugal force	RCF
Minute	min	Second	s

Table of Contents

Table of Figures

Table of Tables

Thesis Preface

The growing incidence of antibiotic resistance globally is a significant public health issue and as previously susceptible bacteria continue to develop resistance, we need to develop novel strategies to counter this trend. *Mycoplasma pneumoniae* is a genome reduced bacteria that is one of the major causes of bacterial pneumonia in close contact settings such as schools and hospitals. Children, the elderly, and the immuno-suppressed are commonly infected due to an under developed or impaired immune system. A successful vaccine against this respiratory pathogen is yet to be developed and treatment options are limited. Additionally, children are limited to one class of antibiotics due to the permanent side effects of other agents.

Antibiotic resistance within *M. pneumoniae* was detected over a decade ago and has now spread to most of the Northern Hemisphere. Though infections are not typically fatal, *M. pneumoniae* can cause secondary co-infections; some of which can be fatal. The work presented within this thesis expands the functional proteome of *M. pneumoniae*, with the goal of discovering potential novel therapeutic or vaccine targets. This was initially achieved by examining the full repertoire of proteins exposed on the surface of *M. pneumoniae*. This thesis then addresses which host antigens these proteins potentially interact with during infection. Although a single protein was not chosen as a vaccine target, the result of the work presented here report a list of potential targets that participate in the colonisation of the respiratory epithelium. This thesis highlights that the interactions between *M. pneumoniae* and host epithelium are complex, and involve a wide range of diverse proteins.

This thesis begins with an introduction to *M. pneumoniae* and what is currently known about the proteins involved during the interaction with the human host.

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Chapter 1.
Introduction on *Mycoplasma pneumoniae* and its interaction with the human host

i3 ithreeinstitute

1.1 Introduction to Mycoplasma pneumoniae

M. pneumoniae is one of the major causes of community acquired pneumonia primarily affecting children, but can also affect the elderly and the immuno-suppressed[1]. It is responsible for up to 40% of community acquired pneumonia in children and 18% of cases require hospitalisations worldwide[1,2]. This respiratory pathogen is implicated to exacerbate asthma in infected patients, though the exact role is yet to be determined[3]. *M. pneumoniae* epidemics are cyclic occurring every 4 – 7 years[4-8]. Climate and season have been ruled out as factors associated with the different epidemic trends between countries[9,10]. Unlike the typical pneumonia produced by S*treptococcus pneumoniae*, *M. pneumoniae* causes an atypical pneumonia called 'walking pneumonia' where hospitalisation is often required, albeit rarely in adults[1]. Infections are transmitted by airborne droplets enabling *M. pneumoniae* to easily enter and colonise the respiratory tract. Due to this ease of transmission, *M. pneumoniae* outbreaks tend to occur more frequently in close contact settings such as schools[11], military barracks[12,13], hospitals[14,15], inpatient institutions[16,17], closed communities[18], and religious settings[19,20]. Once *M. pneumoniae* has adhered to the respiratory epithelium, an array of virulence factors come into play are, targeting the mucociliary escalator and perturbing mucociliary clearance by removing cilia from the epithelium[21]. Although *M. pneumoniae* infections are rarely fatal[22], removal of the cilia can cause secondary co-infections, some of which can be fatal[23]. Only a limited number of antibiotics can be used to treat Mycoplasmal diseases including macrolides, tetracyclines, or quinolones[24]. Macrolides are the therapy of choice to treat infections in children[25]. Resistance to front line antibiotics such as macrolides, began to emerge in 2001[26] and has spread to most parts of the globe. In addition, *M. pneumoniae* also has the ability to change its surface protein topography conceivably to deceive and evade the host immune system[27-32]. With the continued rise of resistance, treatment options are limited and there is a clear need for a successful vaccine.

In addition to studies that seek to find novel vaccine targets, *M. pneumoniae* is gaining popularity as a model organism in different scientific disciplines, particularly those concerning systems biology. The first *M. pneumoniae* genome sequence was completed in 1996 by Himmelreich et al., and was found to be 816,394 bp, encoding 689 predicted open reading frames (ORFs)[33,34]; which has now been extended to 694 ORFs[35]. Three reference *M. pneumoniae* genomes have been sequenced: the M129 strain[33], the FH

strain[36], and the Roo3 strain[37]. Clinical *M. pneumoniae* isolates have also been sequenced[38,39]. Studies by Spuesens et al. found distinct genetic differences between the reference and clinical strains[37]. The transcriptome[40,41], proteome[42-47], metabolome[48], protein interactome[49], acetylation proteome[50], phosphoproteome[50-52], and the DNA methylome[53] have been studied in *M. pneumoniae*. In addition there are studies that have investigated minimal ORFs essential for life[54], as well as transcriptional coordination[55]. All these datasets have been collated into an accessible online database <http://mympn.crg.eu/>[56]. However, despite all these systems biology studies, a vaccine against *M. pneumoniae* infections is yet to be developed. An understanding on how this pathogen interacts with the respiratory epithelium and engages the immune system is considered fundamental for the development of efficacious vaccines and therapeutic targets to counter *M. pneumoniae*.

M. pneumoniae belongs to the Mollicutes class and is the most widely recognised and studied member[1]. Mollicutes distinguish themselves from eubacteria that express a functional cell wall by a process known as degenerative evolution. Mollicutes lack a cell wall and have lost the genes needed for the biosynthesis of amino acids, fatty acids, co-factors, and vitamins[1]. Instead these biosynthetic metabolites are recovered from their host during infection[57]. *M. pneumoniae* is a flask-shaped organism (Figure 1.1) with an elongated cellular extension called the attachment organelle[58] (covered later in this review). This attachment organelle provides motility and allows *M. pneumoniae* to successfully colonise the human respiratory tract.

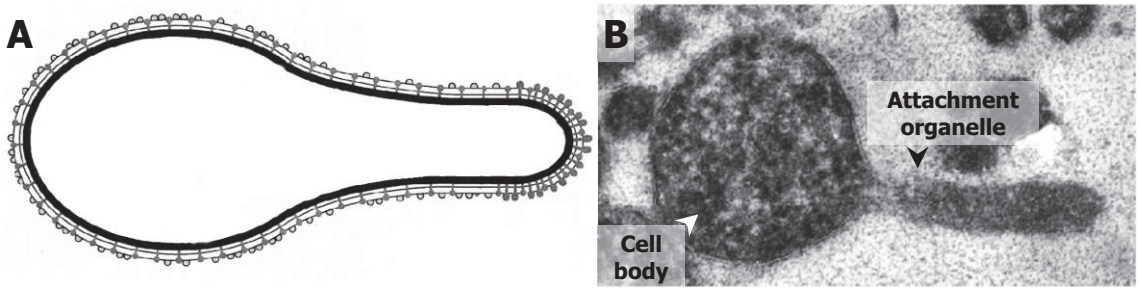


Figure 1.1: *M. pneumoniae* structure. A) Schematic of flask cell shape of *M. pneumoniae* cell (Adapted from[59]). B) Ultrathin section of an intact *M. pneumoniae* cell (Adapted from[45]). Cell body and attachment organelle are indicated.

1.2 Mycoplasma pneumoniae disease

M. pneumoniae symptoms commonly appear in children, the elderly, and the immuno-suppressed due to an under developed or impaired immune system. Typical symptoms include coughing, gradual phlegm development, chills, sore throat, fever, and general malaise[1]. Clinical cases are divided into two groups, Type 1 and Type 2 infections, and the features that define these infection types will be discussed later in this chapter[60,61]. Studies have shown that *M. pneumoniae* is not confined to the human respiratory tract. Extrapulmonary infections are well described and can involve cardiovascular, dermatological, digestive, haematological/hematopoietic, musculoskeletal, neurological, respiratory, sensory organ, and urogenital tract sites (reviewed in[62]). *M. pneumoniae* infections have been suggested to exacerbate asthma[3,63-65] and numerous reports describe the isolation of *M. pneumoniae* from asthmatic patients[66,67]. Common diagnostic techniques used to detect *M. pneumoniae* infection include a combination of serology and Polymerase Chain Reaction[68-72], microbiological culture is not ideal due to *M. pneumoniae*'s slow growth rate[73].

The current treatment of *M. pneumoniae* infections involves the use of macrolides, tetracyclines, or quinolones[24]. Macrolides are the primary option for children due to the undesirable side effects afforded by the other two antibiotics[25]. β-Lactam antibiotics including penicillin, which inhibit cell wall synthesis, are ineffective as *M. pneumoniae* lacks a cell wall[69]. The first instance of macrolide resistant *M. pneumoniae* was recorded in 2001 in Japan[26] and this resistance has now spread to other parts of Asia, Europe, and North America[74-88].

Efficacious vaccines are yet to be developed to counter a *M. pneumoniae* infection, and previous attempts have exacerbated symptoms of disease[89,90], a result stemming from a lack of knowledge of virulence factors, pathogenesis, and how the pathogen engages the host immune response (reviewed in[91]). Inactivated *M. pneumoniae* vaccines generated in the 1960s and 1970s only achieved a reduction in infection cases and reduced symptoms of pneumonia by ~40% without any severe adverse side effects[91]. Over the past several years *M. pneumoniae* vaccines have targeted what are considered to be three major adhesins that play an important role in adherence to epithelial cell receptors.

Immunological cross reactions and exacerbations of pathological symptoms are a major limitation of many immunisation trials and further studies are urgently needed[90,92-96].

1.3 Mycoplasma pneumoniae pathogenicity

Once *M. pneumoniae* has attached to the human bronchial epithelium, host cellular damage begins by the release of several virulence factors including (a pathway is hypothesised by[97]): superoxide and hydrogen peroxide radicals[98,99], the HapE protein to induce haemolysis[100], and an ADP-ribosylating and vacuolating toxin named Community Acquired Respiratory Distress Syndrome (CARDS) toxin[101,102]. However, for this to occur, it is essential that *M. pneumoniae* first establishes adherence to the human respiratory epithelium[103]. Scanning electron microscopy has been used to monitor how *M. pneumoniae* colonises the epithelial cells (Figure 1.2)[21]. Muse et al. visualised hamster tracheal rings that were infected with *M. pneumoniae.* After 72 hours, the host epithelium was partially stripped/denuded of cilia (Figure 1.2D) and tracheal rings were almost completely denuded of cilia 96 hours after infection (Figure 1.2E)[21].

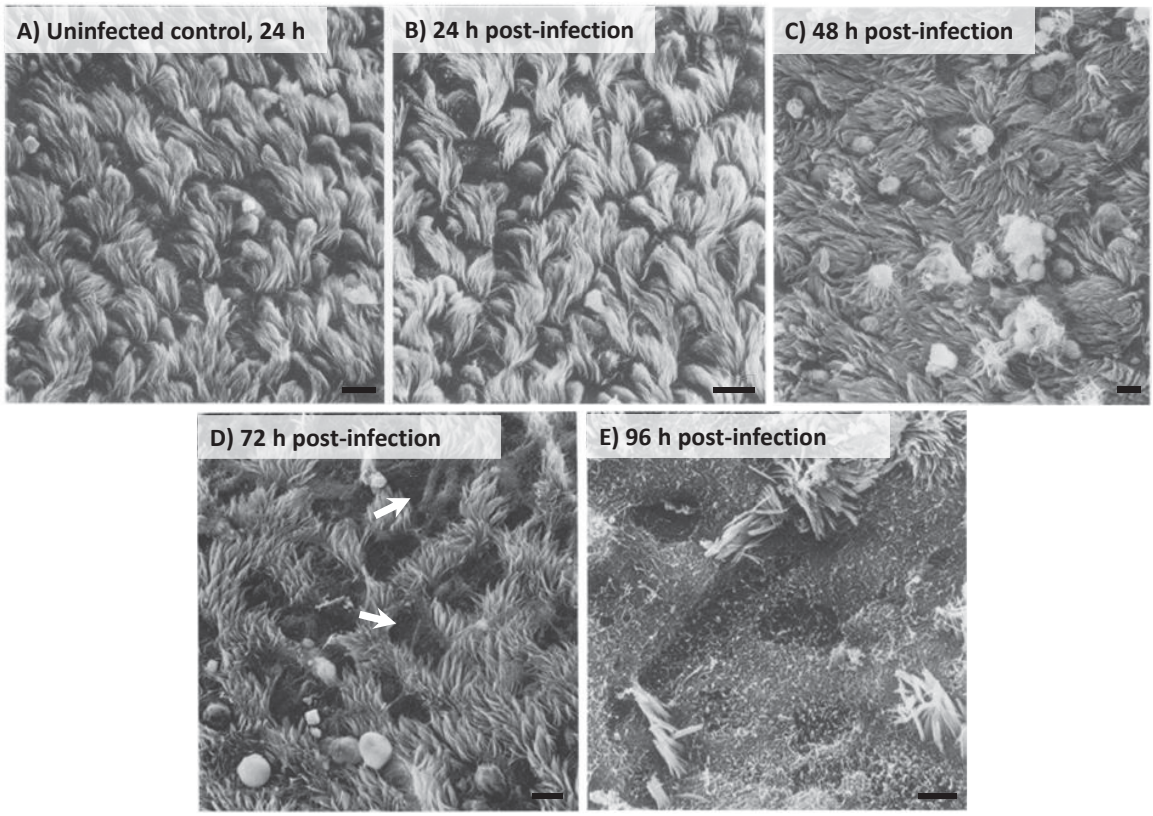


Figure 1.2: Scanning electron microscopy of the progress of pathogenicity caused by *M. pneumoniae* (not visible). A) Uninoculated hamster trachea. B) Hamster trachea 24 hours post-infection with *M. pneumoniae*, C) hamster trachea 48 hours post-infection with *M. pneumoniae*, D) hamster trachea 72 hours post-infection, and E) hamster trachea 96 hours post-infection. Regions denuded of cilia are indicated by arrow in D). Scale bar: 5 µm, adapted from[21].

CARDS toxin was discovered to be a major factor responsible for deterioration of cilia[102]. This surface associated toxin (Figure 1.3) is an ADP-ribosylating, vacuolating toxin that is only secreted after *M. pneumoniae* binds to the bronchial epithelium; once again highlighting the importance of initial adhesin-host interactions[104]. Once secreted into the extracellular milleu, CARDS toxin induces tissue disorganisation, inflammation, vacuolisation, reduced oxygen and amino acid uptake by host cells, and a reduction in glucose metabolism[102,105-109].

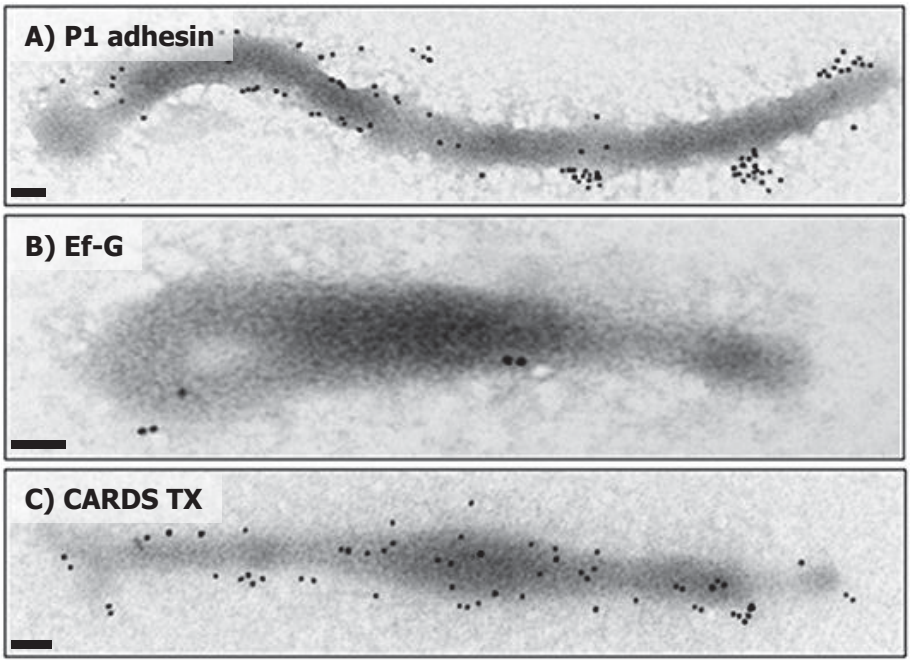


Figure 1.3: Immunogold electron microscopy of antibodies targeting CARDS TX on the surface of *M. pneumoniae*. A) Gold labelled antibodies against the P1 adhesin served as an extracellular control. B) Gold labelled antibodies against elongation factor G (Ef-G) served as an intracellular control. C) Gold labelled antibodies against CARDS TX were distributed across the whole cell. Scale bar = 0.1 µm, adapted from[104].

Phenotypic plasticity is seen in *M. pneumoniae* in two key immunogenic genes (*mpn141* and *mpn142*) which contain a stretch of repetitive DNA sequences called RepMP elements[110-112]. These RepMP elements are homologous to other RepMP elements in the *M. pneumoniae* genome and together comprise about 8% of the entire genome highlighting the importance of these elements for survival[33,111]. Genetic recombination with different RepMP elements provides *M. pneumoniae* with a mechanism for surface antigenic variation. This has been described in three key surface proteins (P1, P40, and P90)[27-32]. It is the pattern of variation of these RepMP elements that subdivides *M. pneumoniae* strains into the Type 1 and Type 2 groups[32]. A 10-year study in Japan found that *M. pneumoniae* cells shift between the two types every 8 – 10 years requiring at least 3 years for a subtype to become dominant[113]. Once it is dominant, it remains so for 7 years[113]. Studies have looked into differences between these two *M. pneumoniae*

types and found that they differ in: biofilm formation[114], hemadsorption inhibitory activity[115], and CARDS toxin production[38,116].

1.4 Structural aspects of *Mycoplasma pneumoniae* pathogenicity

M. pneumoniae possesses a 300 nm cellular protrusion named the attachment organelle (also called the tip structure) that is required for adherence to the host respiratory epithelium (cytadherence) (Figure 1.4, indicated)[58]. In addition to adherence, this attachment organelle provides motility and cellular polarity[117-119]. Remarkably, *M. pneumoniae* glides at speeds approaching 1 µm per second[120,121] enabling it to respond to nutrient gradients, environmental stimuli, and traverse the host epithelium, despite the obstructive force generated by cilial beating (~13 hertz)[122-124]. Two mechanisms been forwarded to describe *M. pneumoniae* motility. One mechanism suggests that: the attachment organelle bends, contracts, and elongates much like an inchworm[125-127]while a second mechanism suggests that protein complexes in the attachment organelle act like "legs" that perform a catch-pull-release scenario across the surface like oars on a boat[121,128]. The attachment organelle comprises an electron dense core consisting of two parallel rods that extended down the length of the tip (Figure 1.4B). This core is connected to the internal cytoskeleton and supports the adhesins (Figure 1.4C) and the accessory proteins that localise to the adhesive tip of *M. pneumoniae*[129,130] (see '1.5 The Cytoskeleton and Electron Dense Core of *Mycoplasma pneumoniae*' in this chapter).

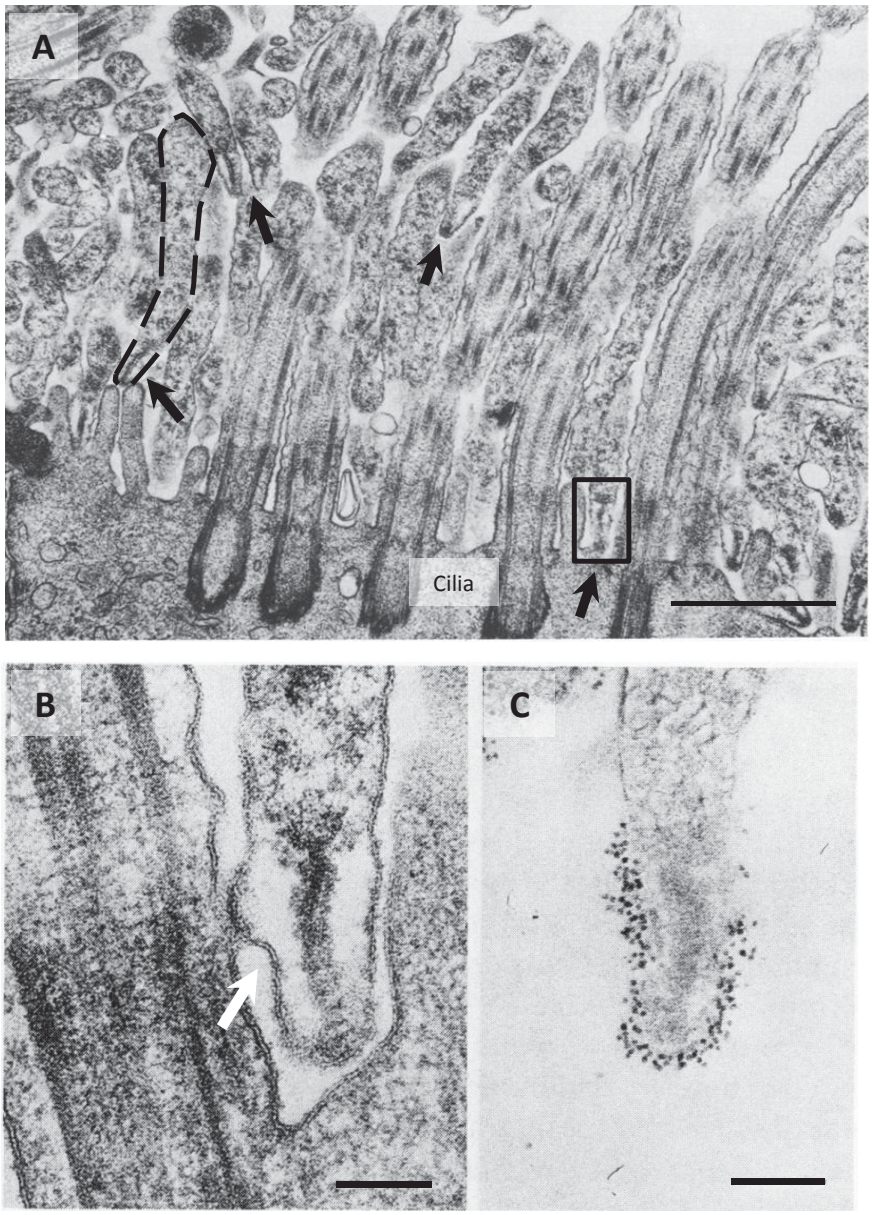


Figure 1.4: Transmission electron micrograph of *M. pneumoniae* interacting with ciliated epithelium. A) Tracheal ciliated cells infected with *M. pneumoniae* 96 hours after infection. Attachment organelles are indicated by arrows. An *M. pneumoniae* cell that has attached to the microvilli has been outlined. The boxed attachment organelle is an example of those seen in B) and C). Scale bar = 1 µm, adapted from[21]. B) Attachment organelle of *M. pneumoniae* adhering to the base of the cilia. C) Immunogold labelled antibodies targeting the P1 adhesin localise to the attachment organelle. For both B) and C): Scale bar = 0.1 µm. Image adapted from[131].

1.5 The cytoskeleton and electron dense core of *Mycoplasma pneumoniae*

Treating *M. pneumoniae* cells with Triton X-100 solubilises the cell membrane leaving behind a detergent-insoluble cytoskeleton. This Triton insoluble network of proteins was termed the "Triton shell"[129]. Electron micrographs of the Triton shell revealed a cytoskeleton-like structure with a protruding electron dense core where the attachment organelle would normally be found (Figure 1.5)[58,118,129]. The electron dense core can be divided into three components: a bowl complex that leads from the internal cytoskeleton, a parallel pair of plates (one thicker than the other), and a terminal button (Figure 1.6)[59]. It is the contraction and expansion of this electron dense core that drives *M. pneumoniae* motility in either of the proposed models mentioned above[128].

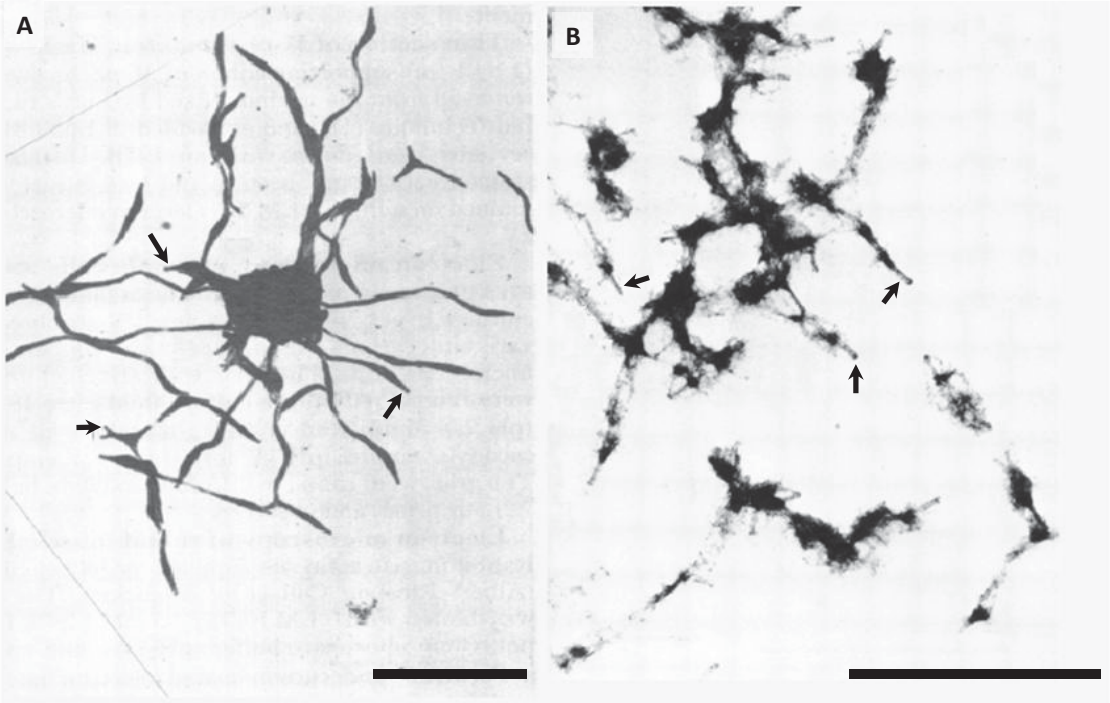


Figure 1.5: Electron micrographs of *M. pneumoniae* microcolonies before (A) and after (B) treatment with Triton X-100. The electron dense core of the attachment organelle is indicated. The cell membrane and the contents of the cytoplasm are removed leaving a network of detergent-insoluble filaments. Scale bar = 2.5 µm, adapted from[129].

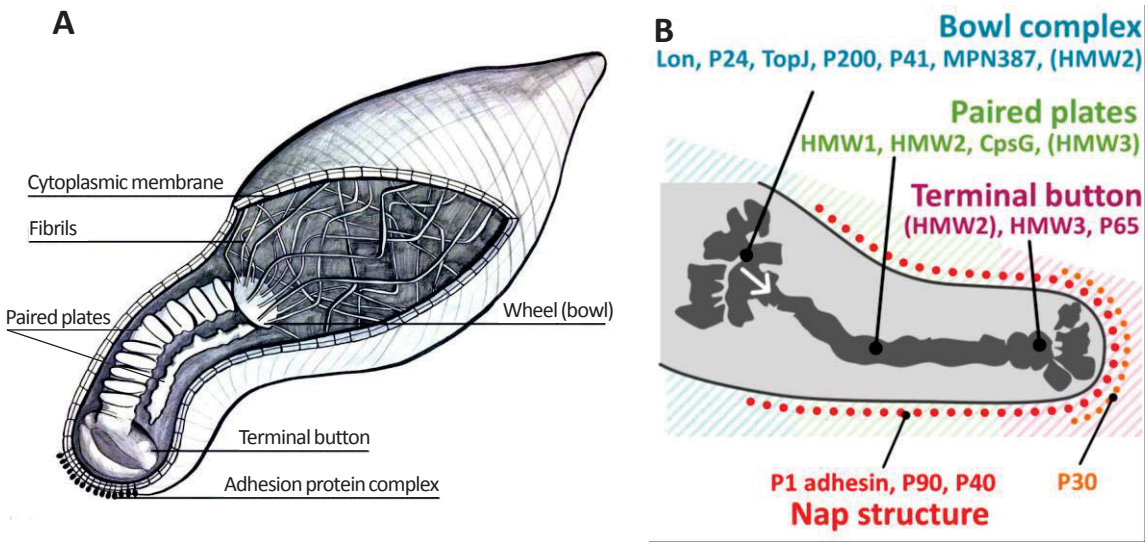


Figure 1.6: Schematics of the electron dense core of *M. pneumoniae* and its three components. A) Basic schematic highlighting how the wheel (bowl) complex extends into the cytoskeleton. Image adapted from[97]. B) Schematic highlighting the precise locations of all fifteen attachment organelle proteins. Image adapted from[132].

At first it was assumed that the *M. pneumoniae* cytoskeleton was composed of an actin-like protein[133], but further analysis failed to identify any actin analogs[33]. Treatment with Triton X-100 removes all the cellular components and assists in the solubilisation of hydrophilic proteins allowing the enrichment of cytoskeletal-associated proteins. Micrographs of Triton X-100 extracted cells revealed a network of Triton X-100 insoluble proteins associated with the cytoskeleton and attachment organelle (Figure 1.5B)[58].

There are fifteen proteins that constitute the attachment organelle (Figure 1.6B)[118,119,132,134]. One of these proteins, HMW1, has been shown to have both an insoluble phase and soluble phase for Triton X-100[135]. This indicates that the molecule has multiple forms. A Triton soluble hydrophilic phase allows the protein to partition on the extracellular side of the cell membrane[135].

The *M. pneumoniae* accessory proteins are essential for cytadherence. These proteins do not play a direct role in cytadherence, but their presence is required. These proteins stabilise the attachment organelle, in addition to localising and maintaining the major adhesins P1 and P30 to the tip of the organelle. P24, P40, P41, P65, P90, P200, Lon

protease, TopJ, MPN387, CpsG, and High Molecular Weight proteins HMW1, HMW2, and HMW3 are the key accessory proteins of *M. pneumoniae*[118,119,132,134]. *M. pneumoniae* mutants defective in either the HMW1, HMW2, HMW3, P40, P90, P24, P41, P65, P200, TopJ, and MPN387 proteins were found to be non-motile or non-adherent highlighting the significance of these accessory proteins for tip structure function[136-143]. The function(s) of the other attachment organelle proteins, Lon protease and CpsG proteins remain unknown.

1.6 Adhesins and accessory proteins of *Mycoplasma pneumoniae*

M. pneumoniae has an affinity for the carbohydrate structures consisting of α2-3-sialated poly (N-acetyllactosamine) on glycoproteins and glycolipids found on human erythrocytes and bronchial cilia[144,145]. Roberts et al. showed that sialylated glycoproteins are recognised by *M. pneumoniae* on human cilia[146]. These long chain poly-N-acetyllactosamine receptors are abundantly expressed and highly polarised at the tips of bronchial ciliated epithelial cells[147]. Sialylated groups carry a negative charge. Early work suggested that the interactions between *M. pneumoniae* and sialylated molecules may not be based on charge interactions[144]. Nonetheless, Mycoplasmas (and other organisms that have a low G+C content) tend to encode for a higher proportion of positively charged lysine, and hydrophobic amino acids[148]. Exposed lysine residues on surface localised proteins may target the negatively charged molecules found along human cilia such as glycosaminoglycans[149]. Negatively charged sulphated groups of glycosaminoglycans are targets for other respiratory pathogens such as *Haemophilus influenzae* and *Mycoplasma hyopneumoniae*[149,150].

Regardless of whether or not positively charged lysine residues are involved, *M. pneumoniae* surface proteins play an important role in adherence to receptors on epithelial cells. *M. pneumoniae* cells pre-treated with trypsin were inhibited in their ability to adhere to a bronchial epithelial layer (Figure 1.7B). Adherence was resolved with prolonged incubation in media indicating that *M. pneumoniae* cells were able to re-establish their surface protein topography sufficiently to regain the capacity to adhere (Figure 1.7D)[151].

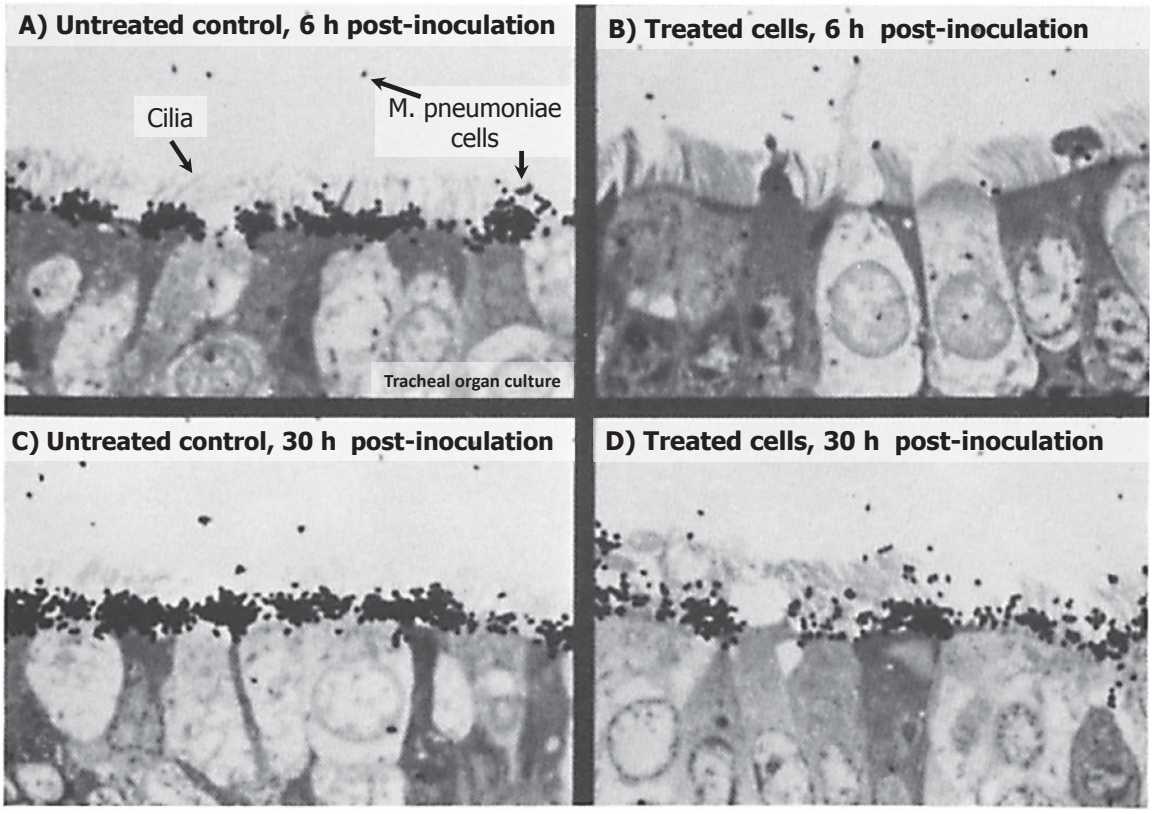


Figure 1.7: Autoradiographs depicting the necessity of surface proteins for the attachment and growth of *M. pneumoniae* cells. Cells were stained with silver grains. A) Tracheal rings were inoculated with *M. pneumoniae* cells and incubated for 6 hours (untreated control) and re-incubated in media for another 24 hours; seen in C). B) Tracheal rings were inoculated with pre-treated *M. pneumoniae* cells (incubation with trypsin for 15 minutes). D) The treated culture from B) was re-incubated in fresh media for 24 hours, the attachment indicates that cells had regenerated new surface proteins and were able to adhere[151]. Note the minimal number of attached cells to the tracheal ring in B), adapted from[151].

Further studies concluded that adhesin and accessory proteins found primarily on the surface of the attachment organelle were responsible for cytadherence of *M. pneumoniae* to the respiratory epithelium (Figure 1.8)[27,136,152]. The major adhesion proteins of *M. pneumoniae* are the P1, P30, and P116 adhesins[151,153-155].

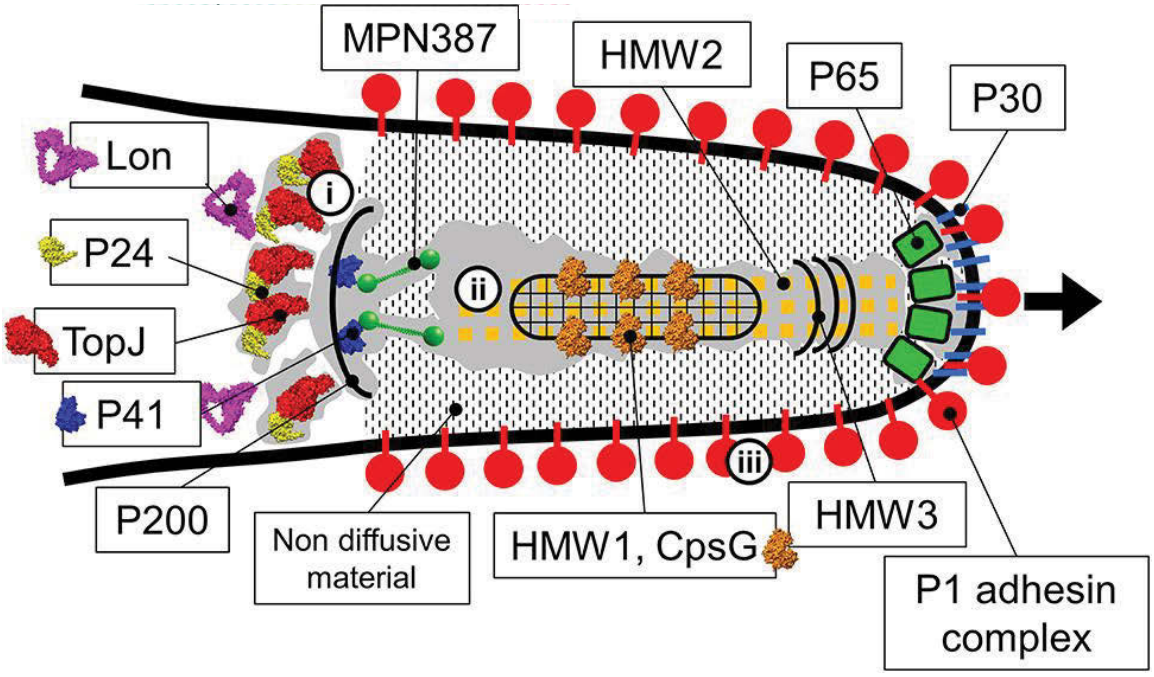


Figure 1.8: Schematic representation of the adhesive attachment organelle of *M. pneumoniae* including the precise locations of the adhesins and accessory proteins. Adapted from[121].

The P1 adhesin is considered to be the major adherence protein in *M. pneumoniae*. Cells that are unable to express P1 are unable to adhere to host cells[151,156-159]. Antibodies generated against P1 will block *M. pneumoniae* adherence to abiotic surfaces[160] and to epithelial cells[27,155,156,159,161,162]. *M. pneumoniae* adherence to epithelial cells was reduced by 75%[27] which indicates adhesins other than P1 play a role in adherence. The P1 adhesin is encoded by *mpn141*[163] which contains the RepMp elements responsible for driving antigenic variation in this protein[33,111]. P30 is another major adhesin in *M. pneumoniae*. *M. pneumoniae* cells that are unable to synthesise P30 are unable to adhere and do not have P1 localised to the attachment organelle (proper localization seen in Figure 1.8)[154,164]. Not much is known about the P116 adhesin other than antibodies raised against P116 are able to reduce *M. pneumoniae* cytadherence[155]. The correct positioning of the P1 and P30 adhesins is a complex process that requires insertion into the cell membrane and correct folding into an adhesive complex. This can only be achieved by interacting with partner accessory proteins[159,165].

The P40 and P90 accessory proteins are cleavage products of Mpn142 encoded by the *mpn142* gene[166-168]. Mutants defective in *mpn142* display a non-adherent phenotype[169]. Both P40 and P90 form a complex with P1 on the attachment organelle named the 'P1 adhesin complex'[170] (Figure 1.8). P65[139] and all three high molecular weight proteins HMW1[171], HMW2[27,152,172], and HMW3[173] are required for attachment organelle development and localisation of the adhesion complex. P24[137], P41[138], and TopJ[141,174] anchor and stabilise the attachment organelle.

Although *M. pneumoniae* has three major surface-exposed adhesins, it requires all 15 attachment organelle proteins for efficient cytadherence. The two adhesins and thirteen accessory proteins are required to develop a functional attachment organelle in order for *M. pneumoniae* to successfully colonise the human respiratory epithelium. Studies have shown that if the function of any of the accessory proteins are inhibited, *M. pneumoniae* cells adhere poorly[136-143]. The function of the Lon protease and CpsG protein in the attachment organelle remains unknown, and the precise cellular location(s) (intracellular or extracellular, or both) of several proteins in the complex are also unclear. It is important to define the surface exposed regions of proteins localising to the attachment organelle. Many of the adhesins have multiple domains and it is becoming clear that certain regions within these adhesins are more desirable as vaccine targets than other regions within the same molecule[93]. It will be important to define regions in adhesins that induce undesirable immune responses that exacerbate inflammatory responses. One of the aims of work presented in Chapter 2 is to identify regions in these and other surface accessible proteins that might serve as a target for vaccine development.

1.7 Moonlighting proteins

Prior to systems-wide genomic, transcriptomic and proteomic studies, there was a widely held assumption that one gene would express one protein with only one function[175]. However, studies characterizing the surface proteome of Gram-positive bacteria[176-180], Gram-negative bacteria[181-184], and Mycoplasmas[185-187] have identified a class of multifunctional proteins dubbed "Moonlighting" proteins[188]. Moonlighting proteins are a subset of multifunctional proteins encoded from a single ORF that perform more than one function[189]. These cannot be from products from gene fusion, alternative splicing, or those that display low secondary functions (promiscuous enzymes)[189,190]. These proteins

typically perform a canonical function in the cytosol. These molecules have no obvious signal peptides but proteome studies report them on the surface of a wide range of Gram-positive and Gram-negative pathogens. How they get to the cell surface has been a topic of controversy and has led to scepticism about the role these proteins might play on the cell surface. However, over the past 25 years there have been many studies attesting to the multifunctional characteristics of these proteins and descriptions of novel mechanisms that might explain how they arrive on the cell surface[191-193]. Many moonlighting proteins are able to be translocated to the bacterial surface to perform an alternate function; a majority of which act as secondary adhesins[188]. Moonlighting proteins are often: 1) enzymes of the glycolytic pathway, 2) enzymes of other metabolic pathways, and 3) molecular chaperones. A vast range of bacteria have been shown to express moonlighting proteins including: *Staphylococcus aureus*, *Bacillus anthracis*, *Escherichia coli*, *Chlamydia pneumoniae*, *H. influenzae*, *Helicobacter pylori*, *Mycobacterium tuberculosis*, *Mycoplasma genitalium*, *Mycoplasma pulmonis*, and *M. hyopneumoniae* just to name a few (Reviewed in[194] and[195]). A database for a majority of the reported moonlighting proteins from both eukaryotes and prokaryotes has been collated and can be found on <http://www.moonlightingproteins.org/>[196].

Several vaccine formulations comprise immunogenic moonlighting proteins including enolase[197-199], GAPDH (reviewed in[200]), 6-Phosphogluconate dehydrogenase[201,202], fructose-biphosphate aldolase[203], and elongation factor Ts[203]. Moonlighting proteins are important components in vaccines designed to elicit immune responses against bacteria and parasites[204].

In *M. pneumoniae*, several cytosolic proteins have been reported to moonlight on the cell surface. Elongation factor Tu (Ef-Tu) and pyruvate dehydrogenase E1 β subunit (Pdh-B) were described to have a cell surface location in 2002[205]. Although both proteins perform their canonical function in the cytosol, putative transmembrane domains have been identified in both molecules[205]. This suggests that both proteins are able to be secreted across the cell membrane to the cell surface via novel pathways. In the study by Dallo et al., it was demonstrated that Ef-Tu and Pdh-B are cell surface-accessible (Figure 1.9) and bind to fibronectin in a dose dependant and saturable manner. Recently, it was suggested that these "cytosolic" moonlighters are anchored to the *M. pulmonis* cell membrane

through a rhamnose moiety near the carboxyl terminus[206]. Only the membrane bound enolase was observed to be post-translationally modified with the addition of rhamnose, cytoplasmasmic enolase had no post-translaitonal modifications[206]. Rhamnose has been detected in a range of Mycoplasmas and might serve to anchor moonlighters in those species[207], however it is possible that other carbohydrates may act as protein anchors. Numerous moonlighting proteins have been identified recently that bind a wide variety of host antigens[208-210] (Table 1.1). Additionally, some of these moonlighting proteins are also able to bind plasminogen. Binding induces a small conformational adjustment in plasminogen that is sufficient to promote activation to the active protease, plasmin, by host plasminogen activators. Capturing plasmin onto the surface of bacterial pathogen represents a powerful strategy to degrade extracellular matrix, trigger a matrix metalloprotease cascade, and degrade a wide variety of essential host clotting and immune molecules (reviewed in[211] and[212]). Several moonlighting proteins in *M. pneumoniae* have been shown to bind plasminogen, and degrade extracellular matrix molecules in the presence of plasminogen activators[209,210,213] (Table 1.1).

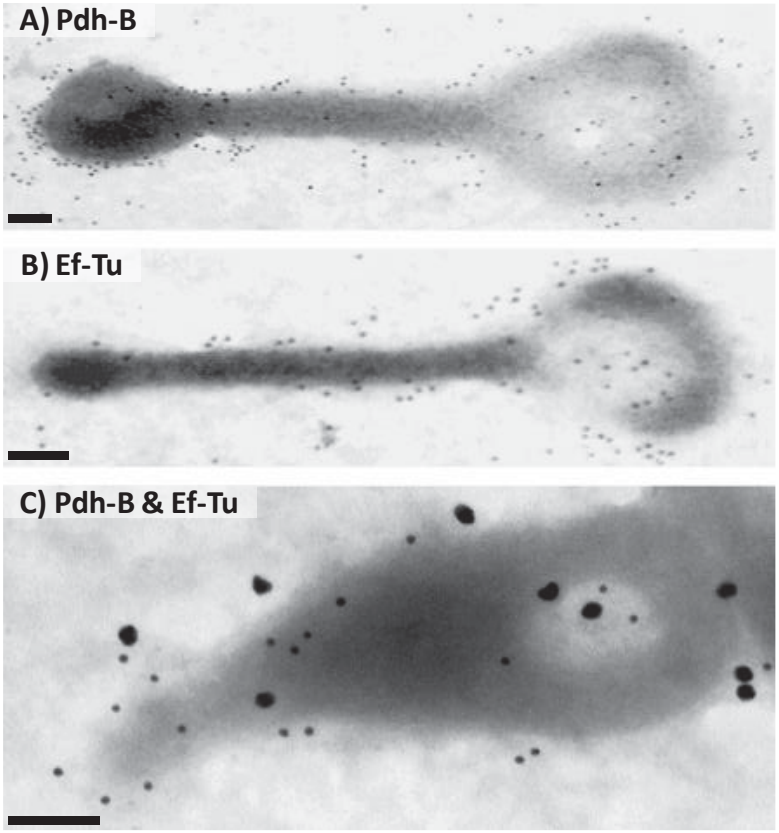


Figure 1.9: Immunogold electron microscopy of the moonlighting proteins on *M. pneumoniae*. A) Immunogold labelled antibodies against Pdh-B. B) Immunogold labelled antibodies against Ef-Tu. C) Dual gold labelling of Pdh-B and Ef-Tu antibodies. Scale bar = 0.1 µm, adapted from[205].

Table 1.1: Metabolic enzymes and chaperones from *M. pneumoniae* with moonlighting functions.

	Fibronectin[f]	A549[e]	Plg[e]	Lactoferrin[e]	Laminin[e]	Fibrinogen[e]	Vitronectin[f]	Plg activation
Ef-Tu	Y[a]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Pdh-B	Y[a]	Y	Y[c]	Y	Y	Y	Y	F&V[d & f]
GAPDH	Y	Y	Y	Y	Y	Y[b]	Y	Vitronectin[f]
Pdh-A	-	Y	Y[d]	Y	-	Y	Y	Vitronectin[f]
Pdh-C	Y	-	Y[d]	Y	Y	Y	Y	-
Lactate dehydrogenase	-	Y	Y	-	Y	Y	Y	Vitronectin[f]
Phosphoglycerate mutase	-	Y	Y	-	Y	Y	Y	F&V[e & f]
Pyruvate kinase	-	Y	Y	Y	Y	Y	Y	Vitronectin[f]
Transketolase	-	Y	Y	-	Y	Y	Y	Vitronectin[f]
GroEL[g]	Y	Y	Y	Y	Y	Y	Y	F&V[g]
DnaK[g]	Y	Y	Y	Y	Y	Y	Y	F&V[g]

Y indicates the protein binds to the specific host cell/protein. N/A indicates the particular interaction was not tested. Abbreviations for *M. pneumoniae* proteins are: elongation factor tu (Ef-Tu), pyruvate dehydrogenase β subunit (Pdh-B), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), pyruvate dehydrogenase α subunit (Pdh-A), and dihydrolipoyllysine acetyltransferase (Pdh-C). (A549) human lung epithelial cell line. (Plg) plasminogen. In the 'Plg activation' column, F&V indicates the protein is able to degrade both fibrinogen and vitronectin in the presence of plasminogen activators. [a]Data collected from Dallo et al., 2002[205]. [b]Data collected from Dumke et al., 2011[208]. [c]Data collected from Thomas et al., 2011[214]. [d]Data collected from Gründel et al., 2015[215]. [e]Data for the entire column was collected from Gründel et al., 2015[209] unless specified. [f]Data for the entire column was collected from Gründel et al., 2016[210] unless specified. [g]Data for the entire row was collected from Hagemann et al., 2017[213].

Post-translational processing may also contribute to the multifunctionality of moonlighting proteins. Processing of surface molecules has been reported in various Mycoplasma species[186,187,216-236] as well as in many Gram-positive[237,238], and Gram-negative bacteria[239-241]. Surface proteome studies have been important in identifying moonlighting proteins as moonlighting proteins cannot be detected during proteomic studies that employ Shotgun approaches to characterise whole cell lysates. In these studies, the lysates are digested with an enzyme before mass spectrometry analysis losing the ability to detect cleavage fragments. Cleavage events are likely to destroy canonical function but represent a mechanism to release peptides or subdomains within modular proteins. Lactate dehydrogenase in *M. hyopneumoniae* is found to be cleaved to create two fragments on the cell surface[227]. In *Campylobacter jejuni*, cleavage of the immunogenic Campylobacter adherence Factor (CadF) generates fragments that no longer generates an immune response but retains affinity for fibronectin[239]. Phage infected *E. coli* cells cleave the N-terminus of Ef-Tu to inhibit translation as a programmed cell death defence mechanism[242].

All the adhesins in the P97 and P102 adhesin families from *M. hyopneumoniae* undergo cleavage to form functional fragments that assist in adhesion to a range of host molecules. *M. hyopneumoniae* adhesin fragments have been shown to bind the glycosaminoglycan mimic heparin[186,187,217,218,221], fibronectin[219,220,226], and plasminogen[219-221]. However, cleavage is not restricted to adhesins in *M. hyopneumoniae* as a recent global proteome study observed that a range of lipoproteins, annotated cytosolic proteins, proteases and uncharacterised proteins are also cleaved[227]. Cleavage fragments can be further modified by the action of aminopeptidases that moonlight on the surface of *M. hyopneumoniae,* but the functional significance of these events remains to be determined[225,227,243,244].

To date, there have been a few examples of unusual cleavage events affecting protein function in *M. pneumoniae*. Perhaps the most well characterised is the major cleavage event that generates P40 and P90 from Mpn142[166-168]. The removal of the signal peptide in both P40[167,168] and the P1 adhesin[163,245,246] has also been observed. An inflammatory response during infection is induced by N-terminal peptides derived from two lipoproteins (Mpn611 and Mpn162)[228] and an ATPase (Mpn602)[229], which are released after proteolysis. These lipopeptides induce NF-κB activation, chemokines, inflammatory cytokines, tumor necrosis factor-alpha, and leukocyte infiltration (in mice models)[228,229].

Though the exact cleavage sites have not been mapped, cleavage fragments of several *M. pneumoniae* proteins have been reported. These include, a DnaK fragment and a C-terminal P1 fragment which forms a complex with full length P1 adhesin and other accessory proteins, observed in paraformaldehyde cross-linking studies[247]. Mass spectrometry analysis of gel spots from two-dimensional gel electrophoresis of *M. pneumoniae* cell lysates also identified fragments of several uncharacterised lipoproteins (Mpn052, Mpn284, Mpn288, Mpn376, Mpn400, Mpn408, Mpn444, Mpn456, Mpn474 and Mpn491)[248] suggesting that processing of lipoproteins is important in this species.

It was recently reported that the accessory proteins, P90 and P40, derived from a major processing event in Mpn142 are targets of a complex array of processing events on the surface of *M. pneumoniae*[230] (see Chapter 4). It is hypothesised that processing events may be more common in *M. pneumoniae* than is currently known. These cleavage fragments could act as adhesins as is observed in *M. hyopneumoniae* or they may perform other functions yet to be characterised. Two-dimensional electrophoresis coupled with mass spectrometry revealed there are *M. pneumoniae* lipoproteins that are cleaved; some of which are cleaved several times[248]. Paraformaldehyde cross-linking studies of *M. pneumoniae* observed a P1 fragment and DnaK fragment complexed to several attachment organelle proteins[247]. However, the full repertoire of cleavage products has not been characterised.

1.8 Thesis aims and Chapter synopses

Despite the application of a wide variety of systems biology approaches to study *M. pneumoniae* (such as genome[33], transcriptome[40,41], proteome[42-47], metabolome[48], protein interactome[49], and phosphoproteome[50-52] just to name a few) there are no efficacious vaccines for this pathogen. However, a significant omission is the lack of knowledge of a surface proteome for *M. pneumoniae*. From a historical perspective, this has probably occurred because several leading laboratories in the field have focussed their efforts in studying the attachment organelle, a remarkably interesting biological structure that is critical for adherence and cell motility. It is clear from extensive proteome studies of *M. hyopneumoniae*, that the surface proteome of Mollicutes are enriched in lipoproteins, dedicated cell adhesins, ABC-like transporters, and a diverse array of proteins that have

canonical functions in the bacterial cytosol. These proteins and regions within them will have essential roles in cytadherence, biofilm formation, cell-cell communication, antimicrobial activity, immunomodulation, and influencing attachment organelle development. Future vaccine and other therapeutic targets are most likely to reside in the surface proteome. As such, it is critically important to attempt to define the collection of proteins that are exposed on the extracellular surface of *M. pneumoniae*, and how they may be modified by post-translational processes. The overarching aim of this thesis was to define proteins and their modified products which comprise the surface proteome of *M. pneumoniae*. High throughput methods were also used to gain insight into the function of these molecules and how they interact with epithelial cells. This aim is further divided into sub-aims:

Aim I: To identify the global surface proteome of *M. pneumoniae* (Chapter 2).

Aim II: To examine the interaction of *M. pneumoniae* proteins and their native complexes with diverse host molecules that reside on host epithelial cell surfaces (Chapter 3).

Aim III: To characterise post-translational cleavage products derived from *M. pneumoniae* surface proteins (Chapter 3).

Aim IV: Provide examples of *M. pneumoniae* proteins that are targets of complex post-translational proteolytic processing events and determine putative functions for the cleavge fragments (Chapter 4, 5, and 6).

This PhD thesis involved the analysis of the complete surface proteome using surface biotinylation and trypsin shaving, two complementary methods that have been used previously in other bacteria (Aim I).

Affinity chromatography methods ('Bait and Prey' experiments) were developed in an endeavour to find *M. pneumoniae* proteins that interact with diverse host cell ligands including heparin, fibronectin, plasminogen, actin, fetuin, as well as protein complexes extracted from the surface of A549 epithelial cells. These protocols were applied in a systems-wide manner to recover proteins from whole cell lysates of *M. pneumoniae* (Aim

II). Proteins identified in 'Bait and Prey' experiments were compared to the list of proteins identified in surface proteome of *M. pneumoniae*.

The preliminary experiments presented here indicated that proteins that reside in the surface proteome are targets of endoproteolytic processing events. Aim III sought to investigate protein proteolysis in a systems wide manner by firstly investigating protein fragments from 'Bait and Prey' experiments and 2D-SDS PAGE spots of *M. pneumoniae* cell lysates. During the course of my PhD, an N-terminomics protocol (developed by Berry, I. J. and others) facilitated the recovery on neo-N-terminal peptides generated as a result of cleavage events. These N-terminomic datasets complimented the collection of protein cleavage fragments identified from the 'Bait and Prey' experiments and 2D-SDS PAGE of cell lysates. To describe how proteins are processed in *M. pneumoniae* (Aim IV), several representative molecules known to play an important role in attachment organelle function (P40 and P90 accessory proteins, Chapter 4), in adherence (the P1 adhesin, Chapter 6), as well as several novel adhesins (Ef-Tu, Chapter 5) were studied in detail. Processing events identified in Ef-Tu from *M. pneumoniae* were compared to processing events in the same protein from *M. hyopneumoniae* and a non-Mollicute species *S. aureus*. (Chapter 5).

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Chapter 2.

The surface proteome of *Mycoplasma pneumoniae*

i3 ithreeinstitute

2.1 Preface

Due to the absence of a cell wall, proteins on the surface of *M. pneumoniae* are predicted to be directly exposed to the extracellular milieu. A comprehensive surface proteome will provide a list of all the proteins that may interact with the respiratory epithelium (Aim I). Two independent methods were employed to explore the surface proteome. The completed analysis of the cell surface revealed essential proteins responsible for a wide variety of functions such as metabolism, protein synthesis, adherence, and motility; and represent future therapeutic targets.

2.2 Introduction

Bacteria interface with the host and environment through interactions involving proteins on their cell surface and those that are secreted into the extracellular milieu. As Mycoplasmas have no cell wall, these surface proteins are crucial for cytadherence, cell invasion, nutrient acquisition and transport, defence, biofilm formation, and growth. For *M. pneumoniae* to survive, it must first adhere to respiratory epithelium. For this, it possesses several dedicated cytadherence proteins including P1[152], P30[249], and P116[155].

Ef-Tu, Pdh-B, GAPDH, Pdh-A, Pdh-C, lactate dehydrogenase, phosphoglycerate mutase, pyruvate kinase, transketolase, GroEL, and DnaK[205,208-210,213-215] have all been shown to reside on the surface of *M. pneumoniae*. Several play a key role in recruiting plasminogen onto the surface of *M. pneumoniae* but many also bind other host proteins with roles in immune function, clot formation, cell structure, and signalling and are likely to be important in cytadherence. Once plasminogen is bound to the cell surface, it is readily activated by plasminogen activators to plasmin, an active serine protease. The capacity to recruit plasmin on the cell surface arms bacteria with the ability to invade cells and disseminate from the original site of colonisation to infect distal tissue sites (reviewed in[211] and[212]). Additionally, this activation of plasminogen and degradation of host molecules produces a cascade affect where matrix metalloproteases continue to degrade the host extracellular matrix[211,212]. Moonlighting proteins play a significant role in this regard, but it is difficult to rank their importance because mutant studies cannot be conducted due to the essential functions performed by many of these proteins.

Over the past 40 years, there has been a sustained effort to characterise the function of proteins associated with the *M. pneumoniae* attachment organelle. Functions of many

proteins in *M. pneumoniae* remain uncharacterised and *M. pneumoniae* may possess other proteins that are required for cytadherence. Despite *M. pneumoniae* being a model organism for systems biology, there have been no studies describing the surface proteome of this species (see Chapter 1). A comprehensive analysis of the surface proteome is required to identify proteins that may become components of future vaccine formulations and identify new targets for therapeutic intervention. Two independent methods, enzymatic cell shaving and surface biotinylation, were used to define the global surface proteome of *M. pneumoniae*.

2.3 Methodology

2.3.1 Strains

M. pneumoniae (M129 strain, ATCC 29342) cells were cultured as described previously[250]. Cells were grown in modified Hayflick's medium in tissue culture flasks (BD Falcon™) at 37°C.

2.3.2 *M. pneumoniae* surface labelling with NHS-biotin and avidin affinity purification

M. pneumoniae cells were grown in a T-175 culture flask as described above until confluent. Cells were washed with three times with PBS (Phosphate Buffered Saline) before being incubated with 10 mM EZ-link sulfo-NHS-biotin (Thermo Fisher Scientific) in PBS on ice for 30 s. The reaction was quenched with a final concentration 50 mM Tris-HCl in PBS for 10 min at 25°C. Labelled cells were washed three times and lysed with 7 M urea, 2 M thiourea, 40 mM Tris-HCl (pH 8.8), 1% (w/v) C7bZ0 (Sigma). Control samples were treated exactly the same as test samples except the cells were not exposed to sulfo-NHS-biotin. This was deemed necessary to determine if *M. pneumoniae* contains naturally biotinylated proteins and acted as a negative control. A positive control comprised cell lysate of *M. pneumoniae* that was labelled with sulfo-NHS-biotin using the same labelling conditions.

Avidin affinity chromatography purification of biotinylated surface proteins was conducted using Avidin Agarose Columns (Thermo Fisher Scientific). Cell lysates of biotinylated *M. pneumoniae* was loaded onto the column and incubated for 60 min at 25°C to allow binding. The eluent was collected and the column was washed with one column volume each of: Wash 1 (2 M urea in 100 mM ammonium bicarbonate); Wash 2 (2 M Urea, 0.5 M

sodium chloride 100 mM ammonium bicarbonate); Wash 3 (2 M urea, 30% acetonitrile in 100 mM ammonium bicarbonate); and Wash 4 (100 mM ammonium bicarbonate). Finally, five column volumes of 30% acetonitrile with 0.2% trifluoroacetic acid was used to ensure all proteins were removed from the column. All column flow throughs were analysed by 1D-SDS PAGE and transferred onto PVDF (polyvinylidene fluoride) membrane. The membrane was probed with a 1:5000 dilution of Extravidin-HRP (Sigma) for 60 min and developed in DAB (3,3'-Diaminobenzidine, Sigma). Duplicate samples were separated by SDS-PAGE and stained with Flamingo™ (Bio-Rad).

Eluents were first pooled and concentrated using a 3K MWCO filter (Pall) after the acetonitrile was evaporated. The sample of pooled eluents was separated by SDS-PAGE and divided into sections listed in Appendix 4: Table 1. The gel sections were dehydrated and incubated with trypsin for 16 h at 37° prior to analysis by LC-MS/MS (liquid chromatography tandem mass spectrometry).

2.3.3 *M. pneumoniae* enzymatic surface shaving

M. pneumoniae cells were grown in a T-175 culture flask (BD Falcon™) as described above until confluent. Cells were washed three times with PBS (pre-warmed at 37°C) before being incubated for 5 minutes with PBS (pre-warmed at 37°C). Cells were exposed to trypsin from bovine pancreas at a final concentration of 50 $\mu g.ml^{-1}$ (Sigma) for 5 min. The supernatant (containing released tryptic peptides and detached cells) were centrifuged for 10 min at 4000 RCF. This peptide supernatant was concentrated using a 3K MWCO filter (Pall).

To decrease sample complexity, both the top sample (> 3 kDa) and bottom flow through (< 3 kDa) of the 3K filter was digested a second time with Trypsin Gold (Promega Corporations) for 16 h at 37°C. The peptide sample above the cut off filter (> 3 kDa) was separated using a 11cm ReadyStrip™ IPG Strip (Bio-Rad) in a PROTEAN IEF Cell Apparatus (Bio-Rad) to further decrease sample complexity. The IPG gel strip was divided into sections and peptides were extracted in 50% acetonitrile, 0.2% trifluoroacetic acid before analysed by LC-MS/MS. Peptides were extracted from the bottom flow through (< 3 kDa) of the 3K filter with an OASIS HLB Extraction Cartridge (Waters Corporation®) and analysed by LC-MS/MS.

2.3.4 Liquid chromatography tandem mass spectrometry (LC-MS/MS)

5 µl of sample containing up to 5 µg of peptide was loaded into an autosampler vial in an Eksigent AS-1 autosampler connected to a Tempo nanoLC system (Eksigent) with a C8 Cap Trap column (Michrom Biosciences). The peptides were washed onto a PicoFrit column (75 µm × 150 mm) packed with Magic C18AQ resin (Michrom Biosciences). Peptides were eluted from the column into the source of a QSTAR Elite hybrid quadrupole-time-of-flight mass spectrometer (Sciex). Eluted peptides were ionized from the PicoFrit at 2300 V. An Intelligent Data Acquisition experiment was performed, with a mass range of 350–1500 Da continuously 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.

Files generated from LC-MS/MS were searched against the MSPnr100 database[251] with the following parameters: Fixed modifications: none; Variable modifications: propionamide, oxidized methionine, deamidation; Enzyme: semi-trypsin; Number of allowed missed cleavages: 3; Peptide mass tolerance: 100 ppm; MS/MS mass tolerance: 0.2 Da; Charge state: 2+, 3+ and 4+. For samples collected from the '*M. pneumoniae* surface labelling with NHS-biotin' methodology listed above, variable modifications also included NHS-LC-Biotin (K) and NHS-LC-Biotin (N-term).

2.4 Results and Discussion

2.4.1 Biotinylation and purification of the *M. pneumoniae* surface proteins

The *M. pneumoniae* surface was biotinylated using sulfo-NHS-biotin. This biotin tag readily labels accessible amines on lysine residues but is hydrophilic and does not permeate the cell membrane. By tagging surface proteins and enriching biotinylated proteins with avidin chromatography, the protein's molecular mass is retained. This is an important consideration because it allows cleavage fragments to be identified on the cell surface. This is in contrast to enzymatic surface shaving protocol which generates tryptic peptides from surface accessible proteins. Isolating intact surface proteins is useful for determining if surface accessible proteins are presented as multiple proteoforms derived by endoproteolytic processing (discussed in Chapter 3). However, the combination is powerful because shaving can identify surface exposed regions within membrane proteins, an important consideration for vaccine design.

To determine if *M. pneumoniae* contains endogenously expressed biotinylated proteins, a blot containing biotinylated and non-biotinylated *M. pneumoniae* cell lysates were reacted to Extravidin-HRP. The blot shown in Figure 2.1B shows that no naturally occurring biotionylated proteins were detected in *M. pneumoniae* (lane 2, 'WCL –') whereas many proteins were labelled with biotin in the test reaction (lane 3, 'WCL +'). The flow through contained non-biotinylated proteins and residual biotinylated proteins (Lane 4, 'Flow through') and washing successfully removed non-biotinylated proteins (Figure 2A). Eluents contained biotinylated proteins that spanned masses from 250 kDa to less than 10 kDa (Eluents in Figure 2.1B). From the 'Flow through' (lane 4) in Figure 2.1B, the presence of biotinylated proteins indicated that the Avidin Agarose Column was overloaded and not all biotinylated surface proteins were retained in the column. To extract more biotinylated proteins, the 'Flow through' sample was passed through avidin chromatography until all biotinylated proteins were collected.

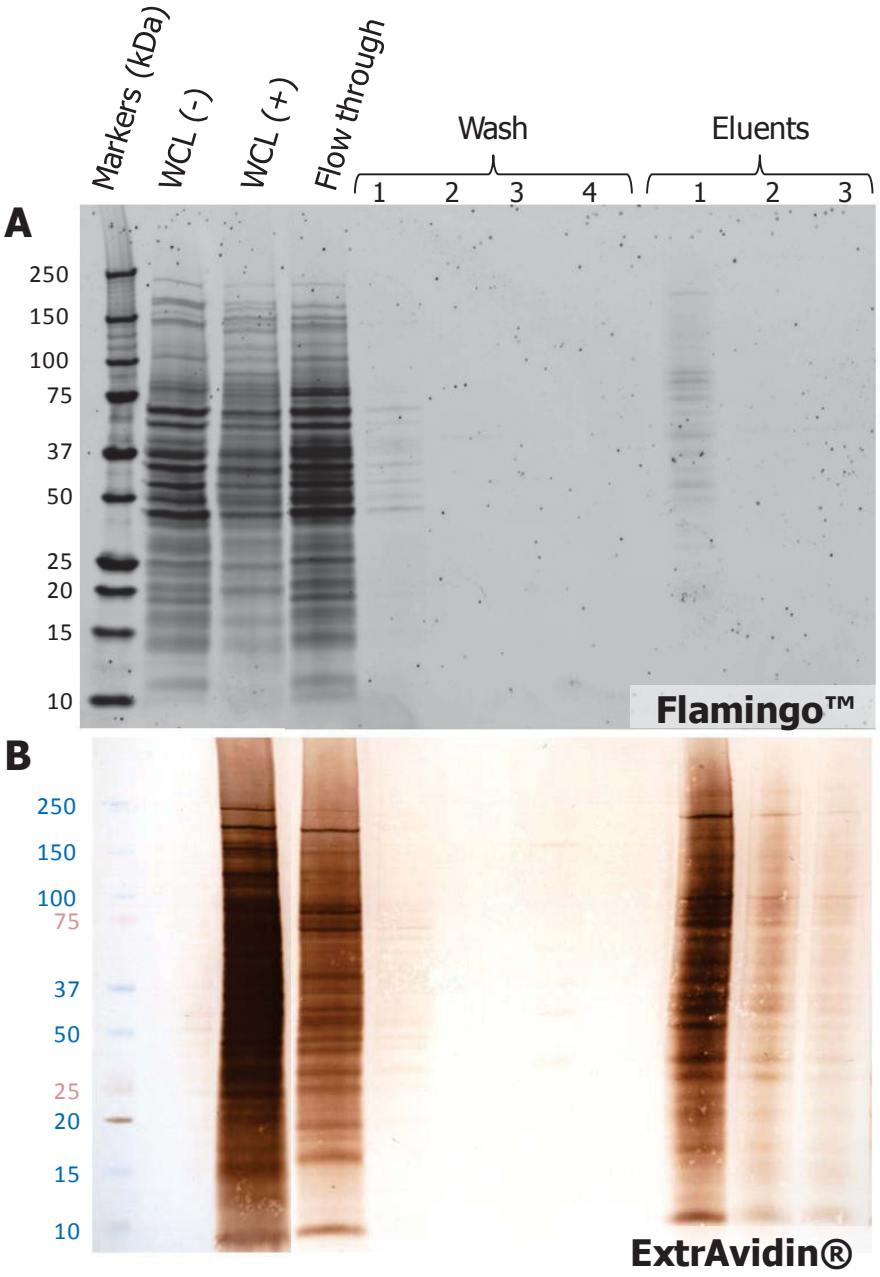


Figure 2.1: Western blot *M. pneumoniae* biotinylated surface proteins. A) Gel stained with Flamingo™. B) ExtrAvidin® HRP western blot of a duplicate gel. An *M. pneumoniae* cell lysate was biotinylated as the positive control (WCL +) and a non-biotinylated negative control was included in both gels. 'Flow through', 'Wash', and 'Eluents' were obtained from avidin affinity chromatography.

2.4.2 The surface proteome of *M. pneumoniae*

By combining both methodologies, LC-MS/MS identified 160 proteins on the *M. pneumoniae* cell surface of which 117 (73%) were identified by both approaches (Figure 2.2). 147 proteins were identified from avidin enrichment of biotinylated proteins, and 130 proteins contained trypsin accessible surface exposed regions that were identified by surface shaving (Figure 2.2). The complete list of proteins identified can be found in Appendix 1. *M. pneumoniae* cells were biotinylated for only 30 seconds on ice. This approach ensured that genuine surface proteins were labelled with biotin. Biotinylation has been conducted on other Mycoplasmas and times vary from 30 seconds for *M. hyopneumoniae*[186] to 30 minutes for *Mycoplasma mobile*[185]; both studies did not report cell lysis. Up to 2 hours has been used to biotinylate Gram-positive bacteria due to the thick cell wall[176,177]. Enzymatic shaving of *M. pneumoniae* was limited to 5 minutes; the same time used for *M. hyopneumoniae*[187]. *M. pneumoniae* has been incubated with trypsin for 30 minutes with no evidence of cell lysis[214]. In earlier studies, cell integrity was monitored using immunoblots with antibodies raised against a cytosolic *M. pneumoniae* protein[214]. Olaya-Abril et al. performed a time course of trypsin shaving on the Gram-postive bacteria, *S. pneumoniae,* ranging between 5 – 60 minutes and found cell lysis occurred after 30 minutes[178].

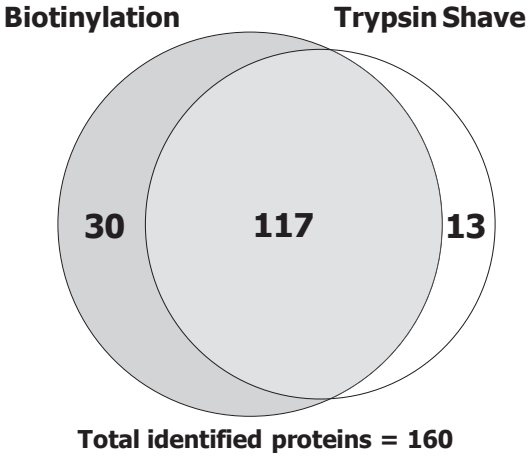


Figure 2.2: Venn diagram of *M. pneumoniae* surface proteins. Diagram was generated using Venny 2.1.0[252] <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.

The labelling time of 30 seconds for *M. pneumoniae*, while limiting cell lysis, may not have allowed sufficient labelling of all proteins; leading to the loss of low abundance proteins. As biotinylation containes a strong enrichment step, samples tend to be much more complex than samples from shaving experiments. In addition, LC-MS/MS analysis involves peptides derived from intact proteins (in biotinylation protocols) as opposed to peptides from only surface exposed regions (shaving protocols). Therefore, low abundance proteins are more likely to be identified using biotionylation protocols. Biotin is a small molecule (0.4 kDa[253] compared to trypsin (24 kDa[254]. It is conceivable that biotin may be able to access lysine residues that would not be accessible to the larger trypsin molecule due to steric hindrance[255]. Alternatively, the structure of the biotin tag used in this study may not always be accessible to the Avidin Agarose. The spacer arm of the sulfo-NHS-biotin may be too short and be blocked from attaching to the avidin by steric hindrance of the biotinylated protein[253].

The release of tryptic peptides provides important insight into protein topography. Trypsin needs to cleave surface exposed lysine or arginine residues to successfully release the peptide from the cell membrane. Depending on the number of transmembrane domains, proteins exposed on the surface of *M. pneumoniae* will either be released into the extracellular milieu or remain tethered to the cell membrane (Figure 2.3). If a protein is only anchored to the cell membrane at the N- or C-terminus, shaving will release the protein from the cell membrane (Figure 2.3A). Proteins with two or more transmembrane domains must have two accessible trypsin cleavage sites in each surface loop in order to be detected (Figure 2.3B). If there is only one accessible site, the protein will remain tethered to the membrane (Figure 2.3C). Lastly, proteins that are fully exposed on the surface should be detected (Figure 2.3D). Moreover, trypsin can target two amino acids (lysine and arginine) as opposed to biotin which can only target lysine[181]. It is possible that the proteins only identified by surface shaving only have surface exposed arginine residues, if so they will not be labelled by biotin. All these reasons could account for the identifications by only one of two methodologies.

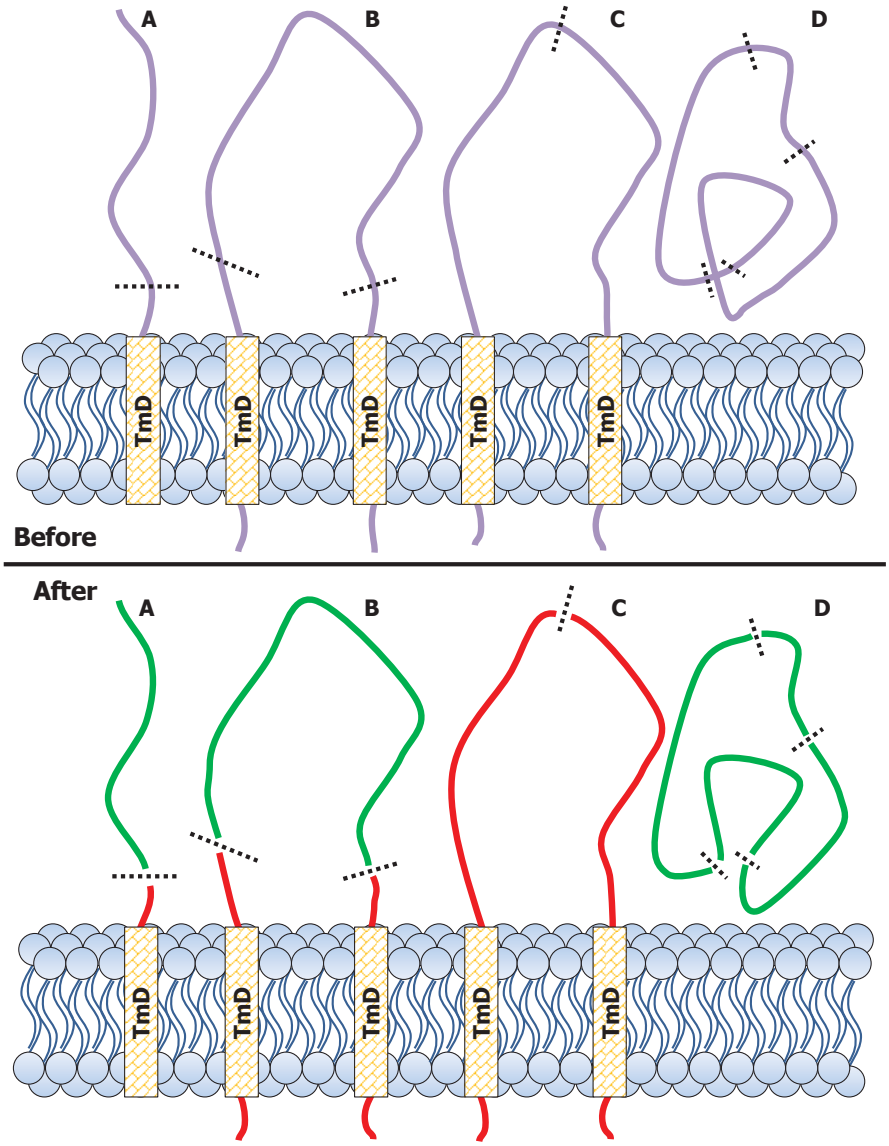


Figure 2.3: Identification of surface exposed proteins from surface shaving. Proteins before (top panel) and after (bottom panel) surface shaving. Trypsin accessible sites are indicated by the dotted lines. Yellow bars labelled 'TmD' represent transmembrane domains within these surface proteins. Green indicates regions that are released into the extracellular milieu to be detected by LC-MS/S whereas red indicates regions that remain tethered to the cell membrane. A) Protein anchored either by the N- or C-terminus, B) Protein with two transmembrane domains that contain two trypsin cleavage sites, C) B) Protein with two transmembrane domains with only trypsin cleavage site, and D) Protein that is fully exposed on the cell surface.

Trypsin releases tryptic peptides of proteins that are accessible from the cell surface. The ATP-dependent zinc metalloprotease was identified by both surface labelling and enzymatic shaving (Appendix 1). It is predicted to contain two transmembrane domains (TMpred predictions) which are located in the N-terminus (Figure 2.4, amino acids 26 – 45 and 175 – 192). Five peptides were identified by LC-MS/MS from shaving experiments which span the C-terminal half of the protein (Figure 2.4). These results suggest that the N-terminus (amino acids 46 – 174) is intracellular or otherwise inaccessible to trypsin and that the region from amino acid 193 onwards is surface exposed.6

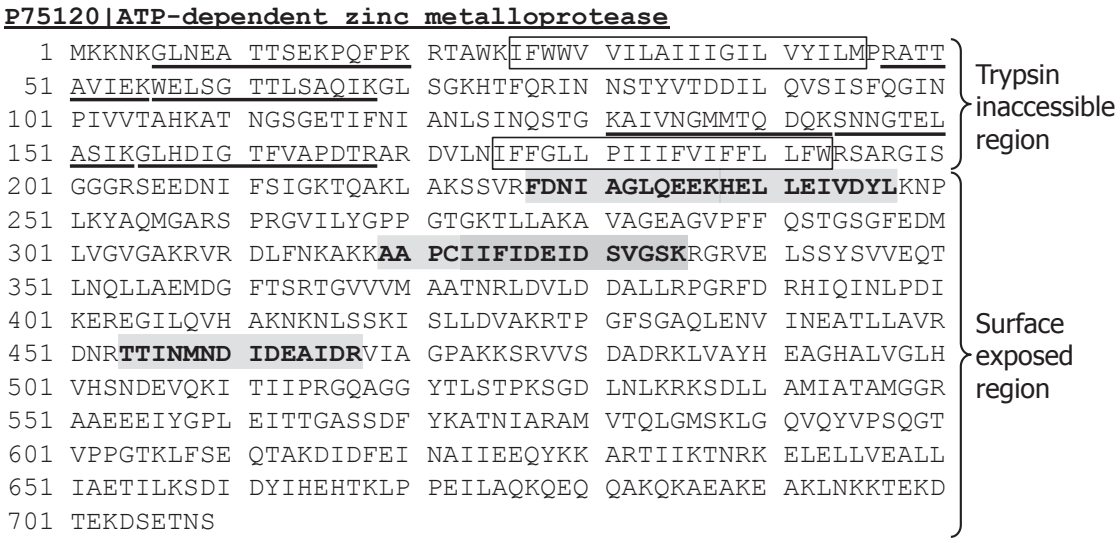


Figure 2.4: Shaved peptides of zinc metalloprotease. The amino acid sequence of zinc metalloprotease has been shown with shaved peptides identified by LC-MS/MS shown in bold text and shaded. Transmembrane domains predicted by TMpred are boxed. Peptides in the N-terminus that are detectable by LC-MS/MS are underlined.

2.4.3 Bioinformatic analyses on *M. pneumoniae* surface proteins

The bioinformatic tools TMpred, SignalP 4.1, and SecretomeP 2.0 were applied to the list of 160 proteins determined experimentally to reside on the cell surface. SignalP 4.1 searches for signal peptides that direct the transport and secretion of the protein from a cell[256]. SecretomeP 2.0 predicts whether or not a protein is secreted via non-classical secretory pathways, proteins identified include growth factors or extracellular matrix binding proteins[257]. Both these tools identified 21 proteins that are classically secreted to the *M. pneumoniae* surface, 37 that are non-classically secreted, and 102 proteins that

lack any signal or secretion sequences (Figure 2.5). TMpred is an algorithm that predicts the region of a protein that passes the cell membrane, and which side of the membrane[258]. The prediction is based on a database of experimentally determined transmembrane proteins and the helical membrane domains within these proteins[258]. TMpred predicts that 74 surface proteins contain at least one transmembrane domain (TMpred score greater than 500). Notably, no statistically significant transmembrane domains were detected for 86 proteins experimentally determined to reside on the cell surface of *M. pneumoniae*. Figure 2.5 depicts a summary of these analyses where the number of transmembrane domains (none, one, two, or more than two) of each protein is shown for each secretion pathway (not secreted, non-classically secreted and classically secreted).

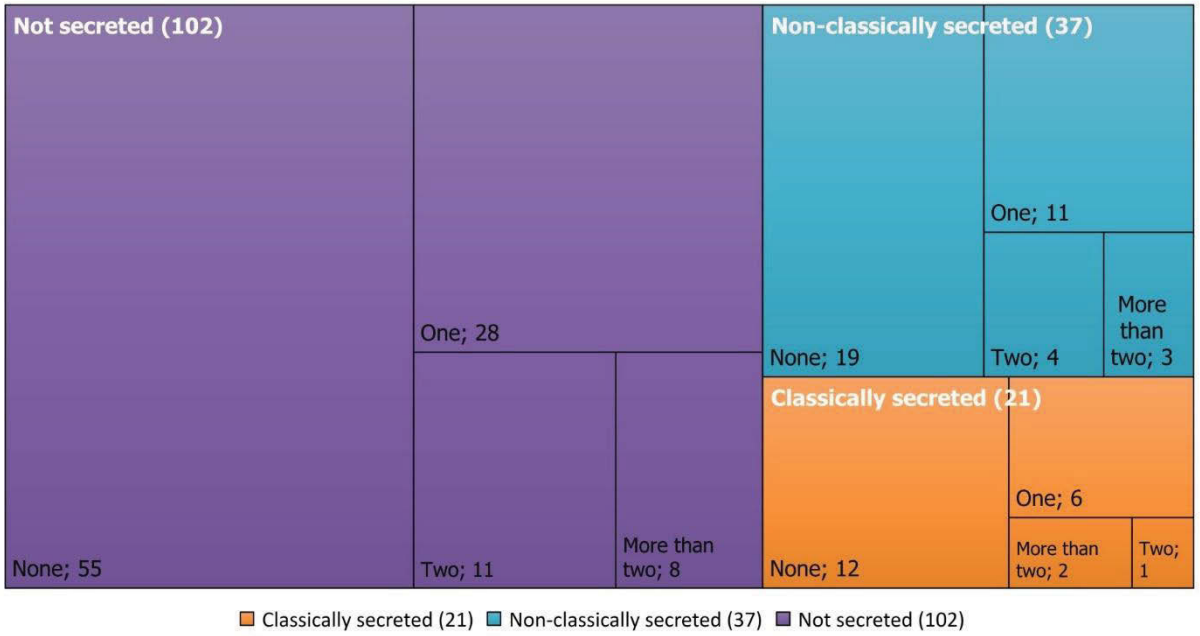


Figure 2.5: Treemap chart of secretion pathway and transmembrane domain predictions of *M. pneumoniae* surface proteins. Based on SignalP 4.1 and SecretomeP 2.0 predictions, surface proteins are divided into classically secreted, non-classically secreted, and not secreted proteins. These are divided further into the number of transmembrane domains (None, One, Two, or More than two) from TMpred predictions for each protein. The exact number of proteins in each box is also given. Predicted classically secreted proteins will have a positive SignalP 4.1 result; Predicted non-classically secreted will have a negative SignalP 4.1 result but a positive SecretomeP 2.0 result; and a protein predicted to not be secreted will have a negative result for both tools.

UniProt and PSORTb 3.0.2 were also utilised to provide insight into what is known about the proteins in the surface proteome of *M. pneumoniae*. UniProt is a universal protein database containing functional information of the proteome of thousands of organisms. UniProt annotates proteins based on published articles, a variety of algorithms and similarities to known proteins[259]. According to UniProt: 37 of the identified surface proteins are localised within the cell cytoplasm, 84 have no known cellular location, and 39 proteins are localised to the surface (Figure 2.6). PSORTb 3.0.2 is an algorithm that predicts the subcellular localisation of proteins by analysing the amino acid sequence. This sequence is applied to multiple algorithms to provide a probability of where the protein might be localised[260]. PSORTb 3.0.2 predicts that 99 proteins are cytosolic, 13 bound to the cellular membrane, and 2 to be on the cellular surface. PSORTb 3.0.2 is unable to predict the location of 46 *M. pneumoniae* surface proteins (Figure 2.6).

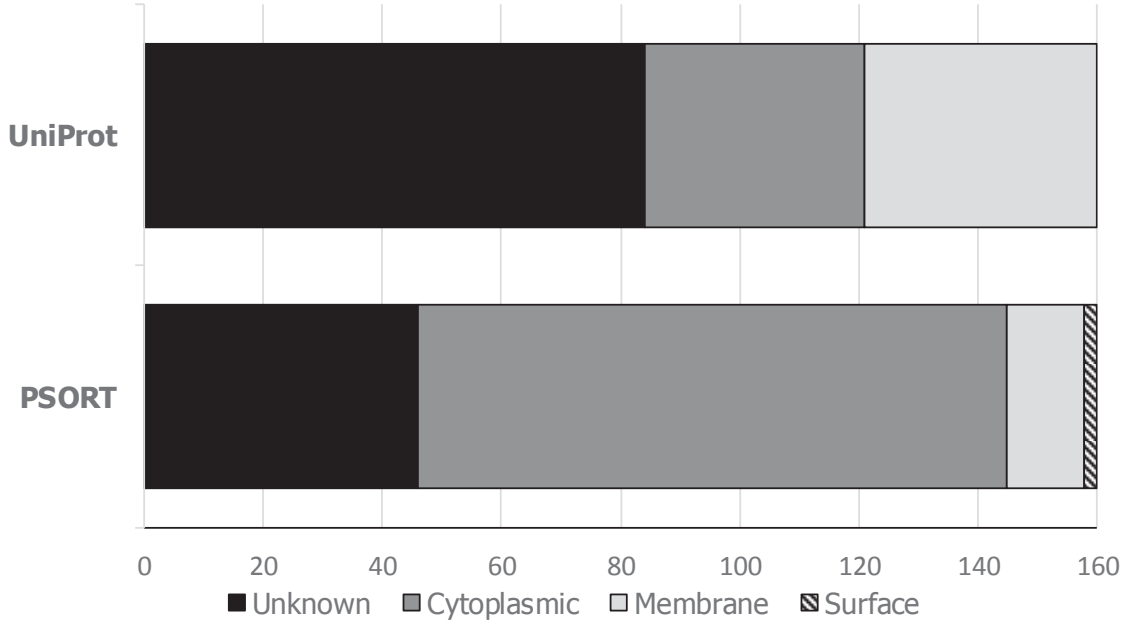


Figure 2.6: Cellular location predictions of *M. pneumoniae* surface proteins. Cellular locations are divided into three and four locations for UniProt and PSORTb 3.0.2, respectively.

Membrane vesicles are known to package cellular proteins for export to the extracellular side of the membrane[191,192]. Cell lysis is an essential part of biofilm formation and this may be an important mechanism for releasing "public goods"[193]. Extracellular DNA and

moonlighting proteins released by lysis are known to play an important role in biofilm formation[261-264]. A review consisting of 22 surface proteomes studies revealed that more than one-third of the proteins identified to be on surface lacked any predicted transmembrane domains or secretion signals[195].

2.4.4 Attachment organelle proteins of *M. pneumoniae*

This study employed complementary approaches to identify the cellular location of *M. pneumoniae* proteins. Of the fifteen attachment organelle proteins, only the P30 adhesin, P200, P24, and the Cpsg protein were not identified by either method to be on the *M. pneumoniae* surface. The P30 adhesin is predicted to be classically secreted, to contain transmembrane domains, and the carboxy-terminal region has been shown to be surface exposed[265,266]. Proteomic data derived from an analysis of *M. pneumoniae* whole cell lysates proteomic data identified only one peptide within P30 (data not shown). Closer inspection of the P30 amino acid sequence showed the presence of a proline residue preceding either a lysine or arginine residue (Figure 2.7). The presence of proline next to lysine or arginine is known to inhibit the ability of trypsin to cleave[267], rendering the C-terminal half of P30 to be trypsin insensitive. Using PeptideCutter[268], which predicts the peptides obtained from digests from different enzymes, P30 has only one tryptic peptide of the suitable length that is identifiable by the QSTAR Elite mass spectrometer used in this study. This tryptic peptide resides in the amino half of P30 which has been shown to be intracellular[266]. Some proteins may be exported but not retained on the cell surface in sufficient quantity to be detected using the methods in this study. Additionally, proteins exposed on the surface may contain large trypsin insensitive regions that are absent of any lysine or arginine residues and therefore would not be identified by the methods used in this study. An example of this is seen in HMW1 which has been shown to be a peripheral membrane protein[135]. HMW1 was identified be a surface exposed protein by both surface biotinylation and trypsin shaving methodologies (Figure 2.8). Upon closer inspection, HMW1 contains a very large trypsin insensitive region (498 amino acids, 49% of the entire protein) that would not be detected by mass spectrometry (Figure 2.8).

```
P75330|P30 Adhesin
  1 MKLPPRRKLK LFLLAWMLVL FSALIVLATL ILVQHNNTEL TEVKSELSPL
 51 NVVLHAEEDT VQIQGKPITE QAWFIPTVAG CFGFSALAII LGLAIGLPIV
101 KRKEKRLLEE KERQEQLAEQ LQRISAQQEE QQALEQQAAA EAHAEAEVEP
151 APQPVPVPPQ PQVQINFGPR TGFPPQPGMA PRPGMPPHPG MAPRPGFPPQ
201 PGMAPRPGMP PHPGMAPRPG FPPQPGMAPR PGMPPHPGMA PRPGFPPQPG
251 MAPRPGMQPP RPGMPPQPGF PPKR
```

Trypsin insensitive region

Figure 2.7: Tryptic sites of the P30 adhesin. The amino acid sequence of P30 has been shown with tryptic cleavage sites (lysine and arginine residues) in red text and underlined. The only peptide that can be identified by LC-MS/MS is in blue bold text. Transmembrane domains predicted by TMpred are boxed. Regions that are trypsin insensitive due to adjacent proline residues are shown in grey boxes.

```
Q50365|Cytadherence high molecular weight protein 1
  1 MKKSKEAVFE DKDYTEENPE QIFGNLYDGK LTVQDGKVKI AYDGDGNGYY
 51 IAFNSETGVY YDPYGDTEYD ISVLFDANGN SFVFADAPTV EVLAGEQEQT
101 EAEPDYLQYV GNEAYGYYDE AGEWVWSGYF EGDQWISTLP QTEAEEKQFG
151 FEDNIETTPT ASEDFGLEAD VPAPEVAEPS YEVQPEVAAE PVYDVQPEVA
201 VEPVGETTAT VEPQAVEIQP EVVVEPIVES QLEQPVEVQA EMVQPEVAVE
251 PQLEVSLDPI GETAPILEQV EPQAVQTQPE IPAEQSAVEL QPEPVAEVQS
301 EMVQPEAAAE PVTEAQQTEP TPVVETIAEI TPQVVTEPVV AVVEHQPEAV
351 AEPLPVEPAV AGVSELIPTE QVQPEVVVES TPVAEVQSEM VQPEVAVEPI
401 VEPQPEQPVE VQPEVITTPE VASVLEVQPE NPVVEVEQVV EPQPETPVEV
451 QPEPVVETVQ EAVAEPTQVV EPQPQAAPQP AVYEWNLTPE AAPVEQPEVI
501 PVTVVESQAT ATAEPQPAVA PVADMDYVLH LTDTVKNQPQ TAPVQPTTPI
551 KIEVAESTPT VTTSPVEPTI APPLFEIELN NTTSSDLPLV EVVDFKHNQH
601 GAVGTHSFDD FTPPEVGMES KTHCHSNSEV VWRVSEPKTV PVPPAVSSIN
651 IQTVNRVVEP TISTPTTPVV ESAPAIEIFV DTPPVETKEA SSNVDVVQQP
701 VKPLMPVMVE QLRTTELQPT TEINLFANSD INSIIAELKQ GRSNPAINFD
751 DIFKMSSYQM VVKKSFVQIS DFITNSKTDI TNRFLLIKKE LQAELTRLIE
801 ENEQLKAEFL NAKDLSVYQK DELLRSLSND FTIAHRPSDS YEQLQKSGEL
851 VRNIQKAILE NESKIKNIQI TLKELKAVYK LCSDTVLNGM AKLDSVLRFN
901 KKEKDPLLLN SMETLSSFET EPQAIIEDLL DFSSSFDKMS NEQLDEFVYQ
951 NLDSGLNLDL DGFDHQLSSM NIHGLEPLDP MKLDDFDFET LTPDKTSNLS
1001 SILDDELMEN GGDFNLDY
```

Trypsin insensitive region

Figure 2.8: Tryptic sites of the HMW1 accessory protein. The amino acid sequence of HMW1 has been shown with tryptic cleavage sites (lysine and arginine residues) in red text and underlined. Peptides identified from trypsin shaving experiments are in blue bold text. The large trypsin insensitive region (no lysine or arginine amino acids) spanning 498 amino acids is shown in grey.

2.4.5 Putative surface proteins of *M. pneumoniae*

Aside from proteins of the attachment organelle, several other proteins have been reported to be localised on the surface of *M. pneumoniae* such as the moonlighting proteins Ef-Tu, Pdh-B, GAPDH, Pdh-A, Pdh-C, lactate dehydrogenase, phosphoglycerate mutase, pyruvate kinase, transketolase, GroEL, and DnaK[205,208-210,213-215]; the uncharacterised Mpn474 protein[269]; and the membrane bound proteins zinc metalloprotease and ATP synthase alpha subunit[270]. The surface proteome data generated here identified all these proteins except for transketolase (Appendix 1).

On the other hand, various *M. pneumoniae* proteins observed to be not being surface exposed in the literature were identified by the surface analysis in this chapter. These proteins include elongation factor G[270], enolase[214], dihydrolipoyl dehydrogenase (Pdh-D)[215], fructose-bisphosphate aldolase, triose-phosphate isomerase, phosphoglycerate kinase, acetate kinase, and phosphate acetyltransferase[209].

The difference in observed cellular locations could arise from the different approaches between previous studies and the methodologies presented here. In the study by Thomas et al., a combination of membrane fractionation, Triton X-100 and Triton X-114 fractionation of *M. pneumoniae* cells, colony blotting, cell shaving, and immunofluorescence was used to determine the location of enolase and Pdh-B. Thomas et al. reported enolase to not be exposed on the surface[214]. However, enolase was identified to be on the surface by both methodologies presented in this chapter. The difference between the cell shaving methodologies used by Thomas et al. and the method presented in this chapter is presented in Figure 2.9. Thomas et al. investigated 'shaved cells' with the aim of recovering proteins imbedded in the cell membrane and those that are retained within the cells after shaving. In that study, *M. pneumoniae* cells were treated with trypsin first, then centrifuged, lysed, and proteins from the 'shaved cells' were separated by SDS-PAGE[214]. Incontrast, the study presented in this chapter aimed to separate and investigate the surface onle 'shaves'. This was achieved by collecting and analysing peptides that are trypsin accessible and released from proteins on the cell surface. This method would recover peripheral membrane and secreted proteins.

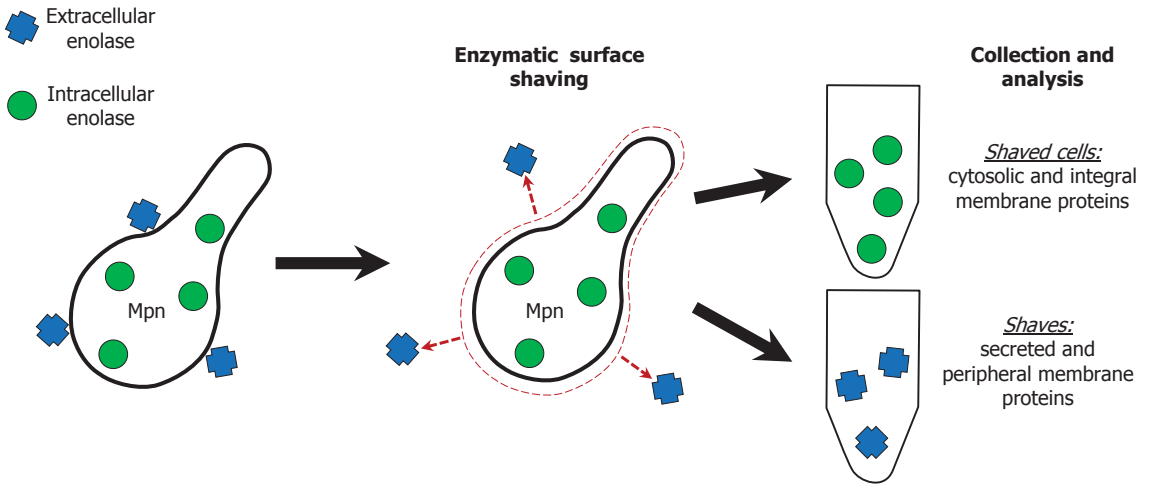


Figure 2.9: Enzymatic surface shaving of *M. pneumoniae* cells. *M. pneumoniae* cells treated with trypsin would release peripheral membrane proteins as the 'shaves'. Integral membrane proteins would be retained in the 'shaved cells'.

Tryptic peptides identified from the 'shaves' fraction spanned the entire protein indicating that the whole enolase enzyme is exposed on the surface of *M. pneumoniae*. It could be that either enolase of *M. pneumoniae* is secreted into the extracellular milieu and only a small percentage is retained on the cell surface, or it is a peripheral membrane protein. In both cases, enolase would not be identified in 'shaved' cells. Surface accessibility and different targets for the enzymes, antibodies, and tags could account for differences in results between surface studies. For instance, the antibodies used in immunofluorescence microscopy would have different sites on acetate kinase[209] compared to accessible lysine and arginine residues for trypsin shaving. As stated in the previous sections, proteins could potentially exist both intra- and extracellularly with different populations for each phase. Proft and Herrmann reported that elongation factor G to be cytosolic but also indicated a small population to be membrane bound[270].

2.5 Conclusion

A complete surface proteome is a valuable resource for the development of novel intervention strategies, such as the identification of novel vaccine candidates. With the identification of several key metaboilic proteins on the surface of *M. pneumoniae*, new vaccines could target those that moonlight as adhesins. The surface exposed regions uncovered by trypsin shaving provide direct targets for vaccines against these metaboloic proteins. Vaccines could also target crucial proteins on the attachment organelle responsible for motility, attachment organelle formation, or adhesin localisation.

The surface proteome of *M. pneumoniae* also lists moonlighting proteins with novel transmembrane domains and signal secretions (addressing Aim I). This chapter highlights the limitations of bioinformatic tools in predicting the cellular locations of proteins. Although useful to predict the location of certain classes of secreted proteins, the bioinformatic analyses in this study predicted 13% of the 160 surface proteins to be classically secreted to the cell, and ~60% of the 160 surface proteins to be localised elsewhere. The observation that most of the proteins do not have a recognisable secretion signals suggests that there are roles for these anchorless proteins on the surface of *M. pneumoniae*. However, this hypothesis warrants further investigation. This chapter presents findings that highlight limitations in the ability of existing algorithms to predict the presence of surface accessible proteins. If the sequences of these domains or signals can be determined, bioinformatics can be better optimised and then employed to find other moonlighting proteins with these sequences in other bacteria. All these moonlighters may potentially have a role as virulence factors or some other function yet to be discovered.

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Chapter 3.
Characterising potential adhesins of *Mycoplasma pneumoniae*

i3 ithreeinstitute

3.1 Preface

With the *M. pneumoniae* surface proteome completed, the next step was to use 'Bait and Prey' affinity chromatography approaches to investigate the binding abilities of proteins in the surface proteome. During the analyses of proteins retained during affinity chromatography, a number of protein fragments were identified suggesting that post-translationally processed proteins may retain the ability to bind to host ligands. An analysis of these processed proteins allowed regions within proteins that potentially bind host molecules to be mapped. Therefore, the number of proteins that also exist as fragments was also investigated (Aim III) in this chapter.

3.2 Introduction

M. pneumoniae has several dedicated adhesins and an assortment of moonlighting proteins with adhesive properties (as mentioned in Chapter 2) that bind to several host molecules. So far, *M. pneumoniae* has been show to interact with a wide range of host target molecules including: sialylated receptors[271], oligosaccharides[145], glycolipids[144], glycoproteins[146], fibronectin[205,210,213], fibrinogen[208,210,213], plasminogen[209,213-215], lactoferrin[210,213], laminin[210,213], and vitronectin[210,213]. Some surface proteins are able to degrade the extracellular matrix protein vitronectin, and the blood clotting protein fibrinogen[209,210,213,215]. The activation of plasminogen to plasmin on the surface of bacterial pathogens represents a mechanism to degrade extracellular matrix of host tissue to assist in colonising and invading epithelial cells (reviewed in[211] and[212]).

To develop a lasting infection, *M. pneumoniae* must overcome clearance by the mucociliary escalator, and also evade the host immune response. The attachment organelle facilitates binding to cilia and other cell surfaces and provides a mechanism to enable *M. pneumoniae* to traffic to a preferred niche[117-119]. *M. pneumoniae* can evade the host immune response by randomly varying the structure of three key surface proteins located at the tip of the attachment organelle by genetic recombination[27-32]. *M. pneumoniae* may also rely on post-translational proteolytic processing (also called cleavage) to release immunomodulatory peptides from lipoprotein precursors[228,229]. Immunomodulatory lipopeptides have been described in several *Mycoplasma spp.* suggesting it used frequently by Mollicutes to regulate the host immune response[231-235].

With the surface proteome of *M. pneumoniae* completed (Chapter 2), a systems-wide 'Bait and Prey' affinity protocol was employed to identify proteins from *M. pneumoniae* that bind a range of host molecules (Aim II). Specifically, the proteins and the complexes they form which display an affinity for lung epithelium surface protein complexes, actin, fetuin, fibronectin, heparin, and plasminogen were investigated to shed insight into *M. pneumoniae* proteins that may be involved in colonising host tissue surfaces.

3.3 Methodology

3.3.1 Strains and cultures

M. pneumoniae (M129 strain, ATCC 29342) cells were cultured as described previously[250]. Cells were grown in modified Hayflick's medium in tissue culture flasks (BD Falcon™) at 37°C.

Human lung carcinoma (A549 cell line; ATCC CCL-185) cells were cultured in RPMI 1640 medium (Invitrogen) supplemented with 10% heat inactivated fetal bovine serum. Cells were grown in tissue culture flasks (BD Falcon™) at 37°C with 5% CO_2.

3.3.2 Host 'Bait 'molecules

Purified fibronectin (Code: 341635) and plasminogen (Code: 528175) from human plasma was supplied by Merck Millipore. Bovine actin (Code: A3653) and fetuin (Code: F3004) was supplied by Sigma.

3.3.3 'Bait and Prey' affinity chromatography using host 'Bait' molecules

1 mg of the purified host molecules listed above was biotinylated and bound to Avidin Agarose (Thermo Fisher Scientific). *M. pneumoniae* cells were grown in a T-175 culture flask as described above until confluent. Cells were washed with PBS and lysed with 1% (w/v) C7bZ0 (Sigma) in PBS. *M. pneumoniae* lysates were then incubated with the Avidin Agarose beads for 16 h at 4°C. The beads were then packed into a glass column by gravity, washed with 5 ml of PBS four times and proteins eluted with 2 ml of 7 M urea, 2 M thiourea, 40 mM Tris-HCl (pH 8.8), 1% (w/v) C7bZ0 four times. Eluents were pooled, concentrated using a 3K MWCO filter (Pall), and precipitated with acetone before separation by SDS-PAGE. The pooled elution lane was divided into sections listed in

Appendix 4: Tables 2 – 5. The sections were dehydrated and incubated with trypsin for 16 h at 37° prior to analysis by LC-MS/MS as described in Chapter 2.

3.3.4 'Bait and Prey' affinity chromatography using heparin as 'Bait'

M. pneumoniae cells were grown in a T-175 culture flask as described above until confluent. Cells were washed with PBS and lysed with 10 mM sodium phosphate, 0.1% Triton TX-100, pH 7. Solubilised protein (300 µg) was loaded at 0.5 $mL.min^{-1}$ onto a 1 mL HiTrap Heparin HP column (GE healthcare) in binding buffer (10 mM sodium phosphate, pH 7) with a Waters 2690 Alliance LC separations module. Flow through from the column was collected at three minute intervals. With a flow rate of 0.5 $mL.min^{-1}$, the column was washed in binding buffer for 20 minutes followed by two elution steps (elution buffer: 10 mM sodium phosphate, 2 M sodium chloride, pH 7): first, a 25 minute gradient of zero to 50% elution buffer; followed by a 10 minute gradient of 50 – 100% elution buffer. Eluents were concentrated and separated by SDS-PAGE. The gel lanes were divided into sections listed in Appendix 4: Table 6. The sections were dehydrated and incubated with trypsin for 16 h at 37° prior to analysis by LC-MS/MS as described in Chapter 2.

3.3.5 'Bait and Prey' affinity chromatography using surface A549 complexes as 'Bait'

A549 and *M. pneumoniae* cells were grown separately in a T-175 culture flask as described above until confluent. A549 cells were washed three times with PBS and biotinylated with 10 mM EZ-link sulfo-NHS-biotin (Thermo Fisher Scientific) in PBS on ice for 30 s. The reaction was quenched with 50 mM Tris-HCl in PBS for 10 min at 25°C. Cells were washed three times, and lysed with 1% (w/v) C7bZ0 (Sigma) in PBS. *M. pneumoniae* cells were washed three times with PBS and lysed with 1% (w/v) C7bZ0 (Sigma) in PBS. Biotinylated A549 protein lysate was allowed to bind to Avidin Agarose beads (Thermo Fisher Scientific) for 16 h at 4°C. The beads were washed with 5 ml of PBS four times before being incubated with the native *M. pneumoniae* lysate for 16 h at 4°C. The beads were gravity packed into a column, washed again with 5 ml of PBS four times and bound proteins were eluted with 2 ml of 7 M urea, 2 M thiourea, 40 mM Tris-HCl (pH 8.8), 1% (w/v) C7bZ0 four times. A second elution step to remove biotinylated A549 proteins was performed using 30% acetonitrile in with 0.2% trifluoroacetic acid. Elutions were pooled using a 3K MWCO filter (Pall) into an 'Urea' fraction and an 'Acid' fraction for the two different elution steps. The two fractions were separated by by SDS-PAGE and divided into sections listed

in Appendix 4: Table 7. The gel sections were dehydrated and incubated with trypsin for 16 h at 37° prior to analysis by LC-MS/MS.

3.4 Results and Discussion

3.4.1 Host molecule binding proteins of *M. pneumoniae*

An *M. pneumoniae* cell lysate prepared under non-denaturing conditions were allowed to interact with different biotinylated host molecules were coupled with Avidin Agarose. Columns were coupled with biotinylated lung carcinoma (A549) surface protein complexes, cytoskeletal actin, sialic acid rich fetuin, the extracellular matrix protein fibronectin, the glycosaminoglycan mimic heparin, or the protease proenzyme plasminogen. Aside from actin and heparin, the other 'Bait' proteins used have been shown to be *M. pneumoniae* targets in previous studies[146,205,208-210,213-215,271]. Both heparin[217,218,224,272] and actin[273,274] (Raymond, B.B.A., unpublished) have been reported to be targets for binding in other Mycoplasmas.

From all six of the 'Bait and Prey' columns used in these studies, a total of 337 *M. pneumoniae* proteins were identified (Appendix 2). As this methodology employs a native extraction of *M. pneumoniae* proteins, it cannot be assumed that all 337 proteins directly bind the host molecules tested as some may belong to a complex where another protein in the complex directly binds the 'Bait'. The number of *M. pneumoniae* proteins identified from each column is seen in Figure 3.1. Eluents from columns coupled with heparin and actin contained 261 and 221 proteins, respectively. Some proteins enriched by heparin chromatography could be due to electrostatic interactions between positively charged amino acids and the negative charge of the heparin[275]. Although these proteins might not be binding directly to the sugar moieties in the glycosaminoglycan chain, positively charged residues have been shown to be essential in heparin interactions[276]. Non-specific binding could also arise from actin chromatography as *M. pneumoniae* complexes could potentially bind with the ATP molecule bound to each actin monomer instead of the monomer directly[277]. These "first pass" experiments are designed to identify potential binding interations between the 'Bait' host molecules and 'Prey' proteins. More sensitive and quantitative methods such as microscale thermophoresis, ELISA, isothermal titration calorimetry, or surface plasmon resonance are needed to examine these molecular interactions.

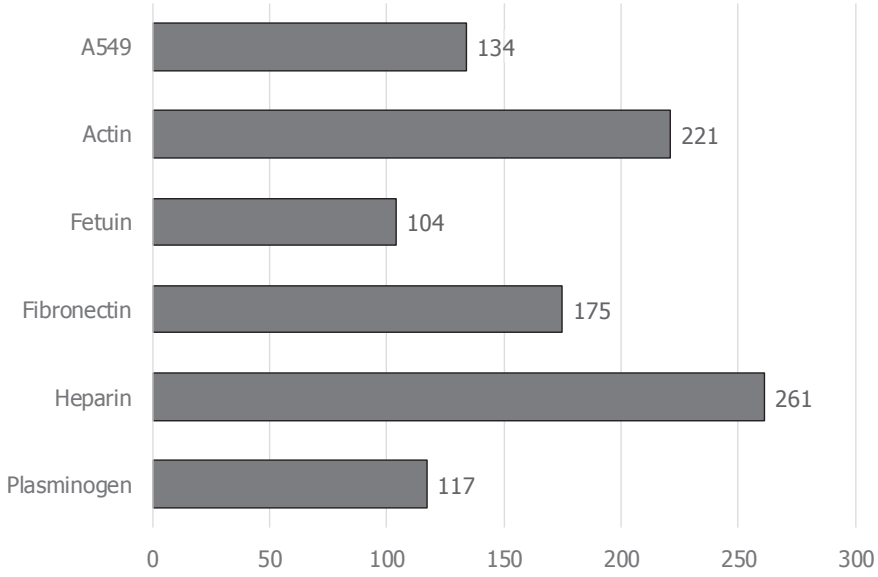


Figure 3.1: Number of *M. pneumoniae* proteins identified from each 'Bait and Prey' affinity chromatography.

Of the 337 proteins, 236 *M. pneumoniae* proteins were identified in at least two of the columns used. 44 proteins were identified in all six columns (Figure 3.2) which suggests that this subset either have multiple binding sites, or are frequent components of different *M. pneumoniae* complexes. These proteins span different protein groups such as attachement organelle proteins, chaperones, glycolytic enzymes, ribosomal proteins, and uncharacterised proteins (Table 3.1). A majority of these proteins have predicted transmembrane domains and are also predicted to not be secreted (Table 3.1) which suggests novel secretion mechanisms discussed in Chapter 2. Interestingly, 35 of the 44 (80%) of these proteins were identified by LC-MS/MS at a lower mass which suggests post-translational processing. All 44 proteins contained at least two putative heparin binding motifs (Table 3.1). Taken together these predictions inidicate the possiblilty of these proteins having multiple binding sites however further analysis is required to verify binding. The three proteins covered in detail in this thesis (Chapters 4, 5, and 6) consists of these 44 proteins.

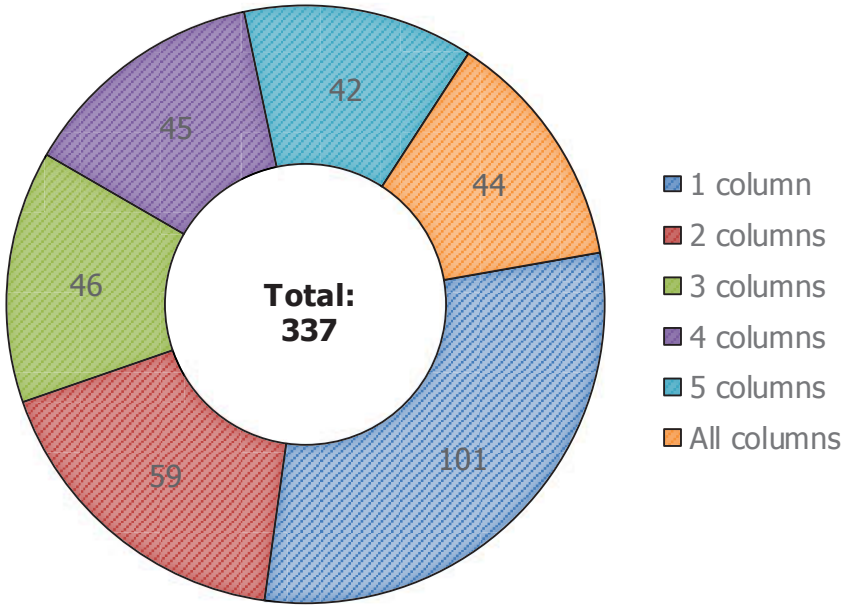


Figure 3.2: Number of *M. pneumoniae* proteins identified from the six 'Bait and Prey' chromatography columns. The numbers indicate the number of proteins identified across all six (orange) columns; at least five (teal), four (purple), three (green), or two (red) columns; or only identified in one column (blue).

Table 3.1: Proteins identified in all six 'Bait and Prey' chromatography experiments.

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Secretion	Tm	Hep motifs	Cleav
P11311	Adhesin P1	*mgpA*	1627	176.2	8.5	Class.	6	13	Y
P22446	DNA gyrase subunit A	*gyrA*	839	93.3	8.5	-	2	14	Y
P22447	DNA gyrase subunit B	*gyrB*	650	73.8	5.7	-	1	13	Y
P23568	Elongation factor Tu	*tuf*	394	43.1	6.1	-	0	12	Y
P75053	Purine nucleoside phosphorylase DeoD-type	*deoD*	238	26.3	6.7	-	1	2	-
P75062	Uncharacterized lipoprotein MG040 homolog	*MPN_052*	657	71.6	9.3	Class.	3	10	Y
P75064	Glycerol kinase	*glpK*	508	56.6	8.7	-	1	10	Y
P75089	Fructose-bisphosphate aldolase	*fba*	288	31.0	6.4	-	0	2	-
P75120	ATP-dependent zinc metalloprotease	*ftsH*	709	77.7	9.1	-	2	14	Y
P75189	Enolase	*eno*	456	49.2	6.1	-	1	5	Y
P75205	10 kDa chaperonin	*groS*	116	12.6	9.0	Non	0	2	-
P75206	Probable cytosol aminopeptidase	*pepA*	445	48.8	8.5	-	1	7	-
P75223	Uncharacterized protein MG377 homolog	*MPN_555*	193	22.4	5.6	-	0	7	Y
P75239	50S ribosomal protein L7/L12	*rplL*	122	13.1	4.9	-	1	2	-
P75247	Chaperone protein ClpB	*clpB*	715	81.3	8.1	-	1	20	Y
P75271	DNA-directed RNA polymerase subunit beta	*rpoC*	1290	144.8	9.3	-	0	21	Y
P75305	FMN-dependent NADH-azoreductase	*azoR*	197	21.5	8.7	Non	3	2	-
P75310	Uncharacterized protein MG328 homolog	*MPN_474*	1140	118.0	5.0	Non	0	16	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Secretion	Tm	Hep motifs	Cleav
P75327	Uncharacterized lipoprotein MG321 homolog	*MPN_456*	1005	110.4	6.3	Class.	4	11	Y
P75344	Chaperone protein DnaK	*dnaK*	595	65.1	5.4	-	0	11	Y
P75358	Glyceraldehyde-3-phosphate dehydrogenase	*gapA*	337	36.8	8.9	-	0	7	Y
P75389	Probable NADH oxidase	*nox*	479	52.8	6.6	-	3	6	Y
P75390	Pyruvate dehydrogenase E1 component subunit alpha	*pdhA*	358	40.6	6.2	-	1	4	Y
P75391	Pyruvate dehydrogenase E1 component subunit beta	*pdhB*	327	35.9	6.5	-	2	4	Y
P75393	Dihydrolipoyl dehydrogenase	*pdhD*	457	49.4	8.4	-	3	6	Y
P75394	Probable lipoate-protein ligase A	*lplA*	339	39.2	6.2	-	0	6	-
P75493	Uncharacterized lipoprotein MPN_284	*MPN_284*	794	87.1	9.2	Class.	2	15	Y
P75539	Replicative DNA helicase	*dnaB*	473	54.5	5.6	-	0	8	Y
P75543	30S ribosomal protein S6	*rpsF*	215	25.4	9.8	Non	0	6	Y
P75556	Uncharacterized protein MG075 homolog	*MPN_213*	1030	115.9	5.9	Class.	2	18	Y
P75569	PTS system glucose-specific EIICBA component	*ptsG*	940	101.6	9.4	Non	10	7	Y
P75591	Transcription termination/antitermination protein NusA	*nusA*	540	60.2	5.9	Non	0	9	Y
P78007	L-lactate dehydrogenase	*ldh*	312	33.9	8.4	-	1	5	Y
P78017	Protein GrpE	*grpE*	217	24.7	7.8	-	0	4	Y
P78018	Phosphoglycerate kinase	*pgk*	409	44.2	7.1	-	2	5	Y
P78027	Ribonucleoside-diphosphate reductase subunit alpha	*nrdE*	721	82.3	6.6	-	1	12	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Secretion	Tm	Hep motifs	Cleav
P78031	Pyruvate kinase	*pyk*	508	57.2	9.2	-	1	8	Y
Q50295	DNA-directed RNA polymerase subunit alpha	*rpoA*	327	36.6	7.0	-	2	4	-
Q50327	ATP synthase subunit b	*atpF*	207	24.0	5.6	Non	0	4	-
Q50329	ATP synthase subunit alpha	*atpA*	518	57.3	6.2	-	1	6	Y
Q50331	ATP synthase subunit beta	*atpD*	475	52.2	5.5	-	0	6	Y
Q50341	Mgp-operon protein 3	*MPN_142*	1218	130.4	8.0	Class.	5	14	Y
Q50360	Cytadherence high molecular weight protein 3	*hmw3*	672	73.7	4.7	Non	0	9	Y
Q50365	Cytadherence high molecular weight protein 1	*hmw1*	1018	112.1	4.0	Non	0	5	Y

Cleav stands for whether or not the same protein was identified at a lower mass on SDS-PAGE from either of the affinity chromatography experiments, a Y in this column represents that a cleavage fragment was identified by LC-MS/MS. Results in the Secretion column are from a combination of SignalP 4.1 and SecretomeP predictions of classically (Class.), non-classically (Non), or not (-) secreted proteins. Results in the Tm column are from TMpred predictions where the number listed indicates the number of predicted transmembrane domains. Results in the Hep motifs column are from ScanProsite searches of clusters of basic residues (either x-[HKR]-x(0,2)-[HKR]-x(0,2)-[HKR]-x or x-[HKR]-x(1,4)-[HKR]-x(1,4)-[HKR]-x) where the number listed indicates the number of putative heparin binding motifs.

The three adhesins of *M. pneumoniae*: P1 (Uniprot #: P11311), P30 (Uniprot #: P75330), and P116 (Uniprot #: P75556) were identified in the eluents from the fetuin chromatography column reinforcing their affinity for sialic acid residues[278] (Appendix 2). Previous studies have shown several moonlighting proteins bind fibronectin, plasminogen, and A549 cells (Table 3.2)[205,209,210,213-215]. Moonlighting proteins were identified in the eluents from columns coupled with fibronectin, plasminogen, or A549 surface proteins (Table 3.2). Four proteins: Pdh-A, Pdh-C, lactate dehydrogenase, and pyruvate kinase, were identified in this study but were not shown to have different binding affinities in previous studies (in Table 3.2). It is possible that due to the native extraction approach used that these four proteins are part of complexes where another protein in that complex binds to the 'Bait'. Alternatively, three proteins that have been previously shown to have binding affinity to the 'Bait' were not identified in the respective 'Bait' chromatography in this study (Phosphoglycerate mutase, transketolase, and GroEL in Table 3.2). These previous studies generated recombinant proteins to measure the binding affinity. Because of the native extraction method used in the study presented in this chapter, the arrangement of these proteins within the complex could differ to the recombinant form used in previous studies. It could also be possible that either of these three proteins do not have an accessible 'Bait' binding site and therefore could not be retained during affinity chromatography. Another reason might be due to the low abundance of either proteins in the samples collected from 'Bait and Prey' chromatography and thus were outside the dynamic range of mass spectrometer used. Phosphoglycerate mutase has significantly lower number of copies per cell when compared to the more abundant Ef-Tu or Pdh-B based on protein quantification data[46,56].

Table 3.2: *M. pneumoniae* moonlighting proteins.

	Published			This study		
	Fn[e]	Plg[d]	A549[d]	Fn	Plg	A549
Ef-Tu	Y[a]	N/A	N/A	Y	Y	Y
Pdh-A	-	Y[c]	Y	Y	Y	Y
Pdh-B	Y[a]	Y[b]	Y	Y	Y	Y
Pdh-C	Y	Y[c]	-	Y	Y	Y
GAPDH	Y	Y	Y	Y	Y	Y
Lactate dehydrogenase	-	Y	Y	Y	Y	Y
Phosphoglycerate mutase	-	Y	Y	-	-	Y
Pyruvate kinase	-	Y	Y	Y	Y	Y
Transketolase	-	Y	Y	-	-	Y
GroEL[f]	Y	Y	Y	Y	-	Y
DnaK[f]	Y	Y	Y	Y	Y	Y

'Bait' proteins were shortened to: Fn for fibronectin, Plg for plasminogen, and A549 for human A549 epithelial cells. Y under any of these columns in the table indicates whether or not the protein identified to bind to the 'Bait'. N/A indicates that the particular interaction was not tested. [a]Data collected from Dallo et al., 2002[205]. [b]Data collected from Thomas et al., 2011[214]. [c]Data collected from Gründel et al., 2015[215]. [d]Data for the entire column was collected from Gründel et al., 2015[209] unless specified. [e]Data for the entire column was collected from Gründel et al., 2016[210] unless specified. [f]Data for the entire row was collected from Hagemann et al., 2017[213].

Lastly MPN554 (Uniprot #: P75224), a single-stranded DNA binding protein[279], was only identified in heparin affinity chromatography. This is expected because heparin and DNA have a net negative charge[275]. Combined, this result and the results reported above suggest that 'Bait and Prey' affinity chromatography is able to successfully enrich for proteins that potentially bind the 'Bait'; though further analysis is needed to validate these preliminary findings and quantify binding.

3.4.2 Potential moonlighting adhesins of *M. pneumoniae*

By combining the *M. pneumoniae* surface proteome (Chapter 2, Appendix 1) with the listed 'Bait and Prey' proteins from this chapter (Appendix 2), 149 of the 160 proteins in the surface proteome display an affinity for at least one of the host molecules used in this study (Figure 3.3, Appendix 3). It should be reiterated again at this point that native cell lysates were used in these experiments so the 337 proteins recovered from affinity chromatography either: i) bind the host molecules directly, or ii) are part of a protein or molecular complex where another protein or biomolecule binds the host molecules. This data suggests that many proteins that localise to the surface of *M. pneumoniae* display the ability to interact with host molecules. Further analyses such as ELISAs, or microscale thermophoresis is required to validate the interactions between proteins in the surface proteome and host proteins used in these studies. Such studies would require purifying recombinant versions of each of the *M. pneumoniae* proteins in the surface proteome.

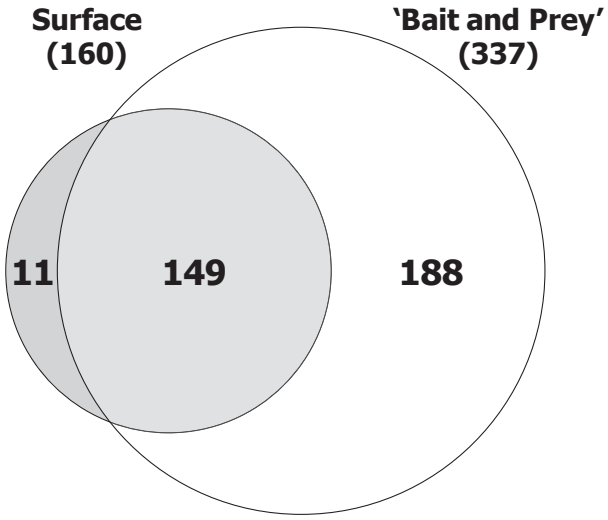


Figure 3.3: Venn diagram of M. pneumoniae surface proteins and affinity host molecules. The 'Bait' host molecules used in this study are A549 surface proteins, actin, fetuin, fibronectin, heparin, and plasminogen. Diagram was generated using Venny 2.1.0[252] <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.

The 149 identified proteins represent ribosomal proteins, lipoproteins, chaperones, glycolytic proteins, proteins of the attachment organelle, elongation factors, proteases, transcription factors, and a considerable number of uncharacterised proteins just to name

a few (Appendix 3). Bioinformatic analyses from Chapter 2 predict that 55 of these proteins are secreted (classical or non-classical) to the cellular surface and 68 contain at least one predicted transmembrane domain. There has only been one novel transmembrane identified in the Mycoplasma genus. It was identified in *Mycoplasma gallisepticum* and carries the sequence (gly-x-x-x-gly-x-x-x-gly)[280]. A recent study in *Mycoplasma pulmonis* suggested that anchorless surface proteins might be tethered to the cell membrane by the sugar rhamnose[206,207]. Further work is needed to determine if sugars play a role in tethering anchorless proteins on the *M. pneumoniae* cell surface; and if the recently described transmembrane domain, or a derivative of it, can be identified in the *M. pneumoniae* surface proteome.

From this list of 149 proteins, 111 proteins were recovered from columns coupled with heparin. Heparin is an *in vitro* substitute for the glycosaminoglycans found abundantly on host cell surfaces and extracellular matrices[281]. Heparin also mimics the negative charges of DNA and therefore can also potentially enrich for nucleic acid binding proteins[275]. Further experiments will be required to determine whether binding was due to electrostatic interactions or via direct binding with sugar residues in the glycosaminoglycan chains. Several pathogens have been shown to interact with glycosaminoglycans on the host cell surface (reviewed in[281,282]). Heparin has also been shown to promote biofilm formation in *S. aureus*[283] and *Lactobacillus rhamnosus*[284] by increasing cell-cell interactions. The presence of heparin binding proteins on the surface of *M. pneumoniae* may facilitate interactions with extracellular DNA to aid in biofilm formation. Extracellular DNA has been shown to be responsible for biofilm formation and is a major constituent of the biofilm matrix[261,264].

Several homologs of these 149 proteins have been shown in a wide range of other pathogens to be virulence factors (Reviewed in[195] &[194]). Some of the moonlighters that function as adhesins in other pathogens are outlined in Table 3.3. Furthermore, moonlighting functions are not limited to adherence as some proteins are able to modulate immune responses such as GAPDH from *Streptococcus agalactiae* that initiates B cell responses[285] and GroEL of *L. johnsonii* stimulating interleukin-8 secretion in macrophages[286].

Table 3.3: Moonlighting proteins previously shown to bind host molecules that were identified in 'Bait and Prey' chromatography experiments.

Moonlighter	Species	Binds	Reference
Ef-Tu	*M. pneumoniae*	Fibronectin	205
Pdh-B	*M. pneumoniae*	Fibronectin and plasminogen*	205,210,214,215
Pdh-A	*M. pneumoniae*	Plasminogen*	210,215
Pdh-C	*M. pneumoniae*	Fibronectin and plasminogen*	209,210,215
Lactate dehydrogenase	*M. pneumoniae*	Plasminogen*	209,210
Pyruvate kinase	*M. pneumoniae*	Plasminogen	209,210
GroEL	*M. pneumoniae*	Fibronectin and plasminogen*	213
GAPDH	*M. pneumoniae*	Fibronectin and plasminogen	209,210
	Candida albicans	Fibronectin	287
	L. crispatus	Plasminogen	288
	Group A streptococci	Actin and fibronectin	289
	Group B streptococci	Actin	290
DnaK	*M. pneumoniae*	Fibronectin and plasminogen*	213
	Bifidobacterium animalis subsp. Lactis	Plasminogen	291
	Neisseria meningitides	Plasminogen	292

* indicates that the moonlighting protein is able to promote the conversion of bound plasminogen to plasmin in the presence of plasminogen activators to degrade host proteins.

3.4.3 Evidence of cleavage in *M. pneumoniae*

M. pneumoniae proteins have been identified in previous studies as cleaved fragments rather than full length proteoforms, but these observations: have not been fully investigated, have rarely been characterised, and precise cleavage sites are yet to be mapped. Paraformaldehyde based cross-linking studies identified a DnaK fragment and a C-terminal P1 fragment forming a complex with full length P1 adhesin and other accessory proteins[247]. Regula et al. found fragments of different lipoproteins (Mpn052, Mpn284, Mpn288, Mpn376, Mpn400, Mpn408, Mpn444, Mpn456, Mpn474 and Mpn491) from mass spectrometry analysis of 2D-SDS PAGE of *M. pneumoniae* cell lysates[248]. Precise cleavage sites in several *M. pneumoniae* proteins have been mapped. The most characterised cleavage event in *M. pneumoniae* is the cleavage of Mpn142 to generate P40 and P90[166-168]. Other cleavage events reported include: the signal peptide for Mpn142 (at the start of P40)[167,168] and the P1 adhesin[163,245,246]; and removal of N-terminal lipo-peptides of lipoproteins, Mpn611, Mpn162[228], and ATPase (Mpn602)[229].

An analysis of proteins recovered from the 'Bait and Prey' experiments revealed that 108 proteins also existed as smaller protein fragments (Appendix 2). In the past, these fragments were reported as artefact or evidence of protein degradation[293]. However, data from 'Bait and Prey' affinity experiments, 2D-SDS PAGE, and N-terminal dimethyl labelling experiments (performed by Berry, I. J.) suggest that protein cleavage may represent an important biological process in *M. pneumoniae*.

An example of a protein that was identified as the full length protein and as cleavage fragments is the uncharacterised lipoprotein MPN_284 (UniProt #: P75493, Figure 3.4). From my datasets, an N-terminal fragment (F_H), a C-terminal fragment (F_B), a fragment spanning the N-terminal half (F_{C2}), and two fragments spanning the C-terminal half (F_{C1} and F_F) were identified (Madonis, K. assisted with the recovery of this data and its compilation). N-terminal dimethyl labelling (performed by Berry, I. J.) identified three cleavage sites: ^{186}T↓S^{187}, ^{235}R↓S^{236}, and ^{552}K↓Q^{553} associated with these five fragments (Figure 3.4). The mass and pI of these fragments were predicted by ProtParam[268] and compared to the observed mass from SDS-PAGE (Table 3.4). Two fragments (F_B and F_F) were identified by LC-MS/MS analysis of 2D-SDS PAGE spots of *M. pneumoniae* cell lysates in a previous study[248] (Figure 3.4 and Table 3.4). Two fragments (F_{C1} and F_B) were

identified from several spots of the same mass but different pI. This suggests different proteoforms of the same fragment (Table 3.4). This example demonstrates how 2D-SDS PAGE, affinity chromatography, and N-terminal dimethyl labelling can be combined to uncover the cleavage fragments of proteins of interest. The function of these fragments remains to be characterised but pulldown experiments indicate that fragments may retain an ability to bind to host bait molecules. Cleavage events in three abundant *M. pneumoniae* surface proteins will be covered in depth in following chapters of this thesis.

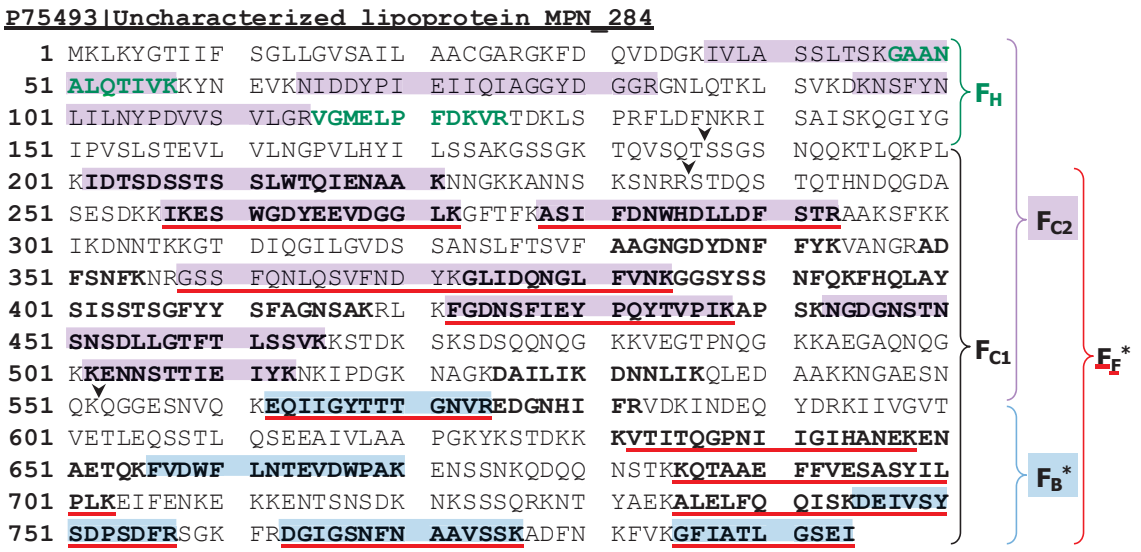


Figure 3.4: Amino acid sequence of the uncharacterized lipoprotein MPN_284 and cleavage fragments derived from this lipoprotein. For simplicity, only five fragments have been included: one fragment was recovered during heparin agarose chromatography (F_H), two from 2D-SDS PAGE of cell lysate (F_{C1} and F_{C2}), one fragment from surface biotin labelling (F_B), and one recovered during fibronectin affinity chromatography (F_F). The range of amino acids that these fragments span and information on molecular weight and pI can be found in Table 3.4. Peptides in green and in bold text were identified from eluents of heparin chromatography. Peptides in black and in bold text were identified from a 2D-SDS PAGE spot of *M. pneumoniae* cell lysates. Peptides in purple boxes were identified from a different 2D-SDS PAGE spot. Peptides in blue boxes were identified from eluents of avidin chromatography of biotinylated surface proteins. Peptides underlined in red were identified from eluents of fibronectin chromatography. Cleavage sites from dimethyl labelling experiments of *M. pneumoniae* are indicated by the black arrows. * indicates two fragments that were also identified by Regula et al.[248]. Figure is adapted from data analysed by Madonis, K.

Table 3.4: Fragments of lipoprotein MPN_284.

Fragment	Total # of amino acids	Range	Predicted mass (kDa)	Predicted pI	Observed mass (kDa)	Observed pI
Full length	794	$M^{1} – I^{794}$	87.2	9.2	~85	4.9
F_H (heparin)	186	$M^{1} – T^{186}$	20.0	9.5	18 – 25	N/A
F_{C1} (2D-PAGE)	608	$S^{187} – I^{794}$	67.2	8.9	~70	$>7^{+}$, 7.0, 7.2, 7.7
F_{C2} (2D-PAGE)	552	$M^{1} – K^{552}$	60.1	9.5	~55	4.7
F_B (biotin)*	242	$Q^{553} – I^{794}$	27.1	5.5	~32	5.0, 5.2, 5.3
F_F (fibronectin)*	559	$S^{236} – I^{794}$	61.8	8.1	50 – 63	N/A

Predicted mass and pI calculated with ProtParam[268]. $^{+}$ identified from 2D-SDS PAGE between pH range of 4 – 7. * fragments also identified by Regula et al.[248].

In addition to the proteolytic processing events in *M. pneumoniae* listed above, cleavage has been well characterised in a variety of different pathogens including: *Campylobacter jejuni*[239], *S. aureus*[237,238], *Rickettsia rickettsia*[241], *Porphyromonas gingivalis*[240], *Mycoplasma fermentans*[232-234], *Mycoplasma gallisepticum*[231], *Mycoplasma genitalium*[235], *Spiroplasma citri*[236], and especially in *M. hyopneumoniae*[186,187,216-227,244]. A majority of these cleavage fragments produced by *M. hyopneumoniae* bind heparin and contain bioinformatically predicted heparin binding regions[186,187,217,218,221,222].

It is possible that cleavage occurs on the *M. hyopneumoniae* cell surface after the full-length parent protein has been translocated across the cell membrane. Six proteases have been identified to reside on the surface of *M. hyopneumoniae*[227]; two of which have been show to moonlight as adhesins[243,244]. The surface proteome of *M. pneumoniae* reveals that there are also six proteases on the surface of *M. pneumoniae* (Appendix 1). More work is needed to determine whether or not these proteases are responsible for the cleavage reported in this chapter.

3.5 Conclusion

'Bait and Prey' affinity chromatography was used to identify proteins in *M. pneumoniae* that potentially bind to host molecules. A large number of *M. pneumoniae* proteins were identified during affinity chromatography studies indicating that protein complexes may complicate the interpretation of binding interactions with host molecules. However, a affinity chromatography approach is useful in efforts to recover cleavage fragments and how they may function. Cleavage sites that give rise to these fragments were in many cases independently verified by mass spectrometry of proteins fractionated by size during SDS-PAGE (1D and 2D) and using the N-terminal dimethyl labelling protocol developed by my colleagues. The value of this approach was exemplified by an analysis of five fragments of lipoprotein MPN_284 used as a model protein here (Figure 3.4).

In this study, 337 *M. pneumoniae* proteins, 48.6% of the expressed proteome, were identified from affinity chromatography experiments using different host molecules as bait (Appendix 2). These experiments are preliminary in nature and further studies are required to confirm binding and quantify binding affinity. To address Aim II, the list of 149 *M. pneumoniae* proteins that are surface exposed and recovered from 'Bait and Prey' experiments provide a list of putative adhesins (Appendix 3). Lipoprotein MPN_284 could function as a potential adhesin as it was recovered in all of the 'Bait and Prey' experiments. Some of the identified moonlighting proteins could be targeted for vaccine development as seen with enolase[197-199] and GAPDH (reviewed in[200]) in other bacteria. Protocols presented in this chapter provide a non-hypothesis driven approach to identify novel protein function.

Though more work is required to characterise the vast number of proteins that undergo post-translational processing (Appendix 2), it is likely that these cleavage fragments assist *M. pneumoniae* during colonisation. 108 proteins were found to undergo cleavage based on data from 'Bait and Prey' and 2D-SDS PAGE of *M. pneumoniae* whole cell lysates (Appendix 2, Aim III). The investigation of proteolytic cleavage events will expand the current knowledge on the molecular interactions that occur on the bacterial cell surface with host cells during infection. This chapter has identified potential players involved in the adherence of *M. pneumoniae* to host cells. The subsequent chapters of this thesis comprise a more in-depth analyses of selected surface molecules identified in this chapter

and in Chapter 2. Three proteins from *M. pneumoniae* (an accessory protein, a putative moonlighter, and an adhesin) were investigated to determine cellular localisation, predict binding motifs, processing frequency, and potential host antigens (Aim IV). Although these are only examples, it is likely that *M. pneumoniae* employs a large number of proteins to assist in interactions with the human respiratory epithelium.

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Chapter 4.

P40 and P90 from Mpn142 are targets of multiple processing events on the surface of *Mycoplasma pneumoniae*

i3 ithreeinstitute

4.1 Preface

The data generated from Chapters 2 (surface proteome) and 3 ('Bait and Prey' chromatography) was combined with dimethyl labelling (performed by Berry, I.J., unpublished) to reveal internal proteolytic cleavage sites in the *M. pneumoniae* proteome. From the list of 149 proteins that were identified from both datasets (Appendix 3), three proteins were chosen and investigated further to determine each protein's role during the interaction of *M. pneumoniae* with the respiratory epithelium. The first protein, Mpn142, is already known to undergo two cleavage events. The article presented as Chapter 4 reports that Mpn142 undergoes secondary post-translational proteolysis on the surface of *M. pneumoniae* (Aim IV). 'Bait and Prey' chromatography coupled with bioinformatic analysis indicate the cleavage fragments of Mpn142 may function as adhesins on the cell surface.

The article presented for Chapter 4 has been peer-reviewed and was accepted for publication in the Special Issue 'Microbial Proteomics' in the 'Proteomes' Journal on the 7th December, 2015.

4.2 Author contribution

Author	Contribution
Michael Widjaja	Performed surface proteome experiments, the 'Bait and Prey' chromatography, and the bioinformatic analysis of Mpn142. Analysed all data generated in this study.
Iain J. Berry	Performed dimethyl labelling for *M. pneumoniae* and assisted in analysing dimethyl labelling data.
Elsa J Pont	Assisted in the dimethyl labelling for *M. pneumoniae*.
Matthew P. Padula	Oversaw the acquisition and assisted with interpretation of mass spectrometry data.
Steven P. Djordjevic	Initiated the study, interpreted the data, and drafted the manuscript with Micahel Widjaja.

Proteomes **2015**, *3*, 512-537; doi:10.3390/proteomes3040512

OPEN ACCESS

proteomes

ISSN 2227-7382

www.mdpi.com/journal/proteomes

Article

P40 and P90 from Mpn142 are targets of multiple processing events on the surface of *Mycoplasma pneumoniae*

Michael Widjaja [1], Iain J Berry [1], Elsa J Pont [1], Matthew P Padula [2] and Steven P Djordjevic [1,2,*]

[1] The ithree institute, The University of Technology Sydney, PO Box 123 Broadway, NSW 2007, Australia; E-Mail: michael.widjaja@student.uts.edu.au; iain.j.berry@student.uts.edu.au; elsa.pont@etudiant.univ-lille1.fr

[2] Proteomics Core Facility, University of Technology Sydney, Cnr Harris and Thomas St, Ultimo, NSW, 2007, Australia; E-Mail: matthew.padula@uts.edu.au

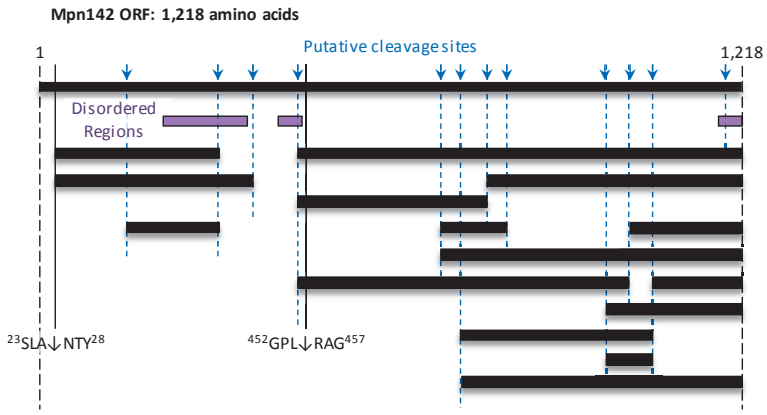
* Author to whom correspondence should be addressed; E-Mail: steven.djordjevic@uts.edu.au; Tel.: +61-2-9514-4127 (ext. 4127); Fax: +61-2-9514-4143

Academic Editor:

Received: 20 September 2015 / Accepted: 7 December 2015 / Published: 16 December 2015

4.3 Abstract:

Mycoplasma pneumoniae is a significant cause of community acquired pneumonia globally. Despite having a genome less than 1 Mb in size, *M. pneumoniae* presents a structurally sophisticated attachment organelle that i) provides cell polarity, ii) directs adherence to receptors presented on respiratory epithelium, and iii) plays a major role in cell motility. The major adhesins, P1 (*Mpn141*) and P30 (*Mpn453*), are localised to the tip of the attachment organelle by the surface accessible cleavage fragments P90 and P40 derived from Mpn142. Two events play a defining role in the formation of P90 and P40; removal of a leader peptide at position 26 (^{23}SLA↓NTY^{28}) during secretion to the cell surface and cleavage at amino acid 455 (^{452}GPL↓RAG^{457}) generating P40 and P90. LC-MS/MS analysis of tryptic peptides generated by digesting size-fractionated cell lysates of *M. pneumoniae* identified 15 cleavage fragments of Mpn142 ranging in mass from 9 – 84 kDa. Further evidence for the existence of cleavage fragments of Mpn142 was generated by mapping tryptic peptides to proteins recovered from size fractionated eluents from affinity columns separately loaded with heparin, fibronectin, fetuin, actin, plasminogen and A549 surface proteins as bait. To define the sites of cleavage in Mpn142, neo-N-termini in cell lysates of *M. pneumoniae* were dimethyl labelled and characterised by LC-MS/MS. Our data suggests that Mpn142 is cleaved to generate adhesins that are auxiliary to P1 and P30.



Abstract Figure: Cleavage map of Mpn142

Keywords: Ectodomain shedding, protein processing; multifunctional proteins

4.4 Introduction

Mycoplasma pneumoniae (*M. pneumoniae*) is a serious respiratory pathogen estimated to cause 100,000 hospitalisations annually in the USA[1] and up to 40% of cases of community-acquired pneumonia globally[294]. *M. pneumoniae* infects patients of all ages and up to 25% of affected individuals develop extrapulmonary complications at neurological, musculoskeletal, haematological and cardiovascular sites. While sporadic infections are common, outbreaks of *M. pneumoniae*-induced disease occur in schools, child-care facilities, inpatient institutions and military barracks[11]. Azithromycin and other macrolides are used for the treatment of infections caused by *M. pneumoniae*[295], but an increase in the frequency of reports of macrolide resistance, particularly in Europe, Asia and North America is of major concern[74-88]. Efficacious vaccines for the prevention of infections caused by *M. pneumoniae* are yet to be developed and are complicated by the presence of antigens that are capable of evoking an autoimmune response[90]. While the infiltration of neutrophils and lymphocytes is a characteristic immunological hallmark of infections caused by *M. pneumoniae*, the severity of the response varies widely. In severe cases, the immune response is known to generate immunopathological sequelae, complicating vaccine design[90,294].

M. pneumoniae has a small genome encoding about 700 ORFs and lacks genes needed for a TCA cycle, cell wall, amino acid and nucleotide biosynthesis[296,297]. Despite having a reduced genome capacity, *M. pneumoniae* is remarkable in that it forms a Triton X-100 insoluble cytoskeleton and complex attachment organelle that is critical for adherence to host epithelium and cellular motility[117]. The attachment organelle comprises the adhesins P1 and P30, High Molecular Weight (HMW) proteins 1, 2 and 3, P40/P90 from *mpn142* (ORF6), P65, P41 and P24[118,125,126,138,247,298]. Adhesins P1, P30 and Mpn142 products (P40/P90) are strictly localised to the extracellular side of the attachment organelle and have N-terminal transmembrane domains that form part of a signal sequence[163,167,168,245,299]. P65 and HMW1 are unusual because they reside intracellularly as part of the cytoskeletal core and on the extracellular side of the attachment organelle, suggesting the existence of different proteoforms[118]. It is not known how HMW1 traffics to the cell surface because it lacks evidence of transmembrane spanning domains and a secretion signal. HMW 1, 2 and 3, P41, P24 and P65 are all integral components of the intracellular cytoskeletal core[300].

Layh-Schmitt et al., (2000)[247] described the use of para-formalydehyde to crosslink proteins in close association with one another and identified protein complexes containing the P1 adhesin by affinity chromatography using P1 antibodies. The identities of the proteins in the complex were determined by a combination of immunoblot analysis and MALDI-TOF MS. P1 complexes contained P40 and P90, P30, P65, DnaK, pyruvate dehydrogenase subunit α, HMW1 and HMW3 proteins[247]. Notably, the P1 adhesin was found to be associated with P30, P40 and P90 when *M. pneumoniae* cells were treated with the membrane impermeable cross-linking reagent DTSSP (3, 3-dithiobis(sulfosuccinimidylproprionate)) suggesting that some of the interactions with P1 are not accessible on the extracellular side of the membrane and may occur with cytoskeletal proteins via transmembrane domains located in the P1 molecule[169]. Notably, a 480 kDa protein complex was isolated by solubilising *M. pneumoniae* proteins after cross-linking with bis[sulfosuccinimidyl] suberate (BS^3) using a non-ionic detergent and Blue Native PAGE. The complex comprises P90 and P1 in a 1:2 molar ratio and forms an appendage that allows *M. pneumoniae* to glide across surfaces[301].

While expression of P1 is essential for adherence, the presence of accessory adherence proteins is critical for the formation of a functional attachment organelle[118]. Insertion of P1 into the membrane and its trafficking to the attachment organelle is largely dependent on P90 and P40[27,302,303]. Mutants defective in the expression of P90 and P40, two important accessory proteins, allow P1 to completely partition to the Triton X-100 soluble phase. In wild type cells, P1 typically partially associates with the Triton X-100 insoluble shell[302]. Mutants defective in the production of *mpn142* are defective in cellular adherence because P1 cannot traffic to the tip structure resulting in random P1 distribution around the cell body[27,303]. The P1 and P30 adhesins concentrate at the tip of the attachment organelle and represent the dominant proteins responsible for adherence[27,58,151,154,156,161,304]. Mutants that express P1 and P30 but that lack P40/P90, or the HMW proteins 1, 2 and 3, are avirulent. As such P1 and P30 are considered to be essential but not sufficient for attachment of *M. pneumoniae* to host cells[136,152,305].

mpn140-mpn141-mpn142 comprise a polycistronic transcriptional unit presumably to ensure equimolar amounts of each of the proteins[170,306]. *mpn140* encodes for a putative phosphoesterase of 28 kDa that has been found to be expressed but still remains

functionally uncharacterised[44,166,306,307]. *mpn141* encodes the P1 adhesin, a 170 kDa protein comprised of 1,627 amino acids[163] while *mpn142* encodes a 130 kDa adherence accessory protein comprising 1,218 amino acids[166]. The function of Mpn142 is complicated by the fact that it undergoes processing shortly after translation at two sites generating an N-terminal 40 kDa protein (P40/protein C) and a C-terminal 90 kDa protein (P90/protein B)[166,167]. Notably, the predicted 130 kDa precursor is not detectable indicating that the cleavage events are efficient[166,167]. P90 and P40 were able to be linked to the P1 protein via a non-permeating, cross-linking reagent indicating that all three proteins co-localise on the extracellular side of the membrane on the tip of the attachment organelle within a distance of approximately 12 Å (1.2 nm) to one other[169]. *M. pneumoniae* mutant strain M29-B176 lacks P40 and P90 proteins and is avirulent (Hu et al., 1984); other *M. pneumoniae* mutants that are unable to express P40 and P90 are also avirulent[303,308,309].

Leader sequences in the N-terminus of P1 and Mpn142 are removed by cleavage at positions 60 and 26, respectively[168,245,246]. Cleavage at position 26 in Mpn142 was confirmed by identifying the semi-tryptic peptide ^{26}NTYLLQDHNTLTPYTPFTTPXDGGXDWR^{54} by N-terminal labelling and ESI-QTOF mass spectrometry of purified P40[168]. Cleavage at this site generates an N-terminal fragment of Mpn142 with a predicted mass of 43.9 kDa but the P40 protein migrates with an apparent mass of 36 kDa, a discrepancy of about 9 kDa. Catrein et al., (2005)[168] hypothesized that further endoproteolytic events removing as much a 9 kDa from the P40 molecule takes place but were unable to accurately determine the location of a cleavage site(s) or confirm the presence of further cleavage fragments of P40. The cleavage event after amino acid 454 creates P40 and P90. Edman degradation of the P90 protein identified the sequence ^{455}RAGNSSEDAL^{465} indicating that P90 spans amino acids 455 – 1,218 with a predicted mass of 83.7 kDa[167].

Adhesins are multifunctional proteins that comprise the different functional domains needed to bind to a range of host molecules during the colonisation of host surfaces. Many adhesins are targets of processing events that not only remove signal secretion sequences but also release a range of functional cleavage products. In support of this view, numerous mycoplasma-derived adhesins have shown to be extensively processed[186,187,216-226,231,232] as well as adhesins found in a range of Gram positive[237] and Gram negative

pathogens[239,240]. *M. pneumoniae* binds to fibronectin[205]; fibrinogen[208]; plasminogen[214,215]; sialylated receptors[27] and oligosaccharides[144,145]; sulfated glycolipids[310]; laminin, fetuin, and human chorionic gonadotropin[146] but the identities of the proteins localised to the attachment organelle that target these molecules are largely unknown. While most studies have focussed on characterising interactions between P40 and P90 with other proteins in the attachment organelle, their presence on the surface of *M. pneumoniae* in close association with P1 indicates that they may also have roles in adherence to host molecules. To investigate this, we applied systems-wide, affinity chromatography methodologies using fibronectin, actin, heparin, plasminogen, fetuin and surface–exposed proteins from A549 cells as bait to determine whether P40 and P90 may play a role in binding key host molecules. We not only recovered P40 and P90 but also found strong evidence that P90 and P40 are subject to further processing events. While further studies are needed to firmly establish the roles of P90 and P40 in adherence to host molecules, LC-MS/MS confirmed that P90 and P40 are processed at multiple sites. To determine precise cleavage sites in P90 and P40, we dimethyl labelled all proteins in whole cell lysates of *M. pneumoniae* and characterised neo-N-termini in P90 and P40 by LC-MS/MS. These studies provide new insights in the structural and functional capabilities of P90 and P40 and identified regions within these two molecules worthy of further study.

4.5 Experimental Section

4.5.1 Strains and cultures and reagents

The M129 *M. pneumoniae* strain was cultured in modified Hayflick's medium at 37°C in tissue culture flasks as described previously[250]. Modified Hayflick's medium contained 21 g PPL broth base without crystal violet, 5 g of D-glucose, 4 ml of 0.5% phenol red, 100 ml of liquid yeast extract (150 g/l), 200 ml heat-inactivated horse serum (56°C, 30 min) supplemented with 1 g ampicillin (Sigma, A5354) per litre.

Carcinoma lung epithelial (A549) cells were cultured in RPMI 1640 medium (Invitrogen) supplemented with 10% heat inactivated fetal bovine serum at 37°C with 5% CO_2 in tissue culture flasks.

Purified fibronectin (Code: 341635) and plasminogen (Code: 528175) from human plasma was supplied by Merck Millipore. Bovine actin (Code: A3653) and fetuin (Code: F3004) was supplied by Sigma.

4.5.2 Enrichment of *M. pneumoniae* surface proteins

4.5.2.1 Biotinylation

Biotinylation of the *M. pneumoniae* cell surface was carried out using a modification of a protocol described previously[186]. In brief, biotinylation using EZ-link sulfo-NHS-biotin (Thermo Fisher Scientific) was added to adherent *M. pneumoniae* cells in culture flasks and performed for 30 seconds on ice to minimise cell lysis after extensive washing of adherent cells with PBS to remove media components. The reaction was quenched with a final concentration 50 mM Tris-HCl and cells were lysed with 7 M urea, 2 M thiourea, 40 mM Tris-HCl (pH 8.8), 1% (w/v) C7bZ0. Biotinylated surface proteins were purified by avidin column chromatography and confirmed to be biotinylated by western blotting using Extravidin- HRP (Sigma).

4.5.2.2 Trypsin shaving

Trypsin shaving of *M. pneumoniae* cells was carried out as described previously[187]. In brief, adherent *M. pneumoniae* cells were extensively washed with PBS and then incubated with trypsin from porcine pancreas (Sigma, 50 $\mu g.ml^{-1}$) for 5 minutes at 37°C to release surface exposed peptides. Peptides were collected, digested a second time with trypsin Gold MS grade (Promega) and analysed by LC-MS/MS.

4.5.3 Preparation of *M. pneumoniae* whole cell lysates for one- and two-dimensional gel electrophoresis

M. pneumoniae whole cell lysates were prepared as previously described[187]. In brief, *M. pneumoniae* cells were extensively washed with PBS and lysed in 7 M urea, 2 M thiourea, 40 mM Tris-HCl, 1% (w/v) C7BzO detergent (Sigma-Aldrich), followed by three 30 s rounds of sonication at 60% power on ice. Proteins were reduced and alkylated with 5 mM tributylphosphine and 20 mM acrylamide monomers for 90 min at room temperature. Insoluble material was removed by centrifugation and five volumes of acetone added to precipitate protein. After centrifugation, the protein pellet was solubilized in 7 M urea, 2 M thiourea, 1% (w/v) C7BzO for one- and two-dimensional gel electrophoresis.

4.5.4 One- and two-dimensional polyacrylamide gel electrophoresis (PAGE)

Protein separation was performed as described in[221,224].

4.5.4.1 1D SDS-PAGE

80 µg of protein was separated on 4-20% Criterion™ TGX™ Gels (Bio-Rad) in Tris-Glycine-SDS buffer (Bio-Rad), fixed and visualized by staining with either Flamingo fluorescent gel stain (Bio-Rad) or Coomassie Blue G-250.

4.5.4.2 2D SDS-PAGE

250 µg of protein was cup-loaded onto 11 cm pH 4-7 IPG strips (BioRad) or 6-11 Immobiline drystrips (GE Healthcare) rehydrated with 7 M urea, 2 M thiourea, 1% (w/v) C7BzO. Focusing was performed in a Bio-Rad Protean IEF cell unit. Following IEF, the strips were equilibrated for 20 minutes with equilibration (2% SDS, 6 M urea, 250 mM Tris-HCl pH 8.5, 0.0025% (w/v) bromophenol blue) solution before running in the second-dimension SDS-PAGE.

4.5.4.3 Trypsin Digest

In-gel trypsin digestion was performed as described in[186]. In brief, gel pieces were destained, dehydrated and incubated with trypsin Gold MS grade (Promega) in 100 mM NH_4HCO_3 at 37°C overnight. Tryptic peptides were then analysed by LC-MS/MS. If necessary, gel pieces were reduced and alkylated with 5 mM TBP, 20 mM acrylamide in 100 mM NH_4HCO_3, destained and dehydrated a second time before the addition of trypsin.

4.5.5 Heparin affinity chromatography

Affinity purification of *M. pneumoniae* heparin binding proteins was performed as described in Raymond *et al.*, (2013)[224] with slight modification. In brief, *M. pneumoniae* cells were extensively washed with PBS and lysed in 10 mM sodium phosphate, 0.1% Triton TX-100, pH 7 with sonication. After centrifugation, ~300 µg of soluble protein was added 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 (GE healthcare) in binding buffer (10 mM sodium phosphate, pH 7). Non-binding proteins were removed with binding buffer. Heparin binding proteins were eluted with an increasing gradient of elution buffer (10 mM sodium phosphate, 2 M sodium chloride, pH 7). Fractionated proteins were separated by 1D-SDS PAGE, in-gel digested with trypsin and analysed by LC-MS/MS.

4.5.6 Avidin purification of host binding proteins

Avidin purification of *M. pneumoniae* host molecule binding proteins was performed as described in Raymond *et al.*, (2015)[226] with slight modifications. In brief, 1 mg of purified host protein was biotinylated and bound to avidin agarose. The avidin beads were incubated for 16 hours at 4°C with *M. pneumoniae* cells lysed with 1% (w/v) C7BzO (Sigma-Aldrich) in PBS (pH 7.8). Non-binding proteins were washed with PBS and host binding proteins were eluted with 7 M urea, 2 M thiourea, 40 mM Tris-HCl and 1% (w/v) C7BzO. Eluants were pooled, concentrated via acetone precipitation for separation by 1D-SDS PAGE, in-gel digested with trypsin and analysed by LC-MS/MS.

4.5.7 Avidin purification of A549 binding proteins

Avidin purification of *M. pneumoniae* proteins that bind A549 surface proteins was performed as described in Raymond *et al.*, (2013)[224] with modifications. In brief, A549 cells were grown to ~80% confluency and biotinylated in the flask after extensive washing with PBS. After quenching excess biotin with a final concentration of 50 mM Tris-HCl, A549 cells were lysed in 1% (w/v) C7BzO (Sigma-Aldrich) in PBS (pH 7.8) with sonication. A549 whole cell lysate was added to avidin agarose beads following centrifugation. The beads were washed four times (5 ml per wash) with PBS to remove non-biotinylated proteins. The beads were then incubated for 16 hours at 4°C with *M. pneumoniae* cells lysed with 1% (w/v) C7BzO (Sigma-Aldrich) in PBS (pH 7.8). Non-binding *M. pneumoniae* proteins were removed by washing four times again (5 ml per wash) with PBS and A549 binding proteins were eluted with 7 M urea, 2 M thiourea, 40 mM Tris-HCl and 1% (w/v) C7BzO (elution 1). Biotinylated surface A549 proteins that bound strongly to avidin-agarose were eluted with 30% acetonitrile and 0.4% trifluoroacetic acid (elution 2). Each of the eluents were concentrated via acetone precipitation for separation by 1D-SDS PAGE, in-gel digested with trypsin and analysed by LC-MS/MS.

4.5.8 Liquid chromatography tandem mass spectrometry (LC-MS/MS)

LC-MS/MS was performed as described in Raymond *et al.*, (2013)[224]. In brief, 15 µl of sample containing up to 5 µg of protein was loaded into an autosampler vial in an Eksigent AS-1 autosampler connected to a Tempo nanoLC system (Eksigent, USA) with a C8 Cap Trap column (Michrom Biosciences, CA). The peptides were washed onto a PicoFrit column (75 µm × 150 mm) packed with Magic C18AQ resin (Michrom Biosciences, CA). Peptides

were eluted from the column into the source of a QSTAR Elite hybrid quadrupole-time-of-flight mass spectrometer (Sciex). Eluted peptides were ionized from the PicoFrit at 2300 V. An Intelligent Data Acquisition (IDA) experiment was performed, with a mass range of 350–1500 Da continuously 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.

4.5.9 MS/MS data analysis

Mascot (Version 6.1) was used to search MS/MS data files as previously described Raymond *et al.*, (2013)[224] with modifications. In brief, files were searched against the MSPnr100 database[251] with the following parameters. Fixed modifications: none. Variable modifications: propionamide, oxidized methionine, deamidation. Enzyme: semi-trypsin. Number of allowed missed cleavages: 3. Peptide mass tolerance: 100 ppm. MS/MS mass tolerance: 0.2 Da. Charge state: 2+, 3+ and 4+. For samples collected from the 'Biotinylation enrichment of surface proteins' and 'Avidin purification of A549 interacting proteins', variable modifications also included NHS-LC-Biotin (K) and NHS-LC-Biotin (N-term). 'Avidin purification of A549 interacting proteins' was also searched against *homo sapiens* entries in MSPnr100 to identify biotinylated surface A549 proteins.

4.5.10 Dimethyl labelling of *M. pneumoniae* proteins

Protein labelling was performed on 1 mg of *M. pneumoniae* protein by the addition of 40 mM formaldehyde (ultrapure grade) (Polysciences, 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 of 4 hours. 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 -80°C for 3 hours. The precipitated protein was then pelleted by centrifugation at 14,000g and washed with 5 volumes of methanol. The protein pellet was resuspended in 50 mM sodium hydroxide, pH 8.0 and digested with trypsin for 16 hours at 37°C prior to clean up by SiliaPrepX™ HLB Polymeric SPE cartridges (Silicycle, Canada) and analysis by LC–MS/MS (Sciex 5600 and Thermo Scientific Q Exactive™).

4.5.11 Liquid chromatography tandem mass spectrometry (LC-MS/MS): Sciex 5600

Peptides from dimethyl labelled *M. pneumoniae* protein were separated by nanoLC using an Ultimate nanoRSLC UPLC and autosampler system (Dionex, Amsterdam, Netherlands). Samples (2.5 µl) were concentrated and desalted onto a micro C18 precolumn (300 µm x 5 mm, Dionex) with H_2O:CH_3CN (98:2, 0.1 % TFA) at 15 µl/min. After a 4 min wash the pre-column was switched (Valco 10 port UPLC valve, Valco, Houston, TX) into line with a fritless nano column (75µ x ~15cm) containing C18AQ media (1.9µ, 120 Å Dr Maisch, Ammerbuch-Entringen Germany). 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 200 nl/min over 240 min. High voltage 2000 V was applied to low volume Titanium union (Valco) with the tip positioned ~ 0.5 cm from the curtain plate (T=150°C) of a 5600^+ mass spectrometer (Sciex, Toronto, Canada). Positive ions were generated by electrospray and the 5600^+ operated in information dependent acquisition mode (IDA).

A survey scan m/z 350-1750 was acquired (PWHH resolution ~30,000, 0.25 sec acquisition time) with autocalibration enabled (at ~6 hr intervals). Up to the 10 most abundant ions (>300 counts) with charge states > +2 and <+5 were sequentially isolated (width m/z ~3) and fragmented by CID with an optimal CE chosen based on m/z (product ion spectra were acquired at a resolution ~20,000 PWHH in 0.15 sec). M/z ratios selected for MS/MS were dynamically excluded for 30 or 45 seconds.

Peak lists were generated using Mascot Daemon/Mascot Distiller (Matrix Science, London, England) or ProteinPilot (Sciex, v4.5) using default parameters, and submitted to the database search program Mascot (version 2.5.1, Matrix Science). Search parameters were: Precursor tolerance 10 ppm and product ion tolerances ± 0.05 Da; oxidation (M), deamidation (NQ), propionamide (C), Dimethyl (K), Dimethyl (N-term) specified as variable modifications; enzyme specificity was semi-ArgC; 1 missed cleavage was possible and the non-redundant protein database from NCBI (Jan 2015) searched.

4.5.12 Liquid chromatography tandem mass spectrometry (LC-MS/MS): Thermo Scientific Q Exactive™

Peptides from dimethyl labelled *M. pneumoniae* protein were separated by nanoLC using an Ultimate nanoRSLC UPLC and autosampler system (Dionex, Amsterdam, Netherlands).

Samples (2.5 µl) were concentrated and desalted onto a micro C18 precolumn (300 µm x 5 mm, Dionex) with H_2O:CH_3CN (98:2, 0.1 % TFA) at 15 µl/min. After a 4 min wash the pre-column was switched (Valco 10 port UPLC valve, Valco, Houston, TX) into line with a fritless nano column (75µ x ~35cm) containing C18AQ media (1.9µ, 120 Å Dr Maisch, Ammerbuch-Entringen Germany). 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 200 nl/min over 30 or 240 min. High voltage 2000 V was applied to low volume Titanium union (Valco) with the column oven heated to 45°C (Sonation, Biberach, Germany) and the tip positioned ~ 0.5 cm from the heated capillary (T=300°C) of a QExactive Plus mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Positive ions were generated by electrospray and the QExactive operated in data dependent acquisition mode (DDA).

A survey scan m/z 350-1750 was acquired (resolution = 70,000 at m/z 200, with an AGC target value of 10^6 ions) and lockmass was enabled (m/z 445.12003). Up to the 10 most abundant ions (>80,000 counts, underfill ratio 10%) with charge states > +2 and <+7 were sequentially isolated (width m/z 2.5) and fragmented by HCD (NCE = 30) with a AGC target of 10^5 ions (resolution = 17,500 at m/z 200). M/z ratios selected for MS/MS were dynamically excluded for 30 or 45 seconds.

Peak lists were generated using Mascot Daemon/Mascot Distiller (Matrix Science, London, England) or Proteome Discoverer (Thermo, v1.4) using default parameters, and submitted to the database search program Mascot (version 2.5.1, Matrix Science). Search parameters were: Precursor tolerance 4 ppm and product ion tolerances ± 0.05 Da; oxidation (M), deamidation (NQ), propionamide (C), Dimethyl (K), Dimethyl (N-term) specified as variable modifications; enzyme specificity was semi-ArgC; 1 missed cleavage was possible and the non-redundant protein database from NCBI (Jan 2015) searched.

4.5.13 Bioinformatic analysis of Mpn142

Bioinformatic analysis of Mpn142 used online resources: ProtParam[268], Clustal Omega[311], TMpred[258], COILS (Addition of 'yes' to 2.5 fold weighting of positions a,d)[312] and PONDR® (VSL2 and VL3 predictors)[313,314]. Using ScanProsite[315], the (X-[HRK]-[HRK]-X-[HRK]-X) motif identified by Cardin & Weintraub (1989)[276] was used to predict putative heparin binding motifs and the (X-[HRK]-X-[HRK]-[HRK]-X) motif is implicated in putative heparin sulphate binding sites[316].

4.6 Results

4.6.1 Defining P40 and P90 on the surface of *M. pneumoniae*

Surface accessible proteins labelled with biotin were recovered by streptavidin chromatography and separated by 2D-PAGE. All protein spots were cut from the gel, digested with trypsin and analysed by LC-MS/MS as described previously[221,224]. Tryptic peptides that mapped to Mpn142 were identified in two spot trains with masses of approximately 37 kDa and 84 kDa (data not shown). Twenty seven tryptic peptides (Mascot scores >50) that mapped to the 40 kDa protein spanned amino acids 26 – 308 of the P40 molecule (Figure 4.1). The first peptide ^{26}NTYLLQDHNTLTPYTPFTTPLNGGLDVVR^{54} was semi-tryptic in composition and represented the mature N-terminus of P40 after removal of the leader peptide (Table 4.1). This result is consistent with an earlier study of Catrein et al., (2005)[168]. A novel cleavage site with the sequence ^{364}NRT↓ASD^{371} defines the largest fragment we identified in the N-terminal half of Mpn142 and delineates the C-terminus of P40 (Fragment 3) consistent with reports of the size of P40[168]. Dimethyl labelling experiments identified semi-tryptic peptides that confirmed that cleavage occurs in multiple sites in this region of Mpn142 suggesting that it is readily accessible to proteases or is further processed after the initial cleavage to create P40 (Table 4.1). Fragment 3 spans amino acids 26 – 368 with a predicted mass of 36.3 kDa (pI = 9.28) and was identified in affinity columns loaded with heparin and fibronectin at masses consistent with this predicted size of the molecule. A putative heparin binding domain with the sequence ^{151}ERKIKL^{156} was identified within the P40 sequence consistent with our ability to recover Fragment 3 from heparin-agarose. Further studies are needed to confirm interactions between Fragment 3 with heparin and fibronectin.

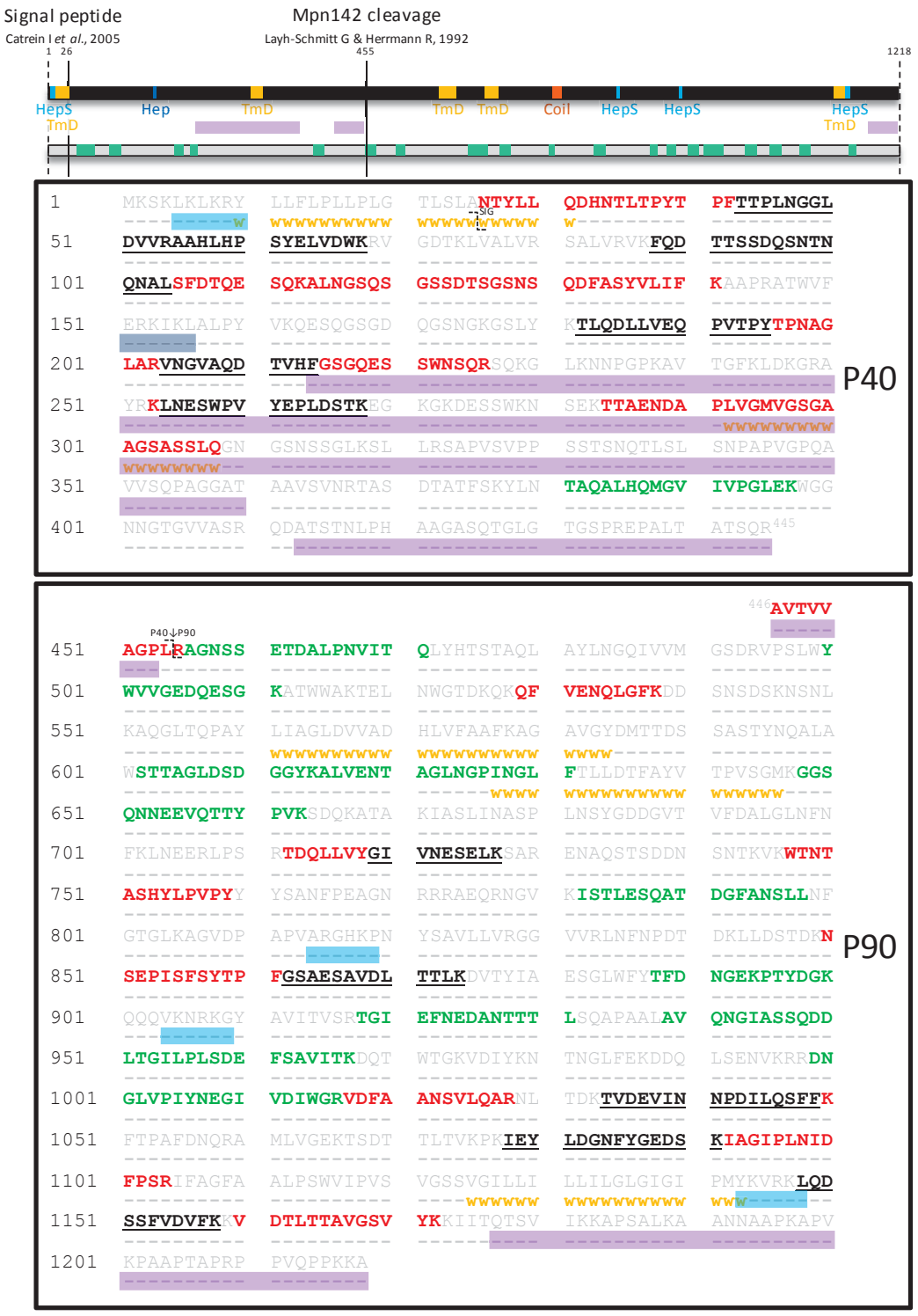


Figure 4.1: Peptides identified in surface proteome analysis of Mpn142. Full length Mpn142 is represented as a black bar and its amino acid sequence (grey) is shown underneath. Tryptic peptides identified by shaving the surface of

M. pneumoniae **are coloured green while tryptic peptides derived from a 2D gel loaded with biotinylated (surface) proteins of** ***M. pneumoniae*** **are coloured red. Tryptic peptides that were common to both surface analyses are coloured black and underlined. Bioinformatics tools were used to predict coiled-coils (orange box; COILS), transmembrane domains (yellow squares/wwwww; TMpred), disordered regions (purple boxes; PONDR®: VSL2 and VSL3 predictor), heparin binding motifs (dark blue box; ScanProsite with "*X-[HRK]-[HRK]-X-[HRK]-X*" motif) and implicated in heparan sulfate binding regions (light blue boxes; ScanProsite with "*X-[HRK]-X-[HRK]-[HRK]-X*" motif). The two previously reported cleavage sites after amino acid 25 and 454 are indicated as unbroken lines across the black bar and with the symbol ↴ in the sequence.**

Table 4.1: N-terminal dimethyl labelled and semi-tryptic peptides identified in Mpn142 (Q50341).

#	Peptide Sequence	Score	E-value	5600	QE
	N-terminal Dimethyl Labelled				
N1	-.^{1}MKSKLKLKR^{9}.Y	11*	0.72	1	1
N2	L.^{25}ANTYLLQDHNTLTPYTPFTTPLNGGLDV VR^{54}.A	36	0.0024	1	3
N3	A.^{26}NTYLLQDHNTLTPYTPFTTPLNGGLDVV R^{54}.A	142	3.5e-13	4	4
N4	R.^{153}KIKLALPYVKQESQGSGDQGSNGKGSLY KTLQDLLVEQPVTPYTPNAGLARVNGV^{207}.A	66	2.3e-5	-	1
N5	R.^{368}TASDTATFSK^{377}.Y	50	0.00068	-	1
N6	R.^{368}TASDTATFSKYLNTAQALHQMGVIVPGL EKWGGNNGTGVVAS^{409}.R	55	1.2e-5	2	-
N7	R.^{368}TASDTATFSKYLNTAQALHQMGVIVPGL EKWGGNNGTGVVASR^{410}.Q	146	1.1e-13	2	3
N8	R.^{368}TASDTATFSKYLNTAQALHQMGVIVPGL EKWGGNNGTGVVASRQ^{411}.D	148	5.2e-14	1	4
N9	R.^{368}TASDTATFSKYLNTAQALHQMGVIVPGL EKWGGNNGTGVVASRQD^{412}.A	163	1.2e-15	-	2
N10	R.^{368}TASDTATFSKYLNTAQALHQMGVIVPGL EKWGGNNGTGVVASRQDA^{413}.T	80	9.8e-8	-	2
N11	R.^{368}TASDTATFSKYLNTAQALHQMGVIVPGL EKWGGNNGTGVVASRQDAT^{414}.S	62	3.9e-6	-	2

N12	T.[369]ASDTATFSKYLNTAQALHQMGVIVPGLE KWGGNNGTGVVASR[410].Q	138	2.7e-13	2	4
N13	R.[446]AVTVVAGPLR[455].A	72	1.2e-5	3	4
N14	A.[447]VTVVAGPLR[455].A	48	0.0024	3	4
N15	L.[696]GLNFNFKLNEER[707].L	13*	1.3	2	2
N16	R.[731]ENAQSTSDDNSNTKVKWTNTASHYLPVP YYYSANFPEAGNRR[772].R	21*	0.015	-	1
N17	A.[775]EQRNGVKISTLESQATDGFANSLLNFGT GLKAGVDPAPVAR[815].G	41	0.00025	-	1
N18	A.[1022]NSVLQARNLTDKTVDEVINNPDILQSF FKFTPAFDNQR[1059].A	84	8.8e-7	3	1
N19	K.[1190]AANNAAPKAPVKPAAPTAPRPPVQPPK KA[1218].-	59	4.7e-6	-	1
N20	A.[1191]ANNAAPKAPVKPAAPTAPRPPVQPPKK A[1218].-	99	1.7e-9	1	3
N21	A.[1192]NNAAPKAPVKPAAPTAPRPPVQPPKKA [1218].-	41	0.00014	-	1
N22	N.[1194]AAPKAPVKPAAPTAPRPPVQPPKKA[121 8].-	44	0.00019	-	2
N23	A.[1195]APKAPVKPAAPTAPRPPVQPPKKA[1218].-	18*	0.051	-	1
N24	P.[1197]KAPVKPAAPTAPRPPVQPPKKA[1218].-	21*	0.029	-	1
	Semi-tryptic Truncated C-terminal Peptides				
S1	R.[253]KLNESWPVYEPLDSTKEGKGKDESSWKN SEKTTAENDAPLVGMVGSGAA[301].G	73	5.4e-7	1	-
S2	R.[253]KLNESWPVYEPLDSTKEGKGKDESSWKN SEKTTAENDAPLVGMVGSGAAGS[303].A	87	1.4e-8	1	-
S3	R.[253]KLNESWPVYEPLDSTKEGKGKDESSWKN SEKTTAENDAPLVGMVGSGAAGSA[304].S	30*	0.0071	1	-
S4	R.[253]KLNESWPVYEPLDSTKEGKGKDESSWKN SEKTTAENDAPLVGMVGSGAAGSAS[305].S	40	0.00072	1	-
S5	R.[253]KLNESWPVYEPLDSTKEGKGKDESSWKN SEKTTAENDAPLVGMVGSGAAGSASS[306].L	39	0.00077	-	1
S6	R.[253]KLNESWPVYEPLDSTKEGKGKDESSWKN SEKTTAENDAPLVGMVGSGAAGSASSLQ[308].G	86	1.4e-7	1	-
S7	R.[1029]NLTDKTVDEVINNPDILQSFFKFTPAF DNQRAMLV[1063].G	45	0.0023	-	1
	Semi-tryptic Truncated N-terminal Peptides				

S8	D.^{810}PAPVARGHKPNYSAVLLVR^{828}.G	65	2.3e-6	-	1
S9	N.^{983}GLFEKDDQLSENVKRR^{998}.D	26*	0.033	-	1
S10	N.^{1193}NAAPKAPVKPAAPTAPRPPVQPPKKA^{1218}.-	25*	0.015	-	1
	P1 adhesin Cleaved Peptides				
C1	R.^{1594}LKQTSAAKPG^{1603}.A (Semi-tryptic)	24*	0.045	-	1
C2	T.^{1598}SAAKPGAPRPPVPPKPGAPKPPVQPPKKPA^{1627}.- (Dimethyl labelled)	58	4.5e-6	-	2

All peptides have a score > 32 and an E-value < 0.05 unless indicated by * which signifies that the peptide was either identified over several replicates or correlates with predicted fragments in this study. Cleavage sites are located beside the bold underlined amino acid (left for N-terminus and right for C-terminus). Amino acid number is written as superscript at the start and end of the peptide. Highest ion score and lowest E-value for the peptide identified across the replicates is listed. The last two columns contain the number of times the peptide was identified by either the Sciex 5600 TripleTOF or Thermo Q Exactive Plus out of a total of six biological replicates. The last two peptides are C-terminal cleavage sites identified for P1 (Uniprot #: P11311).

Twelve tryptic peptides (Mascot scores > 50) spanning amino acids 446 – 1172 were mapped to P90 (Fragment 2). Notably, the first peptide ^{446}AVTVVAGPLR^{455} is semi-tryptic in composition and commenced nine amino acids upstream of the previously determined start site for P90[166,167] (Figure 4.1). The most C-terminal peptide in P90 has the sequence ^{1160}VDTLTTAVGSVYK^{1172}. Tryptic peptides that mapped to a C-terminal 90 kDa fragment of Mpn142 were isolated by affinity chromatography using fetuin, fibronectin, actin and A549 surface proteins as bait indicating that P90 may bind to fibronectin, actin, fetuin and other undefined receptors on the surface of A549 cells (Figure 4.2).

Mpn142 ORF: 1,218 amino acids; 130.46 kDa, pI 8.01

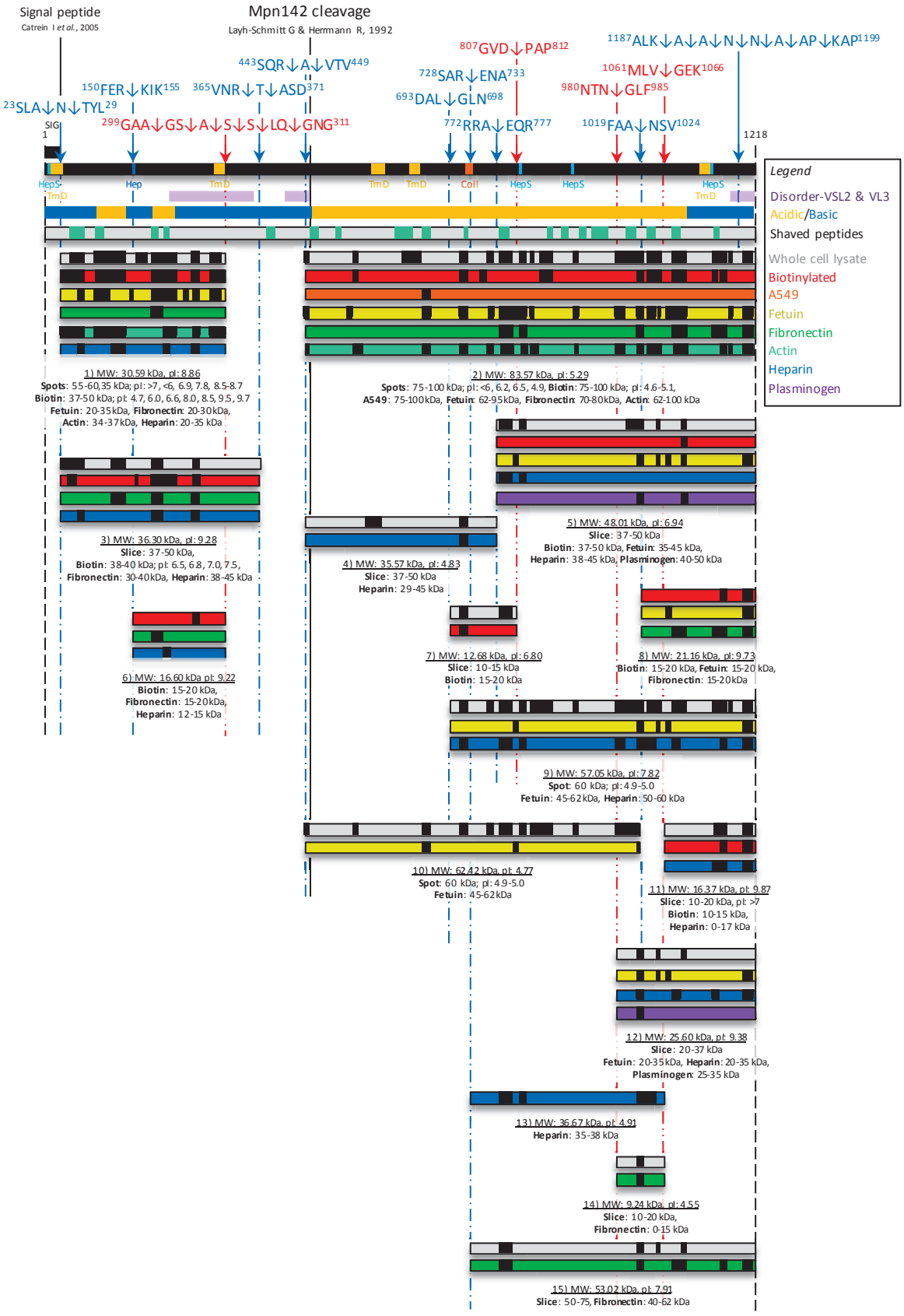


Figure 4.2: Cleavage map of Mpn142. Full length Mpn142 is represented as a black bar and cleavage products shown underneath. Cleavage sites were determined by mapping neo-N-termini generated by N-terminal dimethyl labelling (blue broken lines) and identifying semi-tryptic peptides (red broken lines). Cleavage sites are indicated by the arrows and peptide sequences shown above the bar. Putative glycosaminoglycan binding sites (HepS/Hep), transmembrane domains (TmD), coiled-coils (Coil) and disordered regions identified using ScanProsite, TMpred, COILS and PONDR®, respectively are depicted. Regions within Mpn142 that are enriched in acidic amino acids D/E (yellow) and basic amino acids K/R/H (blue) are shown beneath the black bar representing Mpn142. Shaved peptides are depicted by the green boxes within the grey bar. Peptides spanning fragments of Mpn142 obtained from 1D/2D SDS PAGE of whole cell lysates (grey bars); recovery of biotinyled proteins (red bars); or A549 (orange bar), fetuin (yellow bars), fibronectin (green bars), actin (teal bars), heparin (dark blue bars) or plasminogen (purple bars) based affinity chromatography were identified by LC-MS/MS (small black boxes). Beneath each fragment is the assigned fragment number, predicted theoretical mass and pI predicted by ProtParam (underlined). Most fragments were identified in a gel mass range which is listed. The signal peptide and previously determined Mpn142 cleavage sites are presented as unbroken black lines.

Tryptic peptides spanning Mpn142 that were released during mild trypsin hydrolysis of *M. pneumoniae* cells were identified by LC-MS/MS. Twenty one peptides spanning Mpn142 including six peptides between amino acids 43 – 397 in P40 and fifteen peptides spanning amino acids 456 – 1158 in P90 were identified (see green boxes in the grey bar in Figure 4.2). The peptide ^{381}TAQALHQMGVIVPGLEK^{397} is flanked on either side by predicted disordered regions and is the only peptide that mapped to this region suggesting that it is a preferential site for proteolysis in Mpn142. These data confirm that most of Mpn142 is exposed on the surface of *M. pneumoniae*.

4.6.2 Bioinformatic analysis of Mpn142

The online PONDR® tool (VSL2 and VL3 predictors)[313,314] identified three major disordered regions spanning more than 40 consecutive amino acids from amino acid 214 – 360 and 413 – 453 which aligns with C-terminal half of P40 (Figure 4.1). The third disordered region resides in the C-terminal region of P90 between amino acids 1178 – 1218 and spanned an unusual sequence of amino acids enriched in alanine, proline and lysine residues located at the C-terminus of the molecule (Figure 4.3). Interestingly, a similar run of amino acids was identified in the C-terminus of the P1 adhesin derived from the *mpn141* gene found in the same polycistronic unit (Figure 4.3). Notably, the disordered region spanning amino acids 413 – 453 encompasses the two consecutive cleavage sites ^{443}SQR↓A↓VTV^{449} that separates P40 from P90 while the longest disordered region from 214 – 360 encompasses cleavage sites ^{299}GAA↓GS↓A↓S↓S↓LQ↓^{308} that defines the C-terminus of P40. Evidence to support these cleavage events is derived from the identification of dimethyl labelled semi-tryptic peptides (neo-N-termini) by LC-MS/MS (see Table 4.1). We rarely found tryptic peptides that mapped to amino acids 310 – 440. Processing sites in adhesins often reside within disordered regions presumably because they are accessible to proteases[186,221,222,224,225]. We also identified a putative heparin binding site with the sequence ^{151}ERKIKL^{156} and a series of regions which may be implicated in heparan sulfate binding domains with the sequences: ^{151}ERKIKL^{156}, ^{5}LKLKRY^{10}, ^{814}ARGHKP^{819}, ^{904}VKNRKG^{909} and ^{1143}YKVRKL^{1148} in Mpn142 prompting us to use heparin as bait in affinity chromatography experiments to enrich for cleavage fragments derived from Mpn142.

```
Mpn142      1113 PSWVIPVSVGSSVGILLILLILGLGIGIPMYKVRKLQDSSFVDVFKKVDTLTTAVGSVYK 1172
P1 adhesin  1520 PDYVLPLAITVPIVVIVLSVTLGLAIGIPMHKNKQALKAGFALSNQKVDVLTKAVGSVFK 1579
                 *.:*:*:::   : :::: : ***.*****:* ::  .:.*.   :***.**.*****:*

Mpn142      1173 KIITQTSVIKKAPSALKAANNAAPKAP--VKPAAPTAPRPPVQPPKKA- 1218
P1 adhesin  1580 EIINRTGIS-QAPKRLKQTSAAKPGAPRPPVPPKPGAPKPPVQPPKKPA 1627
                 :**.:*.:  :**. ** :. * * **    *  * **:********
```

Figure 4.3: Alignment of the C-terminal sequence of Mpn142 and P1. A Clustal Omega alignment of the amino acid sequence spanning 1113 – 1218 of Mpn142 against the sequence spanning 1520 – 1627 of P1. Predicted transmembrane domains are highlighted in grey (TMpred score 2,518) for Mpn142 and predicted by Nakane et al. (2010) using the SMART algorithm[301]. Cleavage sites in the sequence are denoted by the symbol ↿. * indicate conserved residues, : indicate similar amino acids and . indicate weakly similar amino acids.

4.6.3 Cleavage fragments in the N-terminus of Mpn142

Dimethyl labelling of neo-N-termini provided evidence that Mpn142 is subject to considerable posttranslational processing. The list of dimethyl labelled peptides identified by LC-MS/MS is shown in Table 4.1. To find evidence that these cleavage events generated functionally important cleavage fragments of Mpn142, we characterised tryptic peptides derived from size-fractionated whole cell lysates of *M. pneumoniae* strain M129 separated by SDS-PAGE and proteins captured by affinity chromatography loaded with different bait including actin, fetuin, plasminogen, heparin, fibronectin and surface proteins from A549 cells (Figure 4.2). In addition, we identified fragments of Mpn142 from 2D SDS-PAGE gels loaded separately with M129 whole cell lysates, and biotinylated surface proteins, and characterised them by LC-MS/MS. Using these combined approaches we identified 15 fragments of Mpn142 (Figure 4.2). Consistent with data described above, we were unable to find a cleavage fragment spanning amino acid 26 – 454 with a predicted mass of 44.9 kDa which represents the largest possible N-terminal cleavage fragment of Mpn142. Fragments 1, 3, and 6 span different regions of the N-terminus of Mpn142. Fragment 1 has a predicted mass of 30.6 kDa (pI = 8.86), spans amino acids 26 – 308 and was identified in a series of gel spots at 35 kDa (pI ranging from 6.9 – 8.7) from 2D gels loaded with whole cell lysates, and biotinylated surface proteins of *M. pneumoniae*, and from affinity columns loaded with fetuin, actin, heparin and fibronectin. These data suggest that Fragment 1 binds actin, fetuin, heparin and fibronectin and is present on the surface

of *M. pneumoniae*. Fragment 6 spans amino acid 153 – 308 with a predicted mass of 16.6 kDa and was also recovered from gel slices in the range of 15 – 20 kDa containing biotinylated surface proteins, *M. pneumoniae* proteins eluted proteins during affinity chromatography using fibronectin as bait, and a SDS-polyacrylamide gel slice (12 – 15 kDa) containing proteins that eluted from a heparin-agarose column. These data suggest that Fragment 6 binds heparin and fibronectin and is present on the surface of *M. pneumoniae*.

4.6.4 Cleavage fragments residing in the C-terminus of Mpn142

Several dimethylated N-terminal peptides were identified in the C-terminal region of Mpn142 (Table 4.1). Four cleavage sites reside between amino acids 695 – 810 and three cleavage sites were identified between amino acids 982 – 1063. In total, we identified 12 fragments spanning different regions of the C-terminal two thirds of Mpn142.

Fragment 5 is predicted to start at position 775 (see dimethyl peptide N17 in Table 4.1) and terminates at amino acid 1,218, generating a protein with a predicted mass of 48 kDa. Fragment 5 was recovered from affinity columns loaded with plasminogen (2 peptides; gel slice 40 – 50 kDa), fetuin (6 peptides; gel slice from 35 – 45 kDa), heparin (two peptides; gel slice 38 – 45 kDa) and from a gel slice (37 – 50 kDa) containing size-fractionated *M. pneumoniae* proteins (6 peptides). Peptide coverage from each affinity chromatography experiment is depicted in Figure 4.2.

Fragment 4 (limited evidence for its existence in this study) is predicted to span amino acids 446 – 774 and has a predicted mass of 35.6 kDa. It was identified (2 tryptic peptides) from a gel slice (37 – 50 kDa) containing size-fractionated *M. pneumoniae* proteins and from size fractionated proteins (29 – 45 kDa) eluted during heparin agarose chromatography.

Fragment 7 spans amino acids 696 – 809, has a predicted mass of 12.7 kDa and was recovered from a gel slice (10 – 15 kDa) containing size-fractionated *M. pneumoniae* proteins (2 tryptic peptides). It is predicted to be derived from a cleavage event at position 696 (^{693}DAL↓GLN^{698}) and ends at a cleavage event at position 809 in the sequence ^{807}GVD↓PAP^{812} (see semi-tryptic peptide S8 and dimethyl peptide N15 in Table 4.1). We

also identified a similar fragment from captured biotinylated surface proteins in a gel slice spanning 15 – 20 kDa (Figure 4.2).

Fragment 8 is delineated by a cleavage event at position 1021 within the sequence ^{1019}FAA↓NSV^{1024} (see dimethyl peptide N18 in Table 4.1) at the N-terminus and ends at the C-terminus of Mpn142 (position 1,218) (Figure 4.2). It has a predicted mass of 21 kDa and was recovered in three separate experiments in gel slices spanning 15 – 20 kDa loaded with biotinylated surface proteins of *M. pneumoniae* (2 peptides) and affinity columns loaded with fetuin (2 peptides) and fibronectin (6 peptides) (Figure 4.2).

Fragments 9 and 10 span different but overlapping regions of P90. Notably, peptide coverage from a protein spot of 60 kDa spanned all of P90 indicating that there are two overlapping fragments co-migrating in the one spot. We predict that Fragment 9 commences at position 696 at the cleavage site ^{693}DAL↓GLN^{698} (see dimethyl peptide N15 in Table 4.1) and spans the C-terminus of Mpn142 generating a theoretical fragment with a mass of 57.1 kDa (p*I* =7.82). Further support was provided by the identification of nine tryptic peptides mapping to a fragment recovered from heparin-agarose in a gel slice spanning 50 – 60 kDa. Six tryptic peptides mapping to the C-terminal 90 kDa region of Mpn142 were also identified by LC-MS/MS in a protein recovered from an affinity column loaded with fetuin as bait, in a gel slice with masses ranging from 45 – 62 kDa (Figure 4.2). As such our data is consistent with Fragment 10 commencing at the true N-terminus of P90 (see dimethyl peptide N13 in Table 4.1) at cleavage site ^{443}SQR↓AVTV^{450} and spanning the central region of Mpn142, ending at position 1,021 at the cleavage site ^{1019}FAA↓NSV^{1024}. As such, Fragment 10 has a theoretical mass of 62.4 kDa (pI=4.77).

Fragment 11 was recovered from SDS-PAGE fractionated *M. pneumoniae* proteins in a gel slice spanning 10 – 20 kDa. Our data is consistent with Fragment 11 commencing at position 1,064 at cleavage site ^{1061}MLV↓GEK^{1066} (see semi-tryptic peptide S7 in Table 4.1) and ending at position 1218. Fragments of a similar size were recovered from heparin agarose in a gel slice containing putative heparin binding proteins up to 17 kDa in size and from a streptavidin column loaded with biotinylated surface proteins of *M. pneumoniae* with masses between 10 – 15 kDa (Figure 4.2). Fragment 11 is predicted to be 16.4 kDa.

Fragment 12 comprises tryptic peptides spanning the C-terminal 25.6 kDa of Mpn142 and resides in gel slices from 20 – 37 kDa in size fractionated *M. pneumoniae* proteins recovered from affinity chromatography experiments using fetuin, heparin and plasminogen as bait. Peptide coverage was consistent with Fragment 12 commencing at position 982 in the sequence ^{980}NTN↓GLF^{985} and ending at position 1,218 of Mpn142, generating a protein with a predicted mass of 25.6 kDa.

Fragment 13 was recovered from a heparin-agarose column and is defined by 4 tryptic peptides residing in a region of Mpn142 spanning amino acids 774 – 1063, consistent with cleavage events commencing at position ^{772}RRA↓EQR^{777} and ending at position ^{1061}MLV↓GEK^{1066} (see dimethyl peptide N17 and semi-tryptic peptide S7 in Table 4.1). Fragment 13 has a predicted mass of 36.7 kDa and encompasses a region of Mpn142 that displays two putative heparan sulfate binding domains with the X-[HRK]-X-[HRK]-[HRK]-X motif (^{814}ARGHKP^{819} and ^{904}VKNRKG^{909}).

Evidence for the existence of Fragment 14 is denoted by the identification of 2 tryptic peptides that map to a small fragment of Mpn142 recovered from gel slices containing *M. pneumoniae* proteins from 10 – 20 kDa and from proteins eluted from an affinity column where fibronectin has been used as bait. Our data is consistent with Fragment 14 commencing at position 982 (^{980}NTN↓GLF^{985}) and ending at position 1063 (^{1061}MLV↓GEK^{1066}) generating a fragment with a predicted mass of 9.2 kDa.

Fragment 15 is predicted to spans amino acids 731 – 1218 (cleavage site ^{728}SAR↓ENA^{733} and dimethyl peptide N16 in Table 4.1) and has a predicted mass of 53 kDa. We identified 5 tryptic peptides spanning this fragment in a gel slice containing proteins 50 – 75 kDa and in size fractioned (40 – 62 kDa) *M. pneumoniae* proteins (8 peptides) eluted from an affinity column loaded with fibronectin (Figure 4.2).

Notably, we identified a series of cleavage events in the C-terminal predicted disorder region from position 1,190 that removed one amino acid at a time from the N-terminus of the peptide with sequence ^{1190}AANNAAPKAPVKPAAPTAPRPPVQPPKKA^{1218} (Table 4.1: dimethyl peptides N18 – N24 and semi-tryptic peptide S10). These events release a C-terminal peptide from Mpn142 that is enriched in alanine/valine (11 residues),

lysine/arginine (5 residues) and proline (9 residues) residues and is highly similar in sequence to the C-terminus of the P1 adhesin (Mpn141) (Figure 4.3).

4.7 Discussion

A growing body of evidence exists to suggest that molecules that reside on the cell surface of mycoplasmal pathogens are processed into discrete functional domains via a process known as ectodomain shedding[186,187,216-226,231,232]. While the majority of this evidence comes from recent studies of adhesin molecules in *M. hyopneumoniae* there is ample evidence that surface-accessible proteins in *M. pneumoniae* including Mpn142 and several uncharacterized lipoproteins (Mpn052, Mpn284, Mpn288, Mpn376, Mpn400, Mpn408, Mpn444, Mpn456, Mpn474 and Mpn491) are processed at multiple sites[248]. However, in most instances, precise cleavage events have not been mapped. Previous studies have shown that *M. pneumoniae* has an affinity for a wide variety of host receptors including fibronectin[205]; fibrinogen[208]; plasminogen[214,215]; sialylated receptors and oligosaccharides[27,144,145]; sulfated glycolipids[310]; laminin, fetuin, and human chorionic gonadotropin[146] but the identity of adhesins that bind them have not been characterized in detail. To characterise the processing events and to determine if the products of cleavage are potential adhesins, we developed systems-wide, affinity chromatography methodologies to recover proteins that interact with important host cell surface molecules such as heparin, fibronectin, actin, plasminogen, fetuin and proteins on the surface of A549 cells and identified them by LC-MS/MS. While this approach suggests a direct interaction between *M. pneumoniae* proteins and the bait, definitive proof is lacking because a subset of the captured proteins may interact with proteins that bind directly to the bait. In *M. pneumoniae* pyruvate dehydrogenase β (PdhB) and elongation factor Tu (Ef-Tu) bind fibronectin[205], glyceraldehyde-3-phosphate dehydrogenase binds fibrinogen[208] and PdhA, PdhB and PdhC subunits bind plasminogen[214,215]on the surface. However, it is not clear how adhesins that localise to the attachment organelle bind host molecules. The data presented here suggests that cleavage fragments of Mpn142 function as adhesins that bind a wide range of host molecules.

Mpn142 comprises 1,218 amino acids and is cleaved generating 40 (P40) and 90 kDa (P90) fragments on the surface of *M. pneumoniae*[166,167,172,317]. Mutants that cannot synthesize Mpn142 are unable to localise the major adhesin, P1 to the tip of the

attachment organelle and lose the ability to adhere to surfaces[152,303]. In addition, P90 forms a 480 kDa protein complex in a 1:2 molar ratio with P1 and forms an appendage that allows *M. pneumoniae* to glide across surfaces[301]. These observations suggest that P90 and P40 function as adhesins either directly via interactions with receptors on cell and abiotic surfaces or via collaborative interactions with P1. We show here that tryptic peptides spanning 15 fragments of Mpn142 were identified by LC-MS/MS in size fractionated lysates of *M. pneumoniae* (Figure 4.2). Previous studies show that a cleavage event at position 455 in Mpn142 was seminal to the creation of the dominant cleavage fragments P90 and P40. Edman degradation of the N-terminus of P90 generated the sequence [455]RAGNSSETDAL[465][167]. Another cleavage event at position 26 removes the leader sequence in the N-terminus[168]. As such, P40 was thought to span amino acids 26 – 454 (theoretical mass of 44.9 kDa) and P90, amino acids 455 – 1218 (theoretical mass of 82.8 kDa). However, during SDS-PAGE, P40 migrates with a mass of 35 – 40 kDa[166,167] prompting speculation that further cleavage events occur in the proposed P40 sequence[80]. In our study, the largest fragment spanning the N-terminal region of Mpn142 spans amino acids 26 – 368. A dimethyl labelled peptide [369]ASDTATFSKYLNTAQALHQMGVIVPGLEKWGGNNGTGVVASR[410] (peptide N12; Table 4.1) commencing at amino acid 369 indicated that P40 spans amino acids 26 – 368 (theoretical mass 36.2 kDa), a size consistent with earlier studies of P40[168]. In support of this hypothesis, we rarely identified tryptic peptides that mapped in the disordered region spanning amino acids 369 – 444 suggesting this region is readily accessible to different proteases. Notably, we identified a series of six semi-tryptic peptides spanning amino acids 253 – 308 and eight semi-tryptic peptides spanning amino acids 368 – 414 that differed by the sequential loss of a C-terminal amino acid (Table 4.1; peptides S1 – S6 and peptides N5 – N12, respectively) indicating that *M. pneumoniae* displays carboxypeptidase activity on the cell surface. The N-terminal peptide consistently identified in P90 ([446]AVTVVAGPLR[455]) started nine amino acids upstream of the predicted start site defined by the sequence [455]RAGNSSETDAL[465] (Table 4.1: peptide N13). A second semi-tryptic peptide [447]VTVVAGPLR[455] was identified (Table 4.1: peptide N14) in this region suggesting aminopeptidase(s) that target hydrophobic amino acid residues are active on the surface of *M. pneumoniae*. This hypothesis was supported by the identification of a series of six dimethyl labelled peptides and one semi-tryptic peptide from 1190 – 1218 each one

differing by the loss of a single N-terminal amino acid (Table 4.1; peptides N19 – N24 and peptide S10). Collectively, our data suggests that P90 commences at position 446 and that peptidase activity can alter the N-terminus generating size variants of P90. We characterised N-terminal dimethyl labelled peptides by LC-MS/MS as a method to define precise cleavage events in Mpn142. Using this approach we identified 17 peptides each indicating the start of a distinct proteoform derived from Mpn142 (Table 4.1, peptides N2 – N5, N12 – N24). The location of these peptides is consistent with processing sites in Mpn142 shown in Figure 4.2 and tryptic peptides that map to all 15 fragments of Mpn142 recovered by affinity chromatography. Other cleavage events were mapped by characterising truncated C-terminal and other semi-tryptic peptides (Table 4.1).

Notably, 21 peptides spanning amino acids 43 – 1158 of Mpn142 were identified when freshly cultured whole cells of *M. pneumoniae* were exposed to mild proteolysis with trypsin. While these data provided further evidence that Mpn142 is on the surface of *M. pneumoniae* we noted that 16 of the 21 peptides were not tryptic. Further analysis showed that eight peptides were semi-tryptic at the N-terminus, eight peptides were semi-tryptic at the C-terminus and five peptides were tryptic. This data is consistent with the presence of peptidase activity on the cell surface. Consistent with these data, we identified a range of proteases (data not shown) on the surface of *M. pneumoniae* in surfaceome studies (our unpublished data). Interestingly, the peptide ^{381}TAQALHQMGVIVPGLEK^{397} is the only peptide we identified in our study to reside within the disordered region spanning amino acids 214 – 453. The identification of six dimethyl labelled peptides and one semi-tryptic peptide (Table 4.1; peptides N19 – N24 and peptide S10) suggests that the C-terminus of Mpn142 spanning amino acids 1190 – 1218 is released into the extracellular milieu. Further work is needed to determine if the peptide ^{1190}AANNAAPKAPVKPAAPTAPRPPVQPPKKA^{1218} and derivatives of it have adhesive or immune-modulatory functions.

4.8 Acknowledgments

M. Widjaja is a recipient of the 'Australian Postgraduate Award' scholarship from University of Technology Sydney. I.J. Berry is a recipient of the Doctoral scholarship from University of Technology Sydney. The authors would like to thank Mark Raftery and the Bioanalytical Mass Spectrometry Facility (BMSF) for access to the Sciex 5600 and Thermo Scientific Q Exactive™ Plus instruments purchased with monies from ARC grant LE130100096 entitled 'Advanced high resolution mass spectrometer for collaborative proteomic and lipidomics research'. The authors would also like to thank Paul Baba for assisting in designing the cleavage map for figures.

4.9 Author Contributions

MW performed all the experiments except the dimethyl labelling experiments; MW interpreted the data, prepared all the Figures and Tables and assisted with writing drafts of the manuscript; IB, EP and MP performed the dimethyl labelling studies and critically evaluated their validity; IB also analysed the data for dimethyl labelling studies and assisted with compiling data for Table 4.1; MP quality controlled the mass spectrometry data generated in this study; SPD initiated and funded the study, interpreted the data and wrote the manuscript. All authors read and edited the final manuscript.

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Chapter 5.
Elongation factor Tu is a multifunctional and processed moonlighting protein

i3 ithreeinstitute

5.1 Preface

Alongside Pdh-B, Ef-Tu was one of the first moonlighting proteins identified in *M. pneumoniae* that was reported to bind fibronectin on the cell surface. Pdh-B has been characterised further and was observed to bind serval more different host proteins, but the binding targets for Ef-Tu remain uncharacterised. Because of this, Ef-Tu was investigated further to search for potential binding targets using microscale thermophoresis and ELISA. The manuscript presented as Chapter 5 demonstrates that Ef-Tu displays an affinity to a wide range of host molecules suggesting a key moonlighting role in adherence. Post-translational proteolysis was also observed in *M. pneumoniae* Ef-Tu and two other pathogens. This manuscript reinforces the multifunctionality of *M. pneumoniae* moonlighting proteins.

The manuscript presented for Chapter 5 has been peer-reviewed and was accepted for publication in 'Scientific Reports' on the 10th August, 2017. It should be noted that text and figures have been rearranged to focus on *M. pneumoniae*. Methods, figures, and tables pertaining to *M. hyopneumoniae* and *S. aureus* have been moved to 'Appendix 5: 8.5.3 Supplementary Methodology'.

5.2 Author contribution

Author	Contribution
Michael Widjaja	Performed surface proteome experiments, the 'Bait and Prey' chromatography, expressed and purified recombinant Ef-Tu, and performed the microscale thermophoresis experiments. Analysed all the data generated for *M. pneumoniae* except those generated by Lisa Hagemann and Anna Gründel, and performed 3D modelling predictions for all three pathogens and compiled Ef-Tu data for *M. pneumoniae.*
Kate L. Harvey	Performed surface proteome experiments and the 'Bait and Prey' chromatography for *S. aureus*. Analysed all data for *S. aureus.* Compiled Ef-Tu data for *S. aureus.*

Lisa Hagemann	Performed ELISA binding experiments for A549 cells and several host proteins, and performed plasminogen binding and activation studies.
Iain J. Berry	Performed dimethyl labelling for *M. pneumoniae* and *M. hyopneumoniae*. Analysed *M. hyopneumoniae* labelling data. Compiled Ef-Tu data for *M. hyopneumoniae*.
Veronica M. Jarocki	Performed bioinformatics analyses specifically SLiM, identified putative heparin binding motifs, disordered regions, and conserved regions of Ef-Tu using *in silico* methods.
Benjamin B.A. Raymond	Performed surface proteome experiments and the 'Bait and Prey' chromatography for *M. hyopneumoniae.*
Jessica L. Tacchi	Performed surface proteome studies for *M. hyopneumoniae.*
Anna Gründel	Produced recombinant Ef-Tu and generated guinea pig antiserum.
Joel R. Steele	Performed dimethyl labelling for *S. aureus* and helped analyse labelling data.
Matthew P. Padula	Oversaw the acquisition of mass spectrometry data.
Ian G. Charles	Provided input on the study.
Roger Dumke	Supervised and secured funding for the binding studies performed by Lisa Hagemann and Anna Gründel.
Steven P. Djordjevic	Initiated and secured funding for this study. Together with Michael Widjaja and Kate Harvey wrote the manuscript.

5.3 Cover page

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Elongation factor Tu is a multifunctional and processed moonlighting protein

Received: 15 May 2017
Accepted: 10 August 2017
Published online: 11 September 2017

Michael Widjaja[1], Kate Louise Harvey[1], Lisa Hagemann[2], Iain James Berry[1], Veronica Maria Jarocki[1], Benjamin Bernard Armando Raymond[1], Jessica Leigh Tacchi[1], Anne Gründel[2], Joel Ricky Steele[1], Matthew Paul Padula[3], Ian George Charles[4], Roger Dumke[2] & Steven Philip Djordjevic[1,3]

[1]The ithree institute, University of Technology Sydney, PO Box 123, Broadway, NSW, 2007, Australia. [2]Technische Universität Dresden, Medizinische Fakultät Carl Gustav Carus, Institut für Medizinische Mikrobiologie und Hygiene, Fetscherstrasse 74, 01307, Dresden, Germany. [3]Proteomics Core Facility, University of Technology Sydney, PO Box 123, Broadway, NSW, 2007, Australia. [4]Quadram Institute Bioscience, Norwich Research Park, Norwich, Norfolk, NR4 7UA, UK.
Michael Widjaja, Kate Louise Harvey and Lisa Hagemann contributed equally to this work.
Roger Dumke and Steven Philip Djordjevic jointly supervised this work.

Many bacterial moonlighting proteins were originally described in medically, agriculturally, and commercially important members of the low G + C Firmicutes. We show Elongation factor Tu (Ef-Tu) moonlights on the surface of the human pathogens *Staphylococcus aureus* ($Sa_{Ef\text{-}Tu}$) and *Mycoplasma pneumoniae* ($Mpn_{Ef\text{-}Tu}$), and the porcine pathogen *Mycoplasma hyopneumoniae* ($Mhp_{Ef\text{-}Tu}$). Ef-Tu is also a target of multiple processing events on the cell surface and these were characterised using an N-terminomics pipeline. Recombinant MpnEf-Tu bound strongly to a diverse range of host molecules, and when bound to plasminogen, was able to convert plasminogen to plasmin in the presence of plasminogen activators. Fragments of Ef-Tu retain binding capabilities to host proteins. Bioinformatics and structural modelling studies indicate that the accumulation of positively charged amino acids in short linear motifs (SLiMs), and protein processing promote multifunctional behaviour. Codon bias engendered by an A + T rich genome may influence how positively-charged residues accumulate in SLiMs.

SCIENTIFIC **REPORTS** | 7: *11227* | DOI:10.1038/s41598-017-10644-z

5.4 Abstract

Many bacterial moonlighting proteins were originally described in medically, agriculturally, and commercially important members of the low G+C Firmicutes. We show Elongation factor Tu (Ef-Tu) moonlights on the surface of the human pathogens *Staphylococcus aureus* ($Sa_{Ef\text{-}Tu}$) and *Mycoplasma pneumoniae* ($Mpn_{Ef\text{-}Tu}$), and the porcine pathogen *Mycoplasma hyopneumoniae* ($Mhp_{Ef\text{-}Tu}$). Ef-Tu is also a target of multiple processing events on the cell surface and these were characterised using an N-terminomics pipeline. Recombinant $Mpn_{Ef\text{-}Tu}$ bound strongly to a diverse range of host molecules, and when bound to plasminogen, was able to convert plasminogen to plasmin in the presence of plasminogen activators. Fragments of Ef-Tu retain binding capabilities to host proteins. Bioinformatics and structural modelling studies indicate that the accumulation of positively charged amino acids in short linear motifs (SLiMs), and protein processing promote multifunctional behaviour. Codon bias engendered by an A+T rich genome may influence how positively-charged residues accumulate in SLiMs.

5.5 Introduction

Elongation factor Thermo unstable (Ef-Tu) is one the most abundant proteins in bacteria[205,318]. It functions as an essential and universally conserved GTPase that ensures translational accuracy by catalysing the reaction that adds the correct amino acid to a growing nascent polypeptide chain[319]. After the incoming aminoacyl-tRNA docks with the mRNA, GTPase activity induces a conformational change releasing Ef-Tu from the ribosome[319-321]. In *E. coli*, Ef-Tu is comprised of three functional domains known as domain I (amino acids 1 – 200), domain II (amino acids 209 – 299) and domain III (amino acids 301 – 393)[322]. Domain I forms a helix structure with Rossmann fold topology, a structural motif found in proteins that bind nucleotides, while domains II and III are largely comprised of beta sheets[319,323]. The GTP/GDP binding domains are housed in domain I, while domains I and II are needed for nucleotide exchange. Domains II and III physically adjust to form an amino acid tRNA binding site[319,321]. Ef-Tu sequences derived from phylogenetically diverse species share considerable sequence identity and have been used to generate phylogenetic descriptions of the tree of life[324]. In eukaryotes, domain III also has a role in actin polymerisation via an actin-bundling domain[325,326]

Despite its highly conserved function in protein synthesis, non-canonical functions have been described for Ef-Tu in all kingdoms of life. Ef-Tu lacks a signal secretion motif yet the ability to execute moonlighting functions often requires the molecule to localise to the cell surface. Ef-Tu is a multifunctional protein in higher order eukaryotes[327-332], parasites[333-336], fungi[337] and it is has been identified on the surface of a wide range of Gram positive and Gram negative pathogenic and commensal bacteria that associate with metazoan species[188,205,338-344]. Bacterial Ef-Tu interacts with nucleolin[345,346], fibrinogen and factor H[339,341], plasminogen and several complement factors[341,342,347], laminin[348], CD21[349], fibronectin[205,348,350,351], is immunogenic[352] and adheres to the surface of Hep-2 cells[348] underscoring the multifunctional adhesive characteristics that have been assigned to this molecule. Ef-Tu binds sulfated carbohydrate moieties found on glycolipids and sulfomucin and promotes the binding of *Lactobacillus reuteri* to mucosal surfaces indicating that Ef-Tu can interact with carbohydrates[353]. Notably, antibodies against Ef-Tu are induced during infections caused by *S. aureus*[354,355] *Mycoplasma capricolum*[356], *Mycoplasma ovipneumoniae*[352], *Chlamydia trachomatis*[357], *Burkholderia pseudomallei*[358] and *M. hyopneumoniae*[359]. Ef-Tu has been identified in six surface proteome studies (excludes

cell membrane and envelope isolations)[179,360-364] performed on *S. aureus* and Ef-Tu is one of twelve proteins consistently identified in the exoproteome of *S. aureus* from patients with bacteraemia[365]. The major staphylococcal autolysin *Alt* is implicated in playing a role in secreting cytosolic proteins including Ef-Tu into the extracellular milieu[340]. Moonlighting proteins are likely to be exported via several mechanisms including within secreted extracellular vesicles[366], during cell lysis[193] and via association with proteins that are secreted by the Sec machinery[367].

The ability of Ef-Tu to be secreted onto the cell surface occurred early in the evolutionary interplay between plant pathogenic bacteria and their eukaryote hosts and is a well described pathogen-associated molecular pattern (PAMP) molecule[368,369]. Plants have evolved pattern-recognition receptors (PRR) in their cell membranes that are designed specifically to recognise PAMP molecules released by bacterial and fungal pathogens[369-375]. An Ef-Tu receptor (EFR) found within Brassica lineages[376,377] recognises the highly conserved N-terminal 18 amino acids (elf18) in the native Ef-Tu molecule[369,376,377]. Binding triggers signal transduction events in plant roots that ensure that pathogenic bacteria are either contained within callose deposits, destroyed by cellular apoptosis, or succumb to an oxidative burst elicited by the production of hydrogen peroxide[376]. A region spanning surface exposed amino acids 176 – 225 in Ef-Tu from the Gram-negative bacterial pathogen *Acidovorax avenae*, interacts with a different PRR in monocotyledonous plants (see Figure 5.1)[378]. EFR has been transferred from the Brassica species *Arabidopsis thaliana* into the monocot species, rice and transgenic rice plants display enhanced innate immune responses when exposed to elf18 from *Xanthomonas oryza*, a major rice pathogen[379]. These studies show that plants have evolved sophisticated molecular machinery to identify Ef-Tu that is released onto the cell surface by diverse plant pathogenic bacteria.

Protein cleavage is emerging as an important post-translational modification that can expand protein function[227,230,239,380]. This is evident in the genome-reduced Mollicutes where species-specific Mycoplasmal adhesins and lipoproteins are targets of complex processing events[186,187,216-227,230-232]. Cleavage fragments are retained on the bacterial cell surface and function as adhesins that bind heparin-like glycosaminoglycans[186,187,217,218,221,222,230], fibronectin[219,220,226,230] and circulatory molecules

such a plasmin(ogen) that regulate the fibrinolytic system[219-221,223,230]. Cleavage motifs have been chemically defined in *M. hyopneumoniae* using mass spectrometry and occur at phenylalanine residues in the motif S/T-X-F↓-X-D/E, within stretches of hydrophobic amino acids, and at trypsin-like sites in diverse molecules including adhesins, lipoproteins and in metabolic enzymes that traffic to the cell surface[186,221,224,225,227]. Cleavage fragments are known to be further processed by aminopeptidases[225,227] that also localise on the cell surface[243,244]. Proposed here is that protein processing represents another layer by which proteins can expand and modify protein function and is under recognised as a post-translational modification in prokaryotes.

In this study, full length Ef-Tu was identified, and an extensive repertoire of processed cleavage fragments of Ef-Tu, on the surface of human pathogens *S. aureus* and *M. pneumoniae,* and the porcine pathogen *M. hyopneumoniae*. Protein cleavage events were mapped using a systems-wide dimethyl labelling protocol that allows for the identification of modified N-terminal peptides (neo-N-termini) by liquid chromatography-tandem mass spectrometry (LC-MS/MS) and thus determine how Ef-Tu is processed and presented on the cell surfaces of these pathogens. This study further characterised the non-canonical functions of Ef-Tu from *M. pneumoniae* ($Mpn_{Ef\text{-}Tu}$) and show that it is a multifunctional protein that can not only bind to and activate plasminogen in the presence of host activators, but is also capable of binding to structurally and chemically diverse host molecules.

5.6 Methodology

Methods pertaining to studies on $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$ can be found in 'Appendix 5: 8.5.3 Supplementary Methodology'.

5.6.1 Strains, cultures, and reagents

M. pneumoniae (M129 strain; ATCC 29342) was cultured in modified Hayflick's medium at 37°C in tissue culture flasks as described previously[250].

Human lung carcinoma cells (A549; ATCC CCL-185) were cultured in RPMI 1640 medium (Invitrogen) supplemented with 10% heat inactivated fetal bovine serum at 37°C with 5% CO_2 in tissue culture flasks.

Host proteins used for affinity chromatography include: purified fibronectin (Code: 341635) and plasminogen (Code: 528175) from human plasma supplied by Merck Millipore; plasminogen from human plasma (Code: P7999) bovine actin (Code: A3653) and fetuin (Code: F3004) supplied by Sigma.

Human proteins used for ELISA include: plasminogen (Code: P7999), lactoferrin (Code: L1294), laminin (Code: L6274), vitronectin (Code: SRP3186), plasma fibrinogen (Code: F3879) and plasma fibronectin (Code: 11051407001) supplied by Sigma.

5.6.2 Enrichment of *M. pneumoniae* surface proteins

5.6.2.1 Biotinylation

Biotinylation of the *M. pneumoniae* cell surface was carried out as described in[230]. Adherent *M. pneumoniae* were incubated with in EZ-link sulfo-NHS-biotin (Thermo Fisher Scientific) for 30 seconds on ice. A detailed method is described in Chapter 2.

5.6.2.2 Trypsin shaving

Trypsin shaving of *M. pneumoniae* cells was carried out as described previously[187] with modifications. Trypsin was added to adherent *M. pneumoniae* cells for 5 minutes at 37°C. A detailed method is described in Chapter 2.

5.6.3 Preparation and separation of whole cell lysates for one- and two-dimensional gel electrophoresis

5.6.3.1 Whole cell lysis preparation

M. pneumoniae whole cell lysates were prepared as previously described[187]. Proteins were reduced and alkylated with 5 mM tributylphosphine and 20 mM acrylamide monomers for 90 min at room temperature. Insoluble material was removed by centrifugation and five volumes of acetone added to precipitate protein. After centrifugation, the protein pellet was solubilized in 7 M urea, 2 M thiourea, 1% (w/v) C7BzO for one- and two-dimensional gel electrophoresis.

5.6.3.2 1D and 2D SDS-PAGE protein separation

Protein separation was performed as described in[221,224]. 80 µg of protein was separated for 1D SDS-PAGE and 250 µg of protein was cup-loaded for 2D SDS-PAGE separation.

5.6.3.3 Trypsin Digest

In-gel trypsin digestion was performed as described in[186]. After digestion, tryptic peptides were stored at 4°C until needed for liquid chromatography tandem mass spectrometry.

5.6.4 Affinity chromatography of host binding *M. pneumoniae* proteins

Affinity purification of *M. pneumoniae* proteins that bind heparin, A549 surface proteins, andthe host proteins listed above (section '5.6.1 Strains, cultures, and reagents') was performed as described in[230]. A detailed method is described in Chapter 3.

5.6.5 Liquid chromatography tandem mass spectrometry (LC-MS/MS) and MS/MS data analysis

LC-MS/MS was performed as described in[224]. Mascot (Version 6.1) was used to search MS/MS data files as previously described[224].

Files were searched against the MSPnr100 database[251] with the following parameters. Fixed modifications: none. Variable modifications: propionamide, oxidized methionine, deamidation. Enzyme: semi-trypsin. Number of allowed missed cleavages: 3. Peptide mass tolerance: 100 ppm. MS/MS mass tolerance: 0.2 Da. Charge state: 2+, 3+ and 4+. For samples collected from the 'Biotinylation enrichment of surface proteins' and 'Affinity chromatography of A549 binding *M. pneumoniae* proteins', variable modifications also included NHS-LC-Biotin (K) and NHS-LC-Biotin (N-term). 'Avidin purification of A549 interacting proteins' was also searched against *homo sapiens* entries in MSPnr100 to identify biotinylated surface A549 proteins. A detailed method is described in Chapter 2.

5.6.6 Expression and purification of $rMpn_{Ef\text{-}Tu}$

Expression and purification of $rMpn_{Ef\text{-}Tu}$ was performed in one of two methods as described by[209] and[244].

The first method was performed as described in[210]. In brief, the *M. pneumoniae tuf* gene (*MPN_665*) was amplified and cloned with a N-terminal hexahistadine tail into the plasmid vector pET30 (Merck Millipore) containing a kanamycin resistance gene. The recombinant construct was transformed in to BL21-DE3 competent *E. coli* cells (Merck Millipore), induced with 1 mM isopropyl-β-D-thiogalactosidase (IPTG, Roth) and purified under denaturing conditions with immobilized metal affinity chromatography Ni^{2+}-charged resin

(Qiagen) as described by the manufacturer. Elutions were concentrated using a 30 kDa Vivaspin centrifugal device (Sartorius). Recombinant protein was assayed and stored at -20°C. $rMpn_{Ef\text{-}Tu}$ was used to produce guinea pig antiserum as reported[210].

The second method was performed as described in[244] with modifications. In brief, the *M. pneumoniae tuf* gene (*MPN_665*) was synthesized with an N-terminal hexahistadine tail and cloned by Blue Heron Biotech (USA) into a plasmid vector (PS100030) with an ampicillin resistance gene. The recombinant construct was transformed in to BL21-DE3 competent *E. coli* cells (Bioline), induced with 1 mM isopropyl-β-D-thiogalactosidase (IPTG, Bioline) and purified under denaturing conditions with Profinity immobilized metal affinity chromatography Ni^{2+}-charged resin (Bio-Rad). BL21cells were lysed in 8 M Urea, 100 mM Na_2HCO_3, 10 mM Tris-HCl, pH 8 with 6 rounds of sonication for 30 seconds on ice. Following centrifugation, the supernatant was added to Ni2+ resin overnight at 4°C. The resin was then loaded onto a column, washed four times with 5 ml 8 M Urea, 100 mM Na_2HCO_3, 10 mM Tris-HCl, pH 6.3 and a 2-step elution: three times with 5 ml 8 M Urea, 100 mM Na_2HCO_3, 10 mM Tris-HCl, pH 5.9 and twice with 10 ml 8 M Urea, 100 mM Na_2HCO_3, 10 mM Tris-HCl, pH 4.5**.** Washes and elutions were monitored by SDS-PAGE. Elutions were concentrated using a 10 kDa Microsep™ centrifugal device (Pall) and dialysed into PBS, 0.5% Tween 20 with 10,000 MWCO SnakeSkin® Dialysis Tubing (Thermo) at 4°C. Recombinant protein was assayed and stored at 4°C.

5.6.7 Binding of $rMpn_{Ef\text{-}Tu}$

5.6.7.1 Binding assays

For this experiment and all subsequent experiments, animal experiments were approved by the ethical board of Landesdirektion Sachsen, Dresden, Germany with the permit no. (permit 24-9168.25-1/2011-1). ELISA experiments were carried out as described in[209]. Guinea pig $rMpn_{Ef\text{-}Tu}$ antiserum (1:750) followed by anti-guinea pig IgG (1:1,000; Dako) dilutions were used. Tetramethylbenzidine (Sigma) was added followed by 1 M HCl and absorbance was measured at 450 nm (620 nm as reference).

5.6.7.2 Influence of anti-rMpn$_{Ef\text{-}Tu}$ on binding

Freshly grown A549 cells were used to coat wells in 96-well microtitre plates for 2 h at 37°C as described in[209]. rMpn$_{Ef\text{-}Tu}$ (10 µg/ml) was incubated with guinea pig rMpn$_{Ef\text{-}Tu}$ antiserum or pre-immune serum (1:100) concentrations were used.

5.6.7.3 Binding of rMpn$_{Ef\text{-}Tu}$ to human proteins in ELISA

Purified human proteins used were supplied by Sigma as described above (section '5.6.1 Strains, cultures, and reagents'). Binding of rMpn$_{Ef\text{-}Tu}$ (15 µg/ml) to extracellular matrix proteins was performed as described previously[209]. The dilutions for the appropriate antisera are: (Sigma) anti-plasminogen: 1:2,500; anti-lactoferrin 1:5,000; anti-laminin 1:750; anti-vitronectin 1:5,000; anti-fibrinogen 1:3,000; anti-fibronectin 1:1,000. Followed by anti-rabbit IgG (Dako) or anti-goat IgG (both 1:2,000).

5.6.7.4 Microscale thermophoresis

Microscale thermophoresis to determine the binding affinities between Ef-Tu and a fluorescently labelled host protein was performed as described in[226]. Time for Microscale thermophoresis was set to 30 s with fluorescence set to 5 s before and 30 s after each run. Each sample was scanned with 40%, 60% and 80% MST Power. Dissociation curves were plotted with hot/cold, jump or thermophoresis settings to determine dissociation constant.

5.6.8 Binding affinity of rMpn$_{Ef\text{-}Tu}$ to plasminogen

5.6.8.1 Effect of NaCl on plasminogen binding

Briefly, 96-well microtitre plates were coated with rMpn$_{Ef\text{-}Tu}$ as described. Plasminogen (2.5 µg) together with increasing concentrations of NaCl were added to the wells and incubated for 1.5 h at 37°C. Wells were incubated with rabbit anti-plasminogen (1:3,000) followed by anti-rabbit IgG (1:2,000). Detection was done as described above (section '5.6.7.1 Binding assays').

5.6.8.2 Effect of ε-aminocaproic acid on plasminogen binding

ELISA was carried out as reported in[209]. In brief, the wells of ELISA plates were coated with rMpn$_{Ef\text{-}Tu}$. 2.5 µg of plasminogen and increasing concentrations of ε-aminocaproic acid were added to the wells and incubated for 1.5 h at 37°C. Wells were incubated with

rabbit anti-plasminogen (1:3,000) followed by anti-rabbit IgG (1:2,000) and OD_{420nm} was measured.

5.6.8.3 Plasminogen activation and degradation of human fibrinogen and vitronectin

Degradation of human fibrinogen and vitronectin by activated plasminogen was carried out as described in[209].10 µg/ml of human plasminogen was added to the wells which were then incubated with fibrinogen or vitronectin (each 15 µg/ml) and urinary plasminogen activator (uPA; Sigma) or tissue plasminogen activator (tPA; Sigma; each 75 ng/ml).

5.6.8.4 Binding of anti- $rMpn_{Ef\text{-}Tu}$ antibodies to *M. pneumoniae* whole cell lysate proteins

Freshly grown *M. pneumoniae* cells were harvested and used to coat wells in 96-well microtitre plate for 2 h at 37°C as described previously[210]. Wells were blocked before adding guinea pig $rMpn_{Ef\text{-}Tu}$ antisera (1:500) followed by anti-guinea pig IgG (1:1,000). As a control wells were incubated with guinea pig antisera raised against total *M. pneumoniae* proteins.

5.6.9 Surface localisation of Ef-Tu on *M. pneumoniae*

5.6.9.1 Localisation of Ef-Tu on the surface of *M. pneumoniae* colonies

M. pneumoniae colonies were grown on PPLO agar plates and blotted onto nitrocellulose as described previously[210]. Antisera to PdhB and 1-phosphofructokinase (FruK) were used as positive and negative controls, respectively.

5.6.9.2 Surface localisation of Ef-Tu on *M. pneumoniae* cells

Immunofluorescence experiments were carried out as described in[210]. Again guinea pig antisera to PdhB and FruK were used as positive and negative controls, respectively.

5.6.10 Dimethyl labelling and LC-MS/MS analysis of *M. pneumoniae, M. hyopneumoniae and S. aureus* proteins

5.6.10.1 Dimethyl labelling of proteins

Dimethyl labelling of proteins was performed as described previously[227,230].

5.6.10.2 LC-MS/MS of dimethyl labelled proteins

Dimethyl labelled proteins were analysed by two mass spectrometers; the Sciex 5600 and the Thermo Scientific Q Exactive™. For full technical set up and method details see 'Appendix 5: 8.5.3 Supplementary methodology'.

5.6.11 Bioinformatic analysis of Ef-Tu

Bioinformatic analysis of Ef-Tu used the online resources: ProtParam[268], Clustal Omega[381], SignalP 4.1 Server[256], SecretomeP 2.0 Server[257], TMpred[258] and COILS (Addition of 'yes' to 2.5 fold weighting of positions a,d)[312]. The amino acid sequences of $Mpn_{Ef\text{-}Tu}$ (Uniprot #: P23568), $Mhp_{Ef\text{-}Tu}$ (Uniprot #: Q4A9G1) and $Sa_{Ef\text{-}Tu}$ (Uniprot #: Q2G0N0) were analysed using a variety of bioinformatics tools. Conservation of amino acid positions in each protein were detected using The ConSurf server[382]. Putative heparin binding sites were identified using the search patterns x-[HKR]-x(0,2)-[HKR]-x(0,2)-[HKR]-x and x-[HKR]-x(1,4)-[HKR]-x(1,4)-[HKR]-x via ScanProsite[315]. Putative protein:protein and protein:nucleic acid interaction sites were identified using ISIS[383]. Intrinsically disordered regions were predicted by Meta-Disorder[384,385], which combines the outputs from original prediction methods NORSnet, DISOPRED2, PROFbval and Ucon. Solvent accessibility of each amino acid position was ascertained using evolutionary information from multiple sequence alignments and a multi-level system[386]. Nucleotide-, DNA-, and RNA-binding regions were predicted by SomeNA[387]. MEROPs was used to identify proteases responsible for the cleavage events of Ef-Tu[388].

5.7 Results

5.7.1 Bioinformatic analysis of $Mhp_{Ef\text{-}Tu}$, $Sa_{Ef\text{-}Tu}$, and $Mpn_{Ef\text{-}Tu}$

The amino acid sequences of Ef-Tu from *M. pneumoniae* ($Mpn_{Ef\text{-}Tu}$), *M. hyopneumoniae* ($Mhp_{Ef\text{-}Tu}$), and *S. aureus* ($Sa_{Ef\text{-}Tu}$) share 60.7% sequence identity. $Mpn_{Ef\text{-}Tu}$ resides on the cell surface of *M. pneumoniae* and binds fibronectin[205]. The fibronectin binding regions have been mapped and are located at the end of domain I and at the beginning of domain II[389,390] and most of domain III is also involved in binding fibronectin[389]. It is not known if sequence conservation in fibronectin binding regions of $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$ is sufficient to afford these Ef-Tu homologs the ability to bind fibronectin. Several *Mycoplasma* spp.[217,272] and *S. aureus*[283,391,392] are known to interact with heparin. Putative heparin binding domains were computationally predicted and mapped onto each of the Ef-Tu molecules

(Figure 5.1). Several of these were conserved in all three Ef-Tu sequences in domains I, II and III.

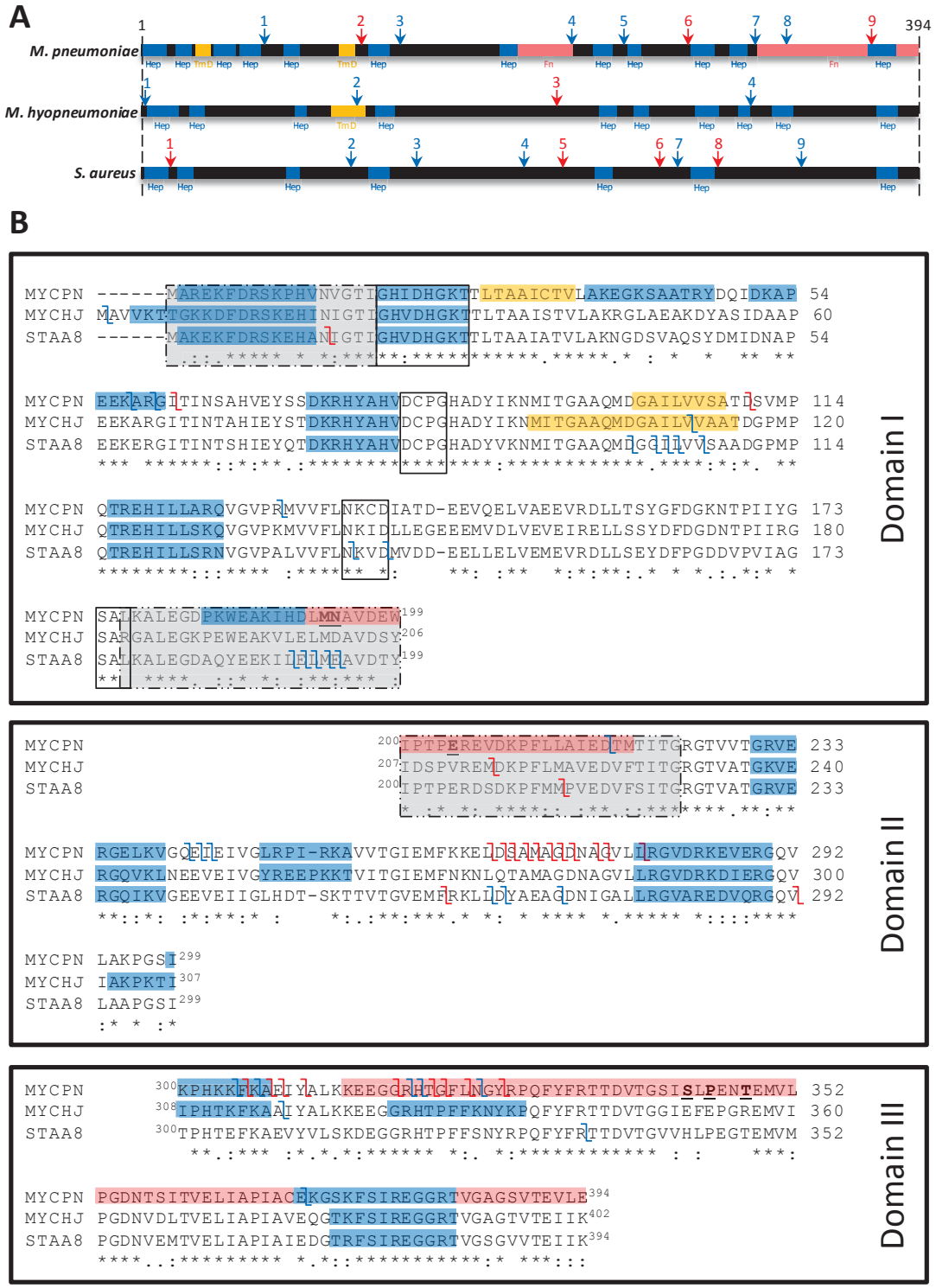


Figure 5.1: Bioinformatic analysis of $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$. A) Schematic of $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$ highlighting putative heparin and fibronectin binding domains, transmembrane domains, and cleavage sites. ScanProsite[315] was used to predict heparin binding motifs (dark blue boxes) by searching clusters of basic residues with the "X-[HRK]-X(0,2)-[HRK]-X(0,2)-[HRK]-X" and "X-[HRK]-X(1,4)-[HRK]-X(1,4)-[HRK]-X" motifs. A putative transmembrane domain (Score 505) was predicted in $Mhp_{Ef\text{-}Tu}$ using TMPred[258] (yellow box). In $Mpn_{Ef\text{-}Tu}$, two fibronectin binding regions (salmon boxes) and two predicted transmembrane domains (scores) are depicted in Panels A and B[389,390]. Key amino acids in $Mpn_{Ef\text{-}Tu}$ involved in binding fibronectin[389,390] are underlined. Cleavage sites identified in this study are shown as arrows above the black bar (blue indicates cleavage sites identified by dimethyl labelling and red indicates cleavage sites identified by the characterisation of semi-tryptic peptides by LC-MS/MS). B) Amino acid sequence alignments of $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$. For consistency, features described in Figure 5.1A are represented by the same colour scheme in Figure 5.1B. Cleavage sites identified in this study are depicted by the symbol ↿. Sequence alignments have been separated into the three domains and the nucleotide binding motifs (boxed regions) and the two pattern recognition receptors (broken black outline grey box from *Acidovorax avenae*[378] and *Brassica*-specific receptors[377]) are shown.

5.7.2 $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$ are accessible on the bacterial surface and are retained during heparin agarose chromatography

LC-MS/MS analysis of tryptic peptides released from the cell surface of *M. pneumoniae*, *M. hyopneumoniae*, and *S. aureus* were separately mapped to $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$, respectively. In other experiments, tryptic peptides generated by digesting biotinylated cell surface proteins that were captured by Avidin Agarose chromatography were also separately mapped to $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$. Peptides identified by mass spectrometry from both techniques spanned the entire length of Ef-Tu, (Appendix 5: 8.5.1: Figure 1) consistent with the hypothesis that a sub-population of Ef-Tu molecules are exposed on the cell surface of the three pathogens (red bars for biotinylation, and green and grey bars for shaving in Figure 2) while the remainder perform an essential function in the cytosol. Tryptic peptides spanning the length of $Mpn_{Ef\text{-}Tu}$ were also

characterised when LC-MS/MS analysis was performed on tryptic digests of high salt (> 500 mM) eluents of proteins that were retained on heparin agarose (blue bars in Figure 5.2). $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$ were also identified by LC-MS/MS of high salt eluents from heparin affinity chromatography (blue bars in Appendix 5: 8.5.1: Figure 2 and Figure 3, respectively).

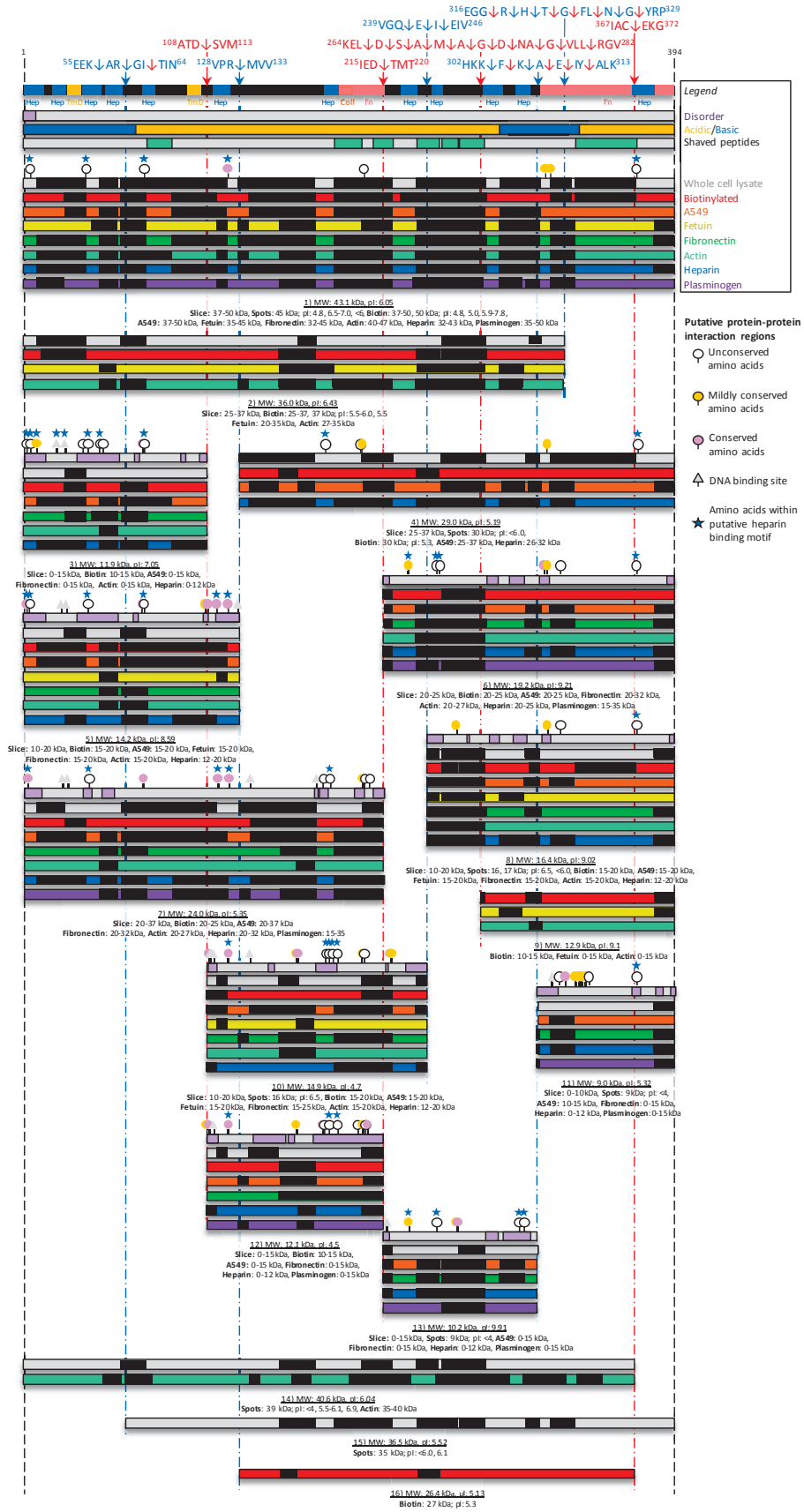

316EGG↓R↓H↓T↓G↓FL↓N↓G↓YRD329
239VGQ↓E↓I↓EIV246
367IAC↓EKG372
108ATD↓SVM113
264KEL↓D↓S↓A↓M↓A↓G↓D↓NA↓G↓VLL↓RGV282
55EEK↓AR↓GI↓TIN64
128VPR↓MVV133
215IED↓TMT220
302HKK↓F↓K↓A↓E↓IY↓ALK313
1
394
Hep
TmD
Coil
Fn
Legend
Disorder
Acidic/Basic
Shaved peptides
Whole cell lysate
Biotinylated
A549
Fetuin
Fibronectin
Actin
Heparin
Plasminogen
1) MW: 43.1 kDa, pI: 6.05
Slice: 37-50 kDa, Spots: 45 kDa; pI: 4.8, 6.5-7.0, <6, Biotin: 37-50, 50 kDa; pI: 4.8, 5.0, 5.9-7.8,
A549: 37-50 kDa, Fetuin: 35-45 kDa, Fibronectin: 32-45 kDa, Actin: 40-47 kDa, Heparin: 32-43 kDa, Plasminogen: 35-50 kDa
2) MW: 36.0 kDa, pI: 6.43
Slice: 25-37 kDa, Biotin: 25-37, 37 kDa; pI: 5.5-6.0, 5.5
Fetuin: 20-35 kDa, Actin: 27-35 kDa
3) MW: 11.9 kDa, pI: 7.05
Slice: 0-15 kDa, Biotin: 10-15 kDa, A549: 0-15 kDa,
Fibronectin: 0-15 kDa, Actin: 0-15 kDa, Heparin: 0-12 kDa
4) MW: 29.0 kDa, pI: 5.19
Slice: 25-37 kDa, Spots: 30 kDa; pI: <6.0,
Biotin: 30 kDa; pI: 5.3, A549: 25-37 kDa, Heparin: 26-32 kDa
5) MW: 14.2 kDa, pI: 8.59
Slice: 10-20 kDa, Biotin: 15-20 kDa, A549: 15-20 kDa, Fetuin: 15-20 kDa,
Fibronectin: 15-20 kDa, Actin: 15-20 kDa, Heparin: 12-20 kDa
6) MW: 19.2 kDa, pI: 9.21
Slice: 20-25 kDa, Biotin: 20-25 kDa, A549: 20-25 kDa, Fibronectin: 20-32 kDa,
Actin: 20-27 kDa, Heparin: 20-25 kDa, Plasminogen: 15-35 kDa
7) MW: 24.0 kDa, pI: 5.35
Slice: 20-37 kDa, Biotin: 20-25 kDa, A549: 20-37 kDa
Fibronectin: 20-32 kDa, Actin: 20-27 kDa, Heparin: 20-32 kDa, Plasminogen: 15-35
8) MW: 16.4 kDa, pI: 9.02
Slice: 10-20 kDa, Spots: 16, 17 kDa; pI: 6.5, <6.0, Biotin: 15-20 kDa, A549: 15-20 kDa,
Fetuin: 15-20 kDa, Fibronectin: 15-20 kDa, Actin: 15-20 kDa, Heparin: 12-20 kDa
9) MW: 12.9 kDa, pI: 9.1
Biotin: 10-15 kDa, Fetuin: 0-15 kDa, Actin: 0-15 kDa
10) MW: 14.9 kDa, pI: 4.7
Slice: 10-20 kDa, Spots: 16 kDa; pI: 6.5, Biotin: 15-20 kDa, A549: 15-20 kDa,
Fetuin: 15-20 kDa, Fibronectin: 15-25 kDa, Actin: 15-20 kDa, Heparin: 12-20 kDa
11) MW: 9.0 kDa, pI: 5.32
Slice: 0-10 kDa, Spots: 9kDa; pI: <4,
A549: 10-15 kDa, Fibronectin: 0-15 kDa,
Heparin: 0-12 kDa, Plasminogen: 0-15 kDa
12) MW: 12.1 kDa, pI: 4.5
Slice: 0-15 kDa, Biotin: 10-15 kDa,
A549: 0-15 kDa, Fibronectin: 0-15 kDa,
Heparin: 0-12 kDa, Plasminogen: 0-15 kDa
13) MW: 10.2 kDa, pI: 9.91
Slice: 0-15 kDa, Spots: 9kDa; pI: <4, A549: 0-15 kDa,
Fibronectin: 0-15 kDa, Heparin: 0-12 kDa, Plasminogen: 0-15 kDa
14) MW: 40.6 kDa, pI: 6.04
Spots: 39 kDa; pI: <4, 5.5-6.1, 6.9, Actin: 35-40 kDa
15) MW: 36.5 kDa, pI: 5.52
Spots: 35 kDa; pI: <6.0, 6.1
16) MW: 26.4 kDa, pI: 5.13
Biotin: 27 kDa; pI: 5.3
Putative protein-protein interaction regions
Unconserved amino acids
Mildly conserved amino acids
Conserved amino acids
DNA binding site
Amino acids within putative heparin binding motif

Figure 5.2: Cleavage map of $Mpn_{Ef\text{-}Tu}$. Full length Ef-Tu is represented as the black bar. Cleavage sites were identified by identifying dimethyl labelled peptides (blue arrows and broken lines) and by characterising semi-tryptic peptides generated after trypsin digestion (red arrows and broken lines). Exact cleavage sites are shown in the amino acid sequences above the black bar. The two fibronectin binding regions (salmon boxes, Fn) and two putative transmembrane domains (TmD, yellow boxes) described by[389,390] were included. Bioinformatic tools such as ScanProsite[315], COILS[312] and Meta-Disorder[384] were used to predict putative heparin binding motifs (Hep, blue boxes), coiled-coils (Coil, orange boxes), and disordered regions (purple boxes in grey bars), respectively. Peptides released from trypsin shaving of cells are shown as the green boxes in the grey bar (labelled 'Shaved peptides'). Tryptic peptides identified by mass spectrometry (black boxes within coloured bars) within Ef-Tu fragments were obtained from 1D- and 2D-SDS PAGE of bacterial whole cell lysates (grey bars); avidin affinity chromatography of: biotinylated surface proteins (red bars), A549 surface proteins (orange bars), fetuin (yellow bars), fibronectin (green bars), actin (teal bars), and plasminogen (purple bars); and peptides identified from heparin agarose affinity chromatography (blue bars). Circles and triangles just above fragments denote amino acid positions that are predicted to be surface exposed and represent putative protein:protein interaction regions (visual cues can be seen on the right of the cleavage map, sites listed in Appendix 5: 8.5.2: Table 1). Those marked with an additional star denote amino acid residues that fall within predicted putative heparin binding domains. White circles mark evolutionary unconserved binding regions, whilst yellow circles are mildly conserved and pink are highly conserved. Amino acid positions marked by grey triangles depict predicted nucleic acid interaction regions.

5.7.3 $Mhp_{Ef\text{-}Tu}$, $Sa_{Ef\text{-}Tu}$, and $Mpn_{Ef\text{-}Tu}$ are cleaved on the bacterial cell surface

As part of a larger study that sought to identify the repertoire of proteins in *M. pneumoniae* that are targets of proteolytic processing events, a dimethyl labelling protocol to tag N-terminal peptides was used to identify precise endoproteolytic cleavage sites in Ef-Tu (Table 5.1; seen as 'ι' in Figure 5.1; and seen as blue and red arrows in Figure 5.2). This was also applied to *M. hyopneumoniae* and *S. aureus* (Appendix 5: 8.5.2: Table 2, Figure 2, and Figure 3). Further evidence that $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$ are targets of protein cleavage events was obtained by LC-MS/MS analysis of: i) SDS-PAGE gel slices separately loaded with biotinylated *M. pneumoniae*, *M. hyopneumoniae* and *S. aureus* surface proteins captured by avidin chromatography, ii) bacterial proteins that eluted from heparin agarose using high salt (> 500 mM NaCl), iii) protein spots representing bacterial whole cell lysates and surface biotinylated proteins separated by 2D-PAGE, and iv) size fractioned whole cell lysate proteins resolved by SDS-PAGE.

Table 5.1: Dimethyl labelled and semi-tryptic peptides identified in $Mpn_{Ef\text{-}Tu}$.

No.	ID	Peptide Sequence	Score	E-value
		Dimethyl labelled peptides		
1	N1	K.^{58}**A**RGITINSAHVEYSSDKR^{75}.H	41	3.00E^{-03}
	N2	R.^{60}**G**ITINSAHVEYSSDKR^{75}.H	94	1.20E^{-08}
3	N3	R.^{131}**M**VVFLNK^{137}.C	57	1.30E^{-03}
5	N4	Q.^{242}**E**IEIVGLRPIR^{252}.K	48	3.30E^{-03}
	N5	E.^{243}**I**EIVGLRPIR^{252}.K	35	0.033
	N6	I.^{244}**E**IVGLRPIR^{252}.K	37	0.014
7	N7	K.^{305}**F**KAEIYALKKEEGGR^{319}.H	110	9.40E^{-09}
	N8	K.^{307}**A**EIYALKKEEGGR^{319}.H	69	2.60E^{-05}
8	N9	R.^{320}**H**TGFLNGYRPQFYFR^{334}.T	61	2.80E^{-05}
	N10	H.^{321}**T**GFLNGYRPQFYFR^{334}.T	39	2.90E^{-03}
	N11	N.^{326}**G**YRPQFYFR^{334}.T	43	1.70E^{-03}
		Semi-tryptic C-terminal Peptides		
6	S1	R.^{253}KAVVTGIEMFKKEL**D**^{267}.S	47	4.50E^{-04}
	S2	R.^{253}KAVVTGIEMFKKELD**S**^{268}.A	56	4.50E^{-05}
	S3	R.^{253}KAVVTGIEMFKKELDSAM**A**^{273}.G	55	7.30E^{-05}

	S4	R.^{253}KAVVTGIEMFKKELDSAMA**G**^{272}.D	95	4.50E^{-09}
	S5	R.^{253}KAVVTGIEMFKKELDSAMAGDN**A**^{275}.G	96	3.30E^{-09}
	S6	R.^{253}KAVVTGIEMFKKELDSAMAGDNA**G**^{276}.V	113	2.90E^{-09}
	S7	R.^{253}KAVVTGIEMFKKELDSAMAGDNAGVL**L**^{279}.R	55	2.00E^{-05}
7	S8	R.^{290}GQVLAKPGSIKPHKK**F**^{305}.K	46	1.50E^{-04}
	S9	R.^{290}GQVLAKPGSIKPHKKFK**A**^{307}.E	61	2.10E^{-06}
	S10	R.^{290}GQVLAKPGSIKPHKKFKA**E**^{308}.I	27*	4.40E^{-03}
	S11	R.^{290}GQVLAKPGSIKPHKKFKAEI**Y**^{310}.A	38	2.40E^{-03}
8	S12	R.^{290}GQVLAKPGSIKPHKKFKAEIYALKKEEG**G**^{318}.R	85	7.00E^{-08}
	S13	R.^{290}GQVLAKPGSIKPHKKFKAEIYALKKEEGGR**H**^{320}.T	24	8.20E^{-03}
	S14	R.^{290}GQVLAKPGSIKPHKKFKAEIYALKKEEGGRH**T**^{321}.G	15*	0.048
	S15	R.^{290}GQVLAKPGSIKPHKKFKAEIYALKKEEGGRHT**G**^{322}.F	48	3.00E^{-04}
	S16	R.^{290}GQVLAKPGSIKPHKKFKAEIYALKKEEGGRHTGFL**N**^{325}.G	25*	9.50E^{-03}
	S17	R.^{320}HTGFLN**G**^{326}.Y	21*	0.058
		Semi-tryptic N-terminal Peptides		
1	S18	I.^{62}**T**INSAHVEYSSDKR^{75}.H	37	4.60E^{-03}
2	S19	D.^{111}**S**VMPQTREHILLAR^{124}.Q	65	7.00E^{-05}
4	S20	D.^{218}**T**MTITGR^{224}.G	41	0.041
6	S21	L.^{267}**D**SAMAGDNAGVLLR^{280}.G	73	2.40E^{-06}
	S22	D.^{268}**S**AMAGDNAGVLLR^{280}.G	85	3.60E^{-06}
	S23	S.^{269}**A**MAGDNAGVLLR^{280}.G	58	1.30E^{-03}
	S24	A.^{270}**M**AGDNAGVLLR^{280}.G	53	7.20E^{-04}
	S25	M.^{271}**A**GDNAGVLLR^{280}.G	75	1.90E^{-05}
	S26	A.^{272}**G**DNAGVLLR^{280}.G	59	7.70E^{-04}
	S27	G.^{273}**D**NAGVLLR^{280}.G	57	1.30E^{-03}
	S28	D.^{274}**N**AGVLLR^{280}.G	42	0.031
8	S29	H.^{321}**T**GFLNGYRPQFYFR^{334}.T	76	6.00E^{-06}
	S30	T.^{322}**G**FLNGYRPQFYFR^{334}.T	47	3.00E^{-03}
	S31	G.^{323}**F**LNGYRPQFYFR^{334}.T	77	3.00E^{-05}
	S32	L.^{325}**N**GYRPQFYFR^{334}.T	62	5.30E^{-05}
	S33	N.^{326}**G**YRPQFYFR^{334}.T	47	2.90E^{-04}
	S34	G.^{327}**Y**RPQFYFR^{334}.T	59	1.40E^{-03}

9	S35	C.^{370}**E**KGSKFSIR^{378}.E	66	1.30E^{-03}
	S36	C.^{370}**E**KGSKFSIREGGR^{382}.T	35	6.60E^{-03}

Identified peptides have a Mascot score > 33 and an E-value < 0.05 unless marked with a *. Peptides marked with a * implies the peptide score was < 33 but still lies within major cleavage site. The exact site of cleavage is to the left of the amino acid that is bold and underlined for N-terminal cleavage fragments and to the right of C-terminal cleavage fragments. Amino acid numbers are written at the start and end of each peptide identified by LC-MS/MS.

The full length (labelled fragment 1) and 15 cleavage fragments of $Mpn_{Ef\text{-}Tu}$ were identified in this study, 12 of which were identified in the biotinylated 1D and 2D SDS-PAGE (red bars in Figure 5.2; fragments: 1 – 10, 12, 16). Notably (including full length), four of the five cleavage fragments derived from $Mhp_{Ef\text{-}Tu}$ (Appendix 5: 8.5.1: Figure 2; fragments: 1, 2, 5, 6) and three of seven fragments of $Sa_{Ef\text{-}Tu}$ (Appendix 5: 8.5.1: Figure 3; fragments: 1, 5, 7) that were enriched during heparin affinity chromatography were also identified in biotinylation experiments. Including the full length Ef-Tu, eleven (blue bars in Figure 5.2; fragments: 1, 3 – 8, 10 – 13), five (blue bars in Appendix 5: 8.5.1: Figure 2; fragments: 1, 2, 4 – 6) and seven (blue bars in Appendix 5: 8.5.1: Figure 3; fragments: 1, 3 – 7, 9) cleavage fragments that span different regions of $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$ respectively were recovered from a heparin agarose chromatography using salt concentrations well above the physiological concentration of 150 mM. All the fragments recovered from heparin affinity chromatography across all three pathogens contained at least one of the predicted heparin binding domains that reside within $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$. These data suggest that the processing events that generate Ef-Tu cleavage fragments, occur on the surface of each of these pathogens and that the fragments may retain an ability to interact with high sulfated glycosaminoglycans such as heparin. To ascertain the nature of the protease(s) responsible for Ef-Tu surface cleavage, the MEROPs database[388] was used to search 56 cleavage events. However, no strong predictions could be made after searching both P4-P3-P2-P1↓P1'-P2'-P3'-P4' and P2-P1↓P1'-P2' cleavage motifs.

5.7.4 Processing events expose new predicted surface macromolecule interaction sites

A single heparin binding consensus motif (XBBBXXBX, where B is a basic residue) with the sequence ^{73}DKRHYAHV^{80} is found within the amino acid sequences of $Mpn_{Ef\text{-}Tu}$, $Sa_{Ef\text{-}Tu}$, and $Mhp_{Ef\text{-}Tu}$ (motif spans amino acids 79 – 86 in $Mhp_{Ef\text{-}Tu}$), yet several Ef-Tu fragments were retained during heparin agarose chromatography that did not span this motif. $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$ sequences were examined for additional motifs enriched with clustered basic residues. 12 putative heparin binding motifs were identified dispersed throughout $Mpn_{Ef\text{-}Tu}$ (Table 5.2); heparin binding motifs were also identified in $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$ (Appendix 5: 8.5.2: Table 3). Many of these putative heparin binding motifs, particularly sequences ^{37}AKEGKSAATRY^{47}, ^{183}PKWEAKIHD^{191} AND ^{248}LRPIRKA^{254} were localised to non-essential regions defined here as evolutionary unconserved regions (Table 5.2 and 'Appendix 5: Supplementary Bioinformatics'). Using ISIS[383], which predicts protein:protein interaction (PPI) sites from sequence information, $Mpn_{Ef\text{-}Tu}$ is predicted to have eight surface-exposed PPI sites that are capable of binding macromolecules (Appendix 5: 8.5.2: Table 1) such as glycosaminoglycans including four that reside within putative heparin binding motifs ^{2}A**RE**KFDRSKPHV^{13}, ^{73}**D**KRHYAHV^{80} and ^{370}E**K**GSKFSIREGGRT^{383} (Table 5.2). Notably, the key residues (underlined and in bold) in the four binding sites were all unconserved residues as determined by ConSurf[382]. Putative heparin binding fragments derived from $Mpn_{Ef\text{-}Tu}$ typically displayed more putative PPI sites and were more intrinsically disordered than the parent molecule and some fragments displayed putative nucleic acid interaction sites, which are absent in the unprocessed, parent molecule. Additionally, three short linear motifs (^{37}**AKEGKS**AAT**RY**^{47}, ^{183}**PKWEAK**I**H**D^{191}, and ^{248}LR**P**I**R**KA^{254}) located in unconserved regions of Ef-Tu that were not predicted binding sites in the parent molecule (fragment 1) were predicted to be exposed in $Mpn_{Ef\text{-}Tu}$ fragments: 3, 4, 5, 6, 7, 10, 12, and 13 (Figure 5.2).

Table 5.2: Putative heparin binding motifs identified in $Mpn_{Ef\text{-}Tu}$ (Uniprot #: P23568).

Sequences	
^{2}AREKFDRSKPHV^{13}	^{183}PKWEAKIHD^{191}
^{19}GHIDHGKT^{26} +	^{230}GRVERGELKV^{293} +
^{37}AKEGKSAATRY^{47}	^{248}LRPIRKA^{254}
^{51}DKAPEEKARG^{60}	^{279}LRGVDRKEVERG^{290}
^{73}DKRHYAHV^{80} *	^{299}IKPHKKFKA^{307}
^{116}TREHILLARQ^{125} +	^{370}EKGSKFSIREGGRT^{383} +

$Mpn_{Ef\text{-}Tu}$ was searched for patters x-[HKR]-x(0,2)-[HKR]-x(0,2)-[HKR]-x and x-[HKR]-x(1,4)-[HKR]-x(1,4)-[HKR]-x using ScanProsite[315]. * indicates the motif (dKRHyaH) which is found in all three pathogens. + indicates the motifs that are highly homologous (up to four non-basic residues different) in all three pathogens.

5.7.5 Molecular modelling of Ef-Tu

The prediction tool MODELLER[393] was used to predict the structures of Ef-Tu for all three pathogens based of Ef-Tu from *E. coli*. For the *M. pneumoniae* prediction, the *E. coli* Ef-Tu (PDB: 4G5G_A) had a structure ID percentage of 70.5% and a zDOPE score of -0.93. For *M. hyopneumoniae*, the *E. coli* Ef-Tu (PDB: 1DG1_H) had a structure ID percentage of 68.6% and a zDOPE score of -0.72. For *S. aureus*, the *E. coli* Ef-Tu (PDB: 1DG1_H) had a structure ID percentage of 75.1% and a zDOPE score of -0.88. All nine distinct cleavage sites for *M. pneumoniae* have been mapped in the ribbon structure (Figure 5.3A). Cleavage sites located in regions that are predicted to release the three domains are mostly surface accessible within the molecule. The location and accessibility of the heparin binding domains in $Mpn_{Ef\text{-}Tu}$, and the two published fibronectin binding domains in $Mpn_{Ef\text{-}Tu}$ are depicted in Figure 5.3B. The cleavage sites and putative heparin binding domains of $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$ have also been mapped in the ribbon (Appendix 5: 8.5.1: Figure 4) and surface structures (Appendix 5: 8.5.1: Figure 5), respectively.

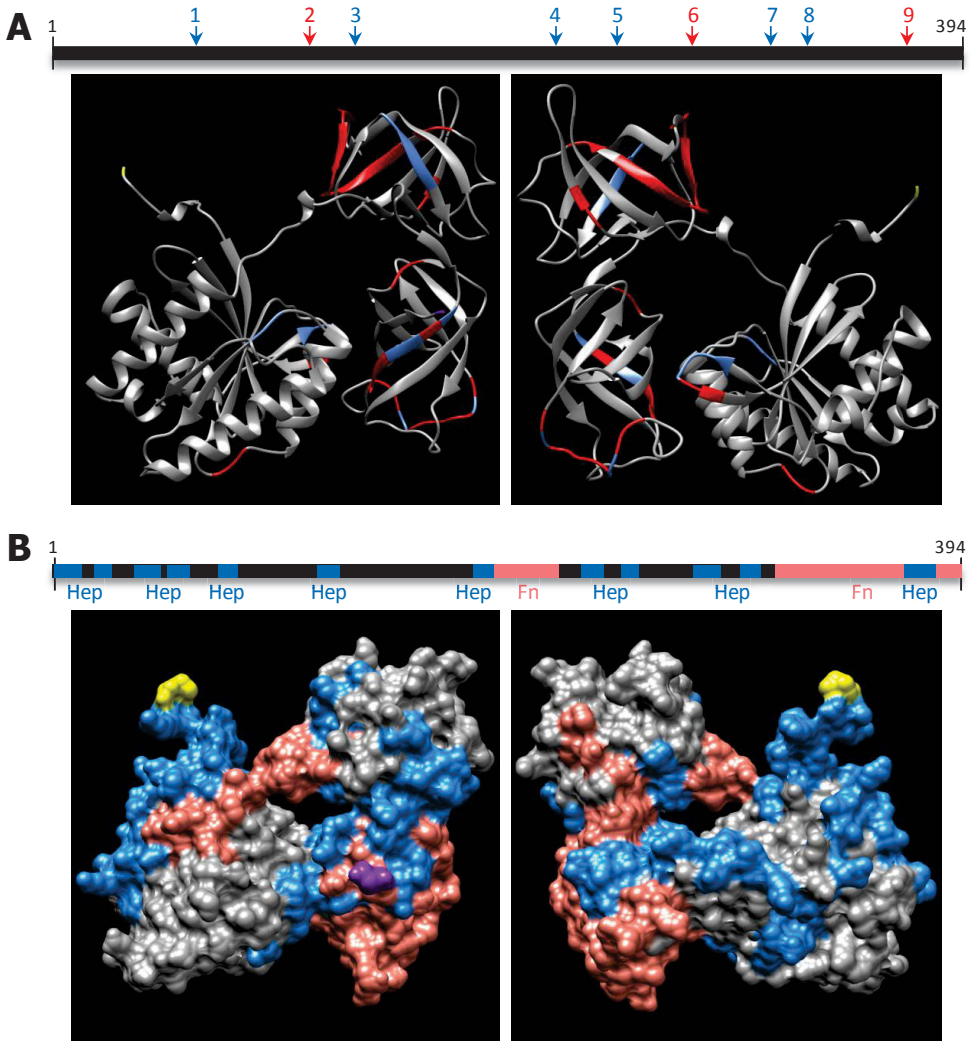


Figure 5.3: Predicted 3D structures of $Mpn_{Ef\text{-}Tu}$. Images are reversed sides for each structure. Ef-Tu is represented as black bars with cleavage sites shown in the ribbon structure (A) and the binding domains in the surface structure (B). The first amino acid (yellow) and last amino acid (purple) have been coloured. In (A), identified cleavage sites are displayed as arrows (blue, dimethyl labelling and red, semi-tryptic) with numbers to indicate each distinct cleavage site. Cleavage sites can also be seen in the ribbon structures as blue and red sections for dimethyl labelling and semi-tryptic sites, respectively. In (B), all 12 predicted heparin binding sites are shown in blue boxes and regions in the surface structure. The two published fibronectin binding domains have also been mapped in pink[389,390]. Structures were predicted by MODELLER and based on *E. coli* Ef-Tu homolog (PDB: 4G5G_A).

5.7.6 $Mpn_{Ef\text{-}Tu}$ and $Mhp_{Ef\text{-}Tu}$ are potential multifunctional binding proteins

It was notable that $Mpn_{Ef\text{-}Tu}$ was recovered from *M. pneumoniae* native cell lysates that were loaded onto affinity columns coupled with A549 epithelial cell surface proteins, fetuin, fibronectin, actin or plasminogen (Figure 5.2). Consistent with these data, $rMpn_{Ef\text{-}Tu}$ bound to immobilized A594 cells in microtitre plate binding assays (Figure 5.4A). Proteins that bind (recombinant pyruvate dehydrogenase subunit B) and that do not bind (P08 fragment of P1 adhesin) to A594 cells were used positive and negative controls respectively[210]. Binding of $rMpn_{Ef\text{-}Tu}$ to A594 cells was partially inhibited when anti-$rMpn_{Ef\text{-}Tu}$ antibodies, but not pre-immune antiserum, was present (Figure 5.4B). $Mhp_{Ef\text{-}Tu}$ was recovered from native cell lysates of *M. hyopneumoniae* that were loaded onto affinity columns coupled with PK15 epithelial cell surface proteins, fibronectin, actin, or plasminogen (Appendix 5: 8.5.1: Figure 2). $Mpn_{Ef\text{-}Tu}$ has previously been shown to bind fibronectin[205] andthis was independently confirmed in microtitre plate binding assays conducted in this study. Furthermore, the fibronectin binding assay conducted in this study suggests that *M. pneumoniae* encodes fibronectin binding proteins other than Ef-Tu (Figure 5.5). Including full length $Mpn_{Ef\text{-}Tu}$, 10 of the 16 fragments of $Mpn_{Ef\text{-}Tu}$ were recovered from affinity columns loaded with fibronectin (green bars in Figure 5.2; fragments: 1, 3, 5 – 8, 10 – 13). Of the ten fragments, seven spanned the known fibronectin binding regions described previously (Figure 5.2; fragments: 1, 6 – 8, 10 – 12)[389,390]. Fragments were identified from columns coupled to fibronectin that spanned the N-terminus of $Mpn_{Ef\text{-}Tu}$ suggesting that other fibronectin binding domains are yet to be identified in this molecule. $Mhp_{Ef\text{-}Tu}$ (fragment 1) and six cleavage fragments of $Mhp_{Ef\text{-}Tu}$ were retained by columns coupled with fibronectin (Appendix 5: 8.5.1: Figure 2; fragments: 1, 2, 4 – 8). The cleavage fragments spanned the N- and C-terminal ends, as well as the central region of $Mhp_{Ef\text{-}Tu}$ suggesting that it may contain fibronectin binding domains.

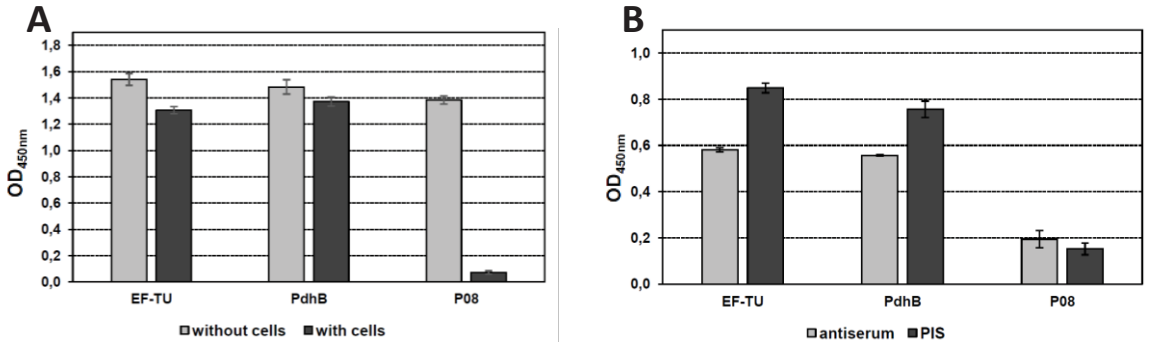


Figure 5.4: Binding of $rMpn_{Ef\text{-}Tu}$ to human A549 epithelial cells. A) A549 cells ('with cells') were bound to wells of a 96-well mictrotitre plate and incubated with rEf-Tu. Bound $rMpn_{Ef\text{-}Tu}$ was detected with antisera raised against $rMpn_{Ef\text{-}Tu}$. rPdhB and rP08 were used as a positive and negative control[210], respectively. Bars represent standard deviation of eight replicates. B) $rMpn_{Ef\text{-}Tu}$ was incubated with either antisera raised against $rMpn_{Ef\text{-}Tu}$ or pre-immune sera (PIS) and added to A549 cells in ELISA plates. rPdhB and rP08 and the corresponding antisera were used as a positive and negative control[210], respectively. Bars represent standard deviation of eight replicates.

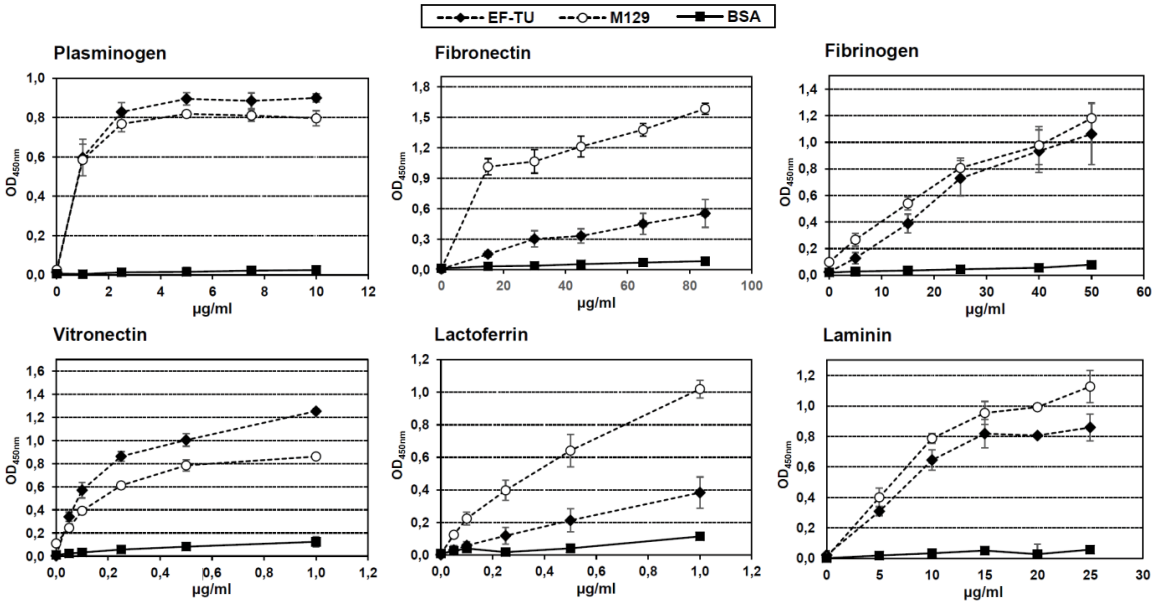


Figure 5.5: Microtitre plate binding assays depicting the interaction of $rMpn_{Ef\text{-}Tu}$ with human proteins. Wells of microtitre plates were coated with $rMpn_{Ef\text{-}Tu}$ and incubated with increasing concentrations of host proteins. Antisera against

each of the host proteins was used to detect interaction with $rMpn_{Ef\text{-}Tu}$. *M. pneumoniae* cells and BSA were used as a positive and negative control, respectively. Bars represent standard deviation of eight replicates.

M. pneumoniae[209,214,215] and *M. hyopneumoniae*[220,223] have both been shown to bind plasminogen onto their cell surface and assist with its conversion to plasmin. In the current study, $Mpn_{Ef\text{-}Tu}$ and $Mhp_{Ef\text{-}Tu}$ were both recovered during plasminogen agarose chromatography. Including the full length protein, six and seven fragments spanning different regions of $Mpn_{Ef\text{-}Tu}$ (purple bars in Figure 5.2; fragments: 1, 6, 7, 11, 12, 13) and $Mhp_{Ef\text{-}Tu}$ (purple bars in Appendix 5: 8.5.1: Figure 2; fragments: 1 – 6, 8), respectively were recovered from plasminogen-couple agarose beads.

Eleven fragments (including full length) spanning different regions of $Mpn_{Ef\text{-}Tu}$ (orange bars in Figure 5.2; fragments: 1, 3 – 8, 10 – 13) were identified from affinity columns coupled with biotinylated A549 surface proteins. Ten fragments (including full length) were recovered from actin-coupled columns (teal bars in Figure 5.2; fragments: 1 – 3, 5 – 10, 14) and six fragments of $Mpn_{Ef\text{-}Tu}$ were recovered during affinity chromatography using fetuin as 'Bait' (yellow bars in Figure 5.2; fragments: 1, 2, 5, 8 – 10). Full length and fragments of $Mhp_{Ef\text{-}Tu}$ were also identified in affinity chromatography with bound PK-15 surface proteins (two fragments; orange bars in Appendix 5: 8.5.1: Figure 2; fragments 1 and 2), and actin (five fragments; teal bars in Appendix 5: 8.5.1: Figure 2; fragments: 1, 2, 4, 5, 6).

5.7.7 $Mpn_{Ef\text{-}Tu}$ is a multifunctional adhesin

Antibodies raised against $rMpn_{Ef\text{-}Tu}$ were used to show that $Mpn_{Ef\text{-}Tu}$ resides on the surface of colonies of *M. pneumoniae* (Figure 5.6). The surface proteome study presented in Chapter 2 (Appendix 1) identified candidate proteins that could be used as a negative control for these studies and antibodies raised against recombinant 1-phosphofructokinase (FruK) from *M. pneumoniae* were used for this purpose (Figure 5.6). To further investigate binding capabilities, the ability of $rMpn_{Ef\text{-}Tu}$ to interact with a range of host molecules was examined by microscale thermophoresis. $rMpn_{Ef\text{-}Tu}$ bound fetuin (K_D = 53 ± 14 nM), actin (K_D = 19 ± 3 nM), and heparin (K_D = 42.5 ± 1.5 nM) in the nanomolar range, and to plasminogen (K_D = 933 ± 388 nM) in the micromolar range (Figure 5.7). These studies were extended using microtitre plate binding assays to confirm that $rMpn_{Ef\text{-}Tu}$ binds

plasminogen and fibronectin, and also show that $rMpn_{Ef\text{-}Tu}$ binds fibrinogen, vitronectin, lactoferrin and laminin in a dose dependent manner (Figure 5.5). Binding of $rMpn_{Ef\text{-}Tu}$ to plasminogen was significantly reduced by the addition of an increasing concentration of NaCl and ε-aminocaproic acid (Figure 5.8A). Notably, ε-aminocaproic acid was effective at blocking interactions between *M. pneumoniae* and plasminogen while high concentrations of NaCl were less effective (Figure 5.8A) suggesting that lysine residues play a significant role in binding interactions between Ef-Tu or *M. pneumoniae* with plasminogen.

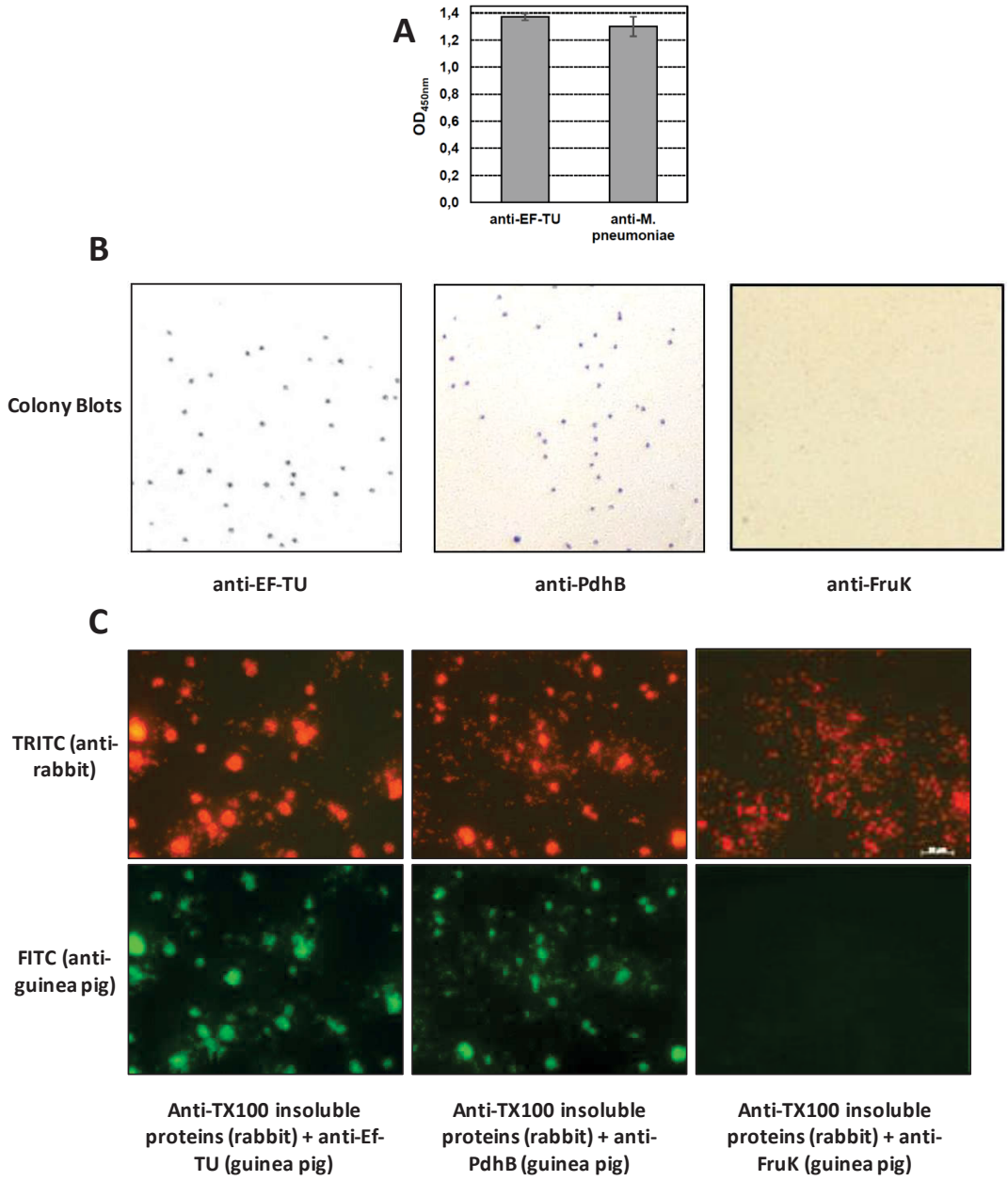


Figure 5.6: $Mpn_{Ef\text{-}Tu}$ resides on the surface of *M. pneumoniae*. A) Anti - $rMpn_{Ef\text{-}Tu}$ antibodies recognise *M. pneumoniae* cells in a whole cell ELISA assay; eight replicates. B) Colony blots of *M. pneumoniae* probed with anti - $rMpn_{Ef\text{-}Tu}$ antibodies, PdhB (positive control) and 1-phosphofructokinase (negative control). C) Immunofluorescence microscopy of *M. pneumoniae* cells probed with antibodies against TX100 insoluble proteins (TRITC; cell control), Ef-Tu, PdhB (positive control) and 1-phosphofructokinase as the negative control (FITC).

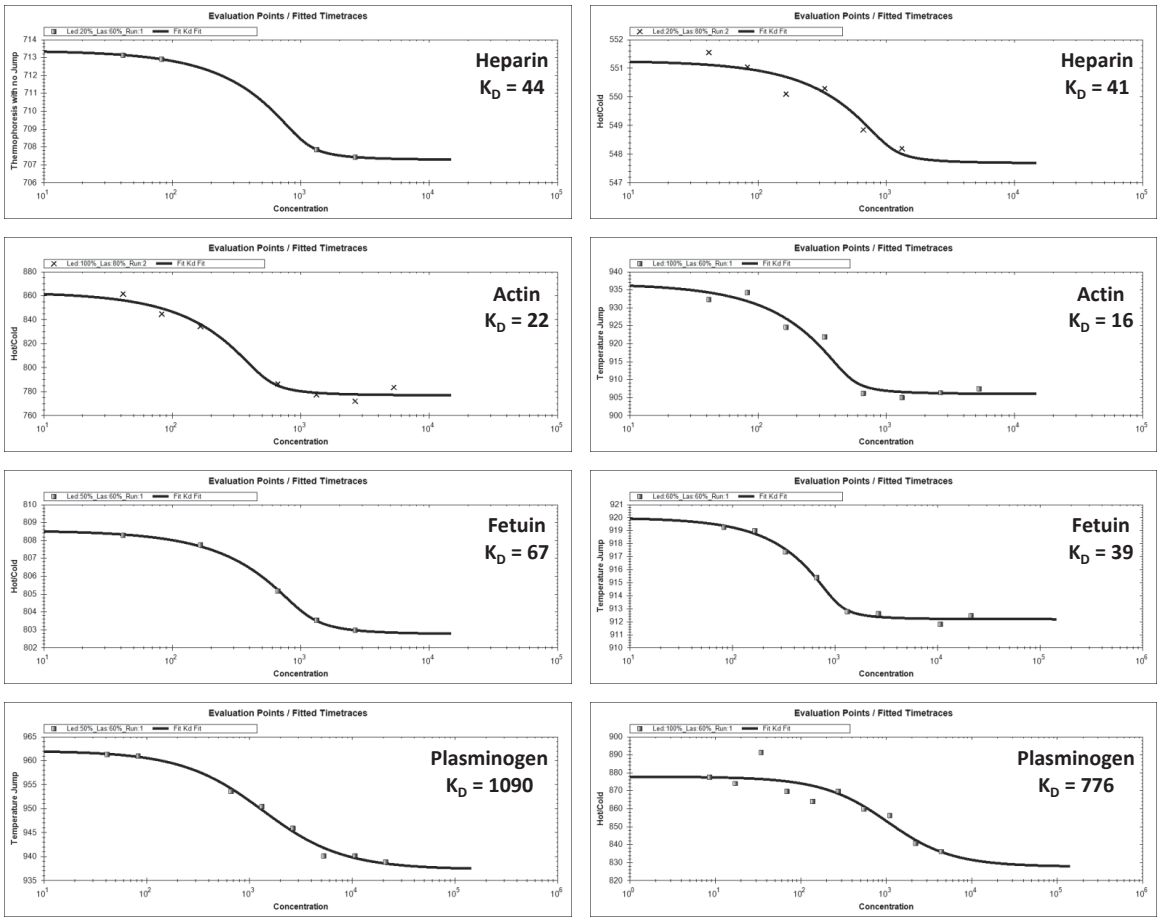


Figure 5.7: Microscale thermophoresis output depicting the interaction of rMpnEf-Tu with human molecules. Concentration of rMpnEf-Tu is plotted against thermophoretic movement of fluorescent human molecules. Experiments performed in duplicate, each panel representing one replicate.

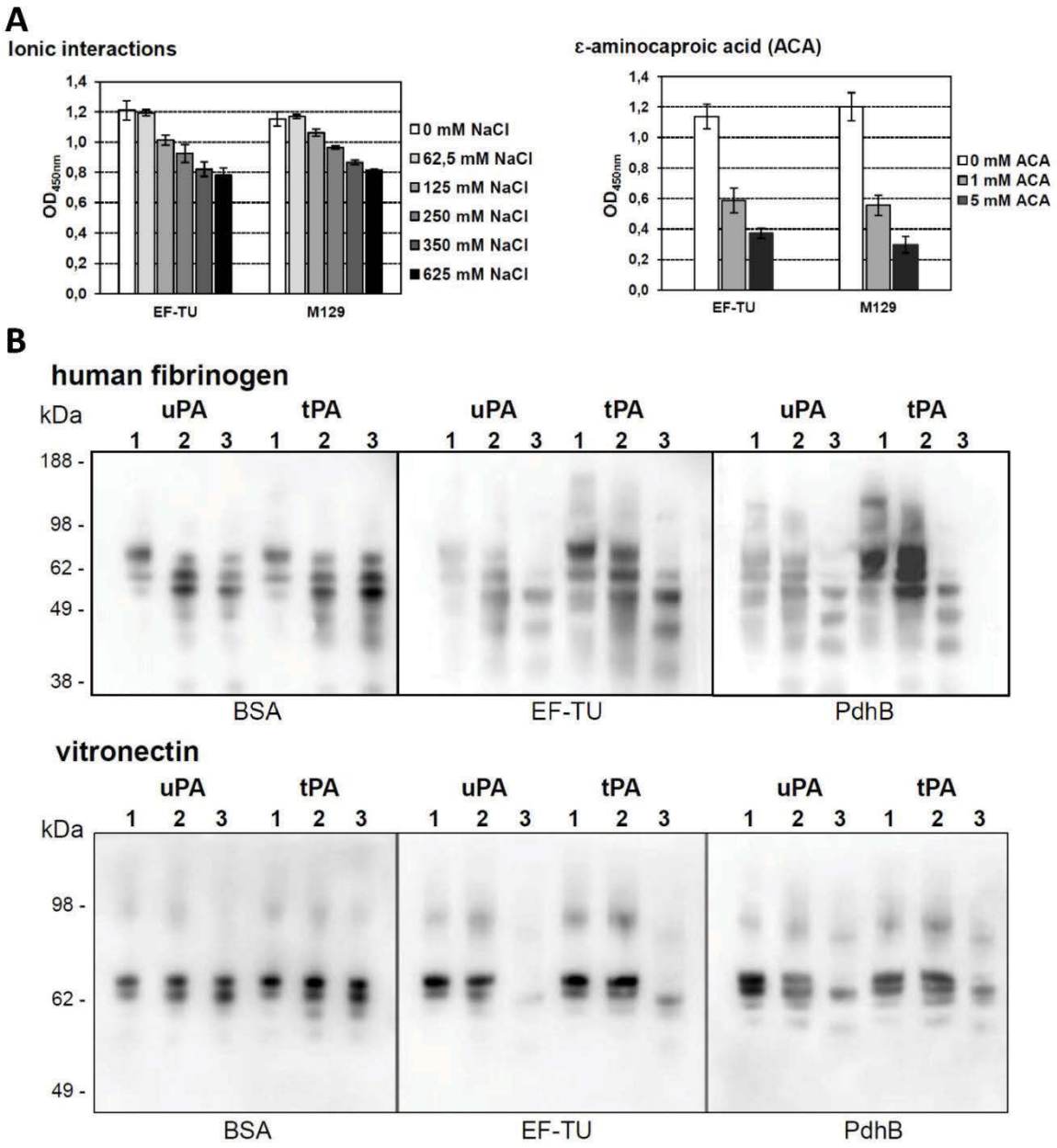


Figure 5.8: Influence of ions and lysine analog ACA on binding of rMpn$_{Ef\text{-}Tu}$ to plasminogen and degradation of human fibrinogen and vitronectin by activated plasminogen. A) Microtitre plate wells were coated with rMpn$_{Ef\text{-}Tu}$ and incubated with plasminogen and increasing concentrations of either NaCl ('ionic interactions') or ε-aminocaproic acid ('ACA'). Bound plasminogen was detected with anti-plasminogen antibodies. In control experiments *M. pneumoniae* cells were coated onto microtitre plates and incubated with plasminogen and increasing concentrations of either NaCl or ACA. Bars

represent standard deviation of eight replicates. B) Fibrinogen or vitronectin was mixed with either urinary plasminogen activator (uPA) or tissue plasminogen activator (tPA) and added to microtitre plates previously coated with $rMpn_{Ef\text{-}Tu}$ and plasminogen. Samples were separated by SDS-PAGE, blotted onto nitrocellulose membrane and probed with anti-vitronectin and anti-fibrinogen antisera. Lane 1 is at 0 hours, lane 2 is after over-night incubation without plasminogen, and lane 3 is after over-night incubation with plasminogen.

In the presence of plasminogen activators tPA and uPA, plasminogen bound to $rMpn_{Ef\text{-}Tu}$ is converted to plasmin and can degrade fibrinogen and vitronectin (Figure 5.8B). Collectively these studies highlight the widespread multifunctional capabilities of Ef-Tu and the cleavage fragments derived from it.

5.8 Discussion

Ef-Tu moonlights on the cell surface of *S. aureus*, *M. pneumoniae* and *M. hyopneumoniae*, three phylogenetically diverse, pathogenic bacteria that belong to the low G+C Firmicutes. Using a combination of microscale thermophoresis and microtitre plate binding assays, $rMpn_{Ef\text{-}Tu}$ was shown to bind strongly to heparin (K_D = 42.5 ± 1.5 nM), fetuin (K_D = 53 ± 14 nM), and actin (K_D = 19 ± 3 nM); as well as to laminin, plasminogen, vitronectin, lactoferrin, fibronectin, and fibrinogen. Plasminogen bound to $rMpn_{Ef\text{-}Tu}$ can be converted to plasmin in the presence of plasminogen activators tPA and uPA (Figure 5.8). This study extends these findings by showing that $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$, are targets of processing events on the cell surface of these bacterial pathogens but the biological significance of this warrants further investigation (see below). Molecules are not strictly confined to compartments in the bacterial cell and can perform novel functions at different cellular locations[188,195,340,367,394-396]. Much remains to be learnt about how proteins, especially those lacking signal motifs, localise on bacterial cell surfaces.

We sought to gain a better understanding of how Ef-Tu has evolved to be a multifunctional binding protein. $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$ all putatively bind heparin, each sharing the consensus heparin binding motif XBBBXXBX (sequence: DKRHYAHV) as well as a number of other heparin binding motifs (see Figure 5.1 and Table 5.2). It is notable that while this motif (DKRHYAHV) is conserved in the Ef-Tu from *M. pneumoniae*, *M. hyopneumoniae*

and *S. aureus* only part of the motif, with the sequence RHYAHV, is conserved in Ef-Tu from other bacterial sources. The addition of DK residues is predicted to impart a putative PPI site. Twelve putative heparin binding motifs identified in $Mpn_{Ef\text{-}Tu}$ (Table 5.2) were predicted to predominantly localise to non-essential, unconserved regions of the molecule that do not unduly influence its ability to function as an elongation factor. Short linear motifs (SLiMs) typically ranging from three to ten amino acids play crucial roles in mediating PPIs[397-399]. In eukaryotes, these motifs are typically located in intrinsically unstructured, disordered regions of proteins that impart plasticity and are reported to favour transient, low affinity and reversible interactions[397,400]. Notably, $Mpn_{Ef\text{-}Tu}$ formed strong interactions with fetuin, heparin, and actin suggesting that the accumulation of SLiMs may be sufficient to form high-affinity interactions.

Positively charged amino acids in SLiMs play a crucial role in interactions between proteins and highly sulphated glycosaminoglycans such as heparin[276], actin[401], plasminogen[401], DNA[402,403] and fibronectin[226,239,404]. Three SLiMs were identified enriched in positively charged amino acids (underlined) in different regions of $Mpn_{Ef\text{-}Tu}$, including sequences ^{37}A<u>K</u>EG<u>K</u>SAAT<u>R</u>Y^{47}, ^{183}P<u>K</u>WEA<u>K</u>I<u>H</u>D^{191}, and ^{248}L<u>R</u>PI<u>RK</u>A^{254}. Eight surface-exposed PPI sites, including three that reside within putative heparin binding motifs ^{2}A**<u>RE</u>**KFDRSKPHV^{13}, ^{73}**<u>D</u>**KRHYAHV^{80}, and ^{370}E**<u>K</u>**GSKFSIREGGRT^{383} were also identified in $Mpn_{Ef\text{-}Tu}$. It is notable that the lysine analog, ε-amino caproic acid was shown to be a potent inhibitor of interactions between $Mpn_{Ef\text{-}Tu}$ and plasminogen, and *M. pneumoniae* M129 whole cells and plasminogen, underscoring the important role played by positively charged amino acids in binding interactions with host molecules (Figure 5.8A). Overlapping SLiMs are frequently identified in multifunctional proteins[397,405]. In *M. hyopneumoniae*, the C-terminal sequence ^{1070}KKSSLKVKITVK^{1081} in the multifunctional cilium adhesin, P97 binds both heparin and fibronectin[226] and overlapping peptides from a region within phosphoglycerate kinase from group B streptococcus strain NCS13 with sequence ^{203}SKVSDKIGVIENLLEKADKV^{222} and ^{213}ENLLEKADKVLIGGGMTYTF^{232} bind both actin and plasminogen[401]. Similarly, SLiMs were identified enriched in positively charged amino acids in $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$. The accumulation of positively charged residues in SLiMs, possibly as a consequence of an A+T rich genome, facilitates binding to a wide range of host molecules in the low G+C Firmicutes. The data presesented here is consistent with the proposition that the

accumulation of surface-exposed SLiMs represents a mechanism to generate protein multifunctionality in bacterial proteins.

S. aureus[283] and *M. hyopneumoniae*[187,217,218,224,226,243,244] display cell surface, heparin binding proteins that are important to the pathogenic potential of these species. Interactions between heparin binding proteins and target receptors in host cell membrane allow microbes to colonise a wide range of niche sites, traverse tissue barriers and disseminate from their initial point of contact and form biofilms[406]. *S. aureus*[407,408], *M. pneumoniae*[114,160] and *M. hyopneumoniae* (our unpublished data) are all capable of forming biofilms. The extracellular matrix of *S. aureus* biofilms is derived from a mixture of eDNA and cytoplasmic proteins[**407,409-414**] and electrostatic interactions between cytoplasmic proteins and eDNA is thought to tether cells together in *S. aureus* and mixed-species biofilms[414]. In *S. aureus*, the addition of heparin increases biofilm production in a protein dependant manner which implies that heparin binding proteins are important for biofilm development[283]. Notably, Ef-Tu has been identified on the surface of *S. aureus* under biofilm inducing conditions[409]. These observations lend weight to the hypothesis that the accumulation of positively charged amino acids in SLiMS represents a powerful mechanism to promote PPIs that underpin essential biological processes such as the formation and maintenance of biofilms.

Bacterial pathogens including *Campylobacter jejuni*[239], *Mycoplasma gallisepticum*[231], and *Chlamydia trachomatis*[415] process molecules that are secreted to the cell surface. In *M. hyopneumoniae*, processing of cilium adhesin families has been reported extensively and cleavage motifs have been mapped[186,217,222,225]. Recently, lactate dehydrogenase was shown to be cleaved on the surface of *M. hyopneumoniae* generating fragments with putative multifunctional binding capabilities[227]. In *M. pneumoniae*, cleavage fragments of the major adhesin P1 and DnaK have been shown to comprise part of the cytoskeletal attachment organelle complex[247] and Mycoplasma derived lipoproteins are targets of processing events that release powerful immunomodulatory peptides[229,232-234]. These observations prompted a systems-wide, protein dimethyl labelling strategy to investigate protein processing. Numerous processing sites were identified and characterised in Ef-Tu derived from all three bacterial pathogens. Furthermore, surface biotinylation studies indicate $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$ were a target of multiple processing events of the

surfaces of *M. pneumoniae*, *M. hyopneumoniae* and *S. aureus*, respectively. This work strongly suggests that the accumulation of positively charged residues in the SLiMs found in Ef-Tu facilitates binding to a wide range of host molecules, and potentially to eDNA, and that protein cleavage events expand the functional complexity of proteins that moonlight on the cell surface. From this, it can be hypothesised that processing is a mechanism that has evolved to promote multifunctional behaviour more broadly and lends itself to the creation of novel binding sites in moonlighting proteins that retain a strict conformational structure needed to execute their canonical function.

Fifteen cleavage fragments of $Mpn_{Ef\text{-}Tu}$ were identified in this study of which eleven reside on the cell surface. Unlike $Mpn_{Ef\text{-}Tu}$, none of the fragments were retained in all six affinity chromatography columns, but five were identified in at least five affinity columns (fragments 5, 6, 7, 8, and 10 in Figure 5.2). Fragments 5, 8, and 10 were retained in columns coupled with: A549 surface proteins, fetuin, fibronectin, actin, and heparin. Fragments 6 and 7 were retained in columns coupled with: A549 surface proteins, fibronectin, actin, heparin, and plasminogen. Fragment 4 was identified in eluents from columns coupled with A549 surface proteins and heparin while Fragment 9 was identified in eluents from columns coupled with fetuin and actin (see Figure 5.2). These data indicate that retention of the fragments during affinity chromatography is dependent on the host molecule that is coupled to the agarose beads and the sequence of the Ef-Tu fragment. Further studies are needed to quantify the binding characteristics of fragments of Ef-Tu with host molecules.

Cleavage fragments of cytosolic proteins that moonlight on the cell surface add another layer of complexity to the concept of multifunctional proteins. The study presented here shows that processing exposes SLiMs that would otherwise be inaccessible for interactions with potential binding partners. Recently, a peptidome study of a protease deficient strain of *Lactococcus lactis* identified 1800 distinct peptides fragments in spent growth medium that were derived from proteolytic activity targeting both surface accessible and cytosolically derived proteins[416]. Similar studies by the same group indicated that surface accessible proteins in other Firmicute species including *Listeria monocytogenes*, *Enterococcus faecalis* and *Streptococcus thermophilus* were also targeted by complex processing events[416]. Processing events play an important role in the maturation of key

adhesin families in pathogenic *Mycoplasma* spp.[186,187,216-227,230,231]. The findings in this study show that surface proteolysis is critical in shaping the surface proteome more broadly and that processing represents a novel and under recognised mechanism to expand protein function.

In summary, Ef-Tu moonlights on the surface of bacteria where it is a target of proteolytic processing events. Computational analysis of fragments of $Mpn_{Ef\text{-}Tu}$ suggest that they are inherently more disordered and display putative PPI sites that are inaccessible in the parent molecule, generating unprecedented functional diversity on the cell surface. Further studies, using systems-wide methodologies, are needed to determine how processing generates biologically important effector molecules and if protein processing is fundamental to the expansion of protein function in bacteria belonging to different phylogenetic clades.

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Chapter 6.

The P1 adhesin in *Mycoplasma pneumoniae* is extensively processed and binds multiple host molecules

i3 ithreeinstitute

6.1 Preface

The P1 adhesin of *M. pneumoniae* is the major adhesin used by this pathogen for adherence and is always identified by *M. pneumoniae* infected patient sera. It has a role in immune evasion and cellular motility. The manuscript presented for Chapter 6 reports that the P1 adhesin is subject to significant proteolytic processing and generates a C-terminal peptide that plays expands on the binding targets of the respiratory epithelium. Due to the highly immunogenic nature of P1, it is possible that these fragments function as immune decoy targets. The results illustrated here broadens the significant role that P1 performs during the interaction of *M. pneumoniae* surface with the respiratory epithelium.

Chapter 6 was written as a manuscript that will be submitted in the near future.

6.2 Author contribution

Author	Contribution
Michael Widjaja	Performed surface proteome experiments, the 'Bait and Prey' chromatography, P1 immunoblots, microscale thermophoresis, and the bioinformatic analysis of P1. Analysed all data generated in this study.
Iain J. Berry	Performed dimethyl labelling for *M. pneumoniae* and assisted in analysing dimethyl labelling data.
Matthew P. Padula	Oversaw the acquisition of mass spectrometry data.
Steven P. Djordjevic	Initiated and designed the study and together with Michael Widjaja interpreted the data. Secured fundingfor the project.

6.3 Abstract

Mycoplasma pneumoniae is a genome reduced pathogen that is responsible for community acquired pneumonia. It has an elongated cellular extension termed the attachment organelle that is responsible for motility and adherence to the respiratory epithelium. The major adhesin, P1, localises to the tip of the attachment organelle forming a complex with P40 and P90 from Mpn142, and other adhesin molecules. Here we show that P1 is extensively processed. LC-MS/MS analysis of protein gel spots derived from whole cell lysates of *M. pneumoniae* separated by 2D-SDS PAGE and eluents derived from 'Bait and Prey' affinity chromatography using a range of host molecules as bait, 22 fragments of the P1 adhesin were identified. Analysis of N-terminome data identified precise location of seventeen cleavage sites within the adhesin. Immunoblots generated using a bank of rabbit polyclonal antisera raised to different recombinant regions of P1 identified the full length P1 adhesin and many smaller proteoforms, and provided another layer of evidence in support of the hypothesis that this adhesin is extensively processed. One of the cleavage events removed the C-terminal 30 amino acids of P1, an unusual region enriched in lysine and proline residues. The carboxyl tail of P1 was synthesised chemically and coupled to agarose resin. The peptide selectively recovered cytoskeletal intermediate filament proteins cytokeratin 7, cytokeratin 8, cytokeratin 18, and vimentin from A549 cell lysates. Microscale thermophoresis was able to quantify the binding affinity between the synthetic peptide and heparin, fetuin, and plasminogen. Results from this study suggests that the P1 adhesin is a multifunctional, modular protein that is targeted by proteases that release fragments of P1 that retain an ability to interact with diverse host molecules.

6.4 Introduction

The attachment organelle is an essential feature of the *M. pneumoniae* cell. The organelle is important for localising two major adhesins P1 and P30 as well as accessory proteins derived from Mpn142 (P40 and P90) that are also needed for cell adherence[27,117,132,154,161,165,172,317]. The P1 and P30 adhesins are encoded by the *mpn141*[163] and *p30* (*mpn453*)[164,304] genes, respectively. *M. pneumoniae* has a third adhesin, P116 (encoded by *mpn213*), but the exact cellular location is yet to be identified[155,417]. Together these adhesins and various moonlighting surface proteins allow *M. pneumoniae* to bind a range of different host cell surface receptors such as sialylated host molecules[271], oligosaccharides[145], glycolipids[144], glycoproteins[146], fibronectin[205,210,213], fibrinogen[208,210,213], plasminogen[209,213-215], lactoferrin[210,213], laminin[210,213], and vitronectin[210,213].

The gene *mpn141* encoding the major adhesin P1 is located in the same operon along with *mpn140* and *mpn142*[170,306]. *mpn140* encodes for a 28 kDa phosphoesterase[44,166,306,307] and *mpn142* generates a 130 kDa product (Mpn142) that is cleaved into two fragments (P40 and P90) shortly after translation[166,167]. It was recently reported that both P40 and P90 undergo multiple cleavage events on the *M. pneumoniae* cell surface[230]. P1, P40, and P90 are related at levels of transcription, translation, stability and are localised within very close proximity on the attachment organelle[170,172,317]. Using a reversible cross-linking agent, DTSSP (3,3'-dithiobis(sulfosuccinimidylpropionate)), P1 was identified to form a complex with P30, P40, and P90[169]. P1 also forms a 480 kDa globular complex with P90 which is suggested to drive motility in *M. pneumoniae*[301]. Further cross-linking studies with paraformaldehyde, a more permanent cross-linker, found P1 forms more complexes with a smaller truncated P1 fragment, P30, P40, P90, P65, DnaK, a truncated DnaK fragment, pyruvate dehydrogenase α subunit (Pdh-A), HMW1, and HMW3[247].

Both *mpn141* and *mpn142* contain strings of repetitive DNA sequences named RepMP elements[110-112]. Within *mpn141*, there are two of these RepMP elements (RepMP 2/3 and RepMP4) that are homologous to another ten and eight RepMP elements in the *M. pneumoniae* genome, respectively[33,111]. RepMP2/3 and RepMp4 span the amino acid positions 778 – 1333 and 46 – 464 of P1, respectively[301]. Together, all the RepMP elements compromise 8% of the *M. pneumoniae* genome highlighting its significance[33]. It is suggested that these regions can undergo genomic recombination (promoted by a RecA

homolog (MPN490)[418]) with the other homologous RepMP elements to generate antigenic variation[27-32]. It is the difference in these RepMP elements in *mpn141* that distinguish the two phylogenetic groups of *M. pneumoniae* known as Type 1 and Type 2 variants[28,32]. Epidemics of *M. pneumoniae* seem to alternate predominantly between the two types[419,420]. These two groups are quite diverse in that they differ in the formation of biofilms[114], hemadsorption inhibitory activity[115] and CARDS toxin production[38,116]. To add to the mechanisms for immune evasion, *M. pneumoniae* is able to form biofilms, as seen in other bacteria, generating resistance against both the host immune system and antibiotics[160,421]. It was reported that both the P1 adhesin and the availability of sialic acid residues are required for biofilm formation[160].

The P1 adhesin is highly immunogenic and often detected in *M. pneumoniae* infected patient sera[159,422,423]. Antibodies raised against P1 reduce adherence of *M. pneumoniae* to abiotic and host cell surfaces[155,156,159-162] and *M. pneumoniae* mutants that lack this adhesin are unable to adhere, highlighting the necessity for the P1 adhesin in adherence[151,156-159]. Translocation to the surface and stability of P1 requires a handful of partner proteins called accessory proteins that work together in a very precise and specific manner within the attachment organelle[118]. Though there are a number of accessory proteins, P40, P90, and HMW1 play a significant role in P1 localisation and stability[27,302,303,424]. There have been several attempts at creating an *M. pneumoniae* vaccine associated with P1. The first is a chimeric recombinant protein which contains regions from all three adhesins (P1, P30 and P116) and is currently under investigation[96]. Zhu et al. generated a fusion protein that improved protective efficacy in mice against *M. pneumoniae* by combining the carboxyl terminal of P1 with an *E. coli* toxin subunit[95] compared to the P1-carboxyl terminal only recombinant. Lastly is another recombinant created by Schurwanz et al. outlined below[93].

Several studies have shown that the carboxyl half of P1 is highly immunogenic and crucial for its function as an adhesin[93,155,425-430]. To identify other significant regions in P1, Schurwanz et al. generated recombinant proteins to fifteen different regions throughout P1, and tested these recombinants against fourteen patient sera[93]. This study found three (one N-terminal and two C-terminal regions) of the fifteen recombinants to be extremely antigenic, with greater than 90% patient reactivity[93]. Schurwanz et al. also found that

antibodies generated to one of the C-terminal regions significantly reduced binding of *M. pneumoniae* to the human cell lines HBEC (Primary bronchial epithelial cells), MRC-5 (fetal lung fibroblasts), and HeLa (cervical carcinomas)[93]. These data informed the creation of a chimeric recombinant protein which included this carboxyl region of P1 and another region in the P30 adhesin. Antibodies raised against this chimeric protein reduced *M. pneumoniae* adherence to human bronchial epithelial cells by more than 95%[93] and also successfully reduced *M. pneumoniae* colonisation in animal models[94].

P1, has been shown to undergo post-translational modifications such as phosphorylation and acetylation[35,50,431]. Post-translational proteolytic processing has been documented to remove the signal peptide of P1 which has been mapped to be the carboxyl side of threonine[59163,245,246].The study presented here sought to determine if P1 is processed on the surface of *M. pneumoniae*. Tryptic peptides that mapped to different regions within P1 were frequently encountered when characterising size fractionated eluents generated during affinity chromatography using different host molecules as bait. These peptides were recovered and a detailed map was generated. The identity of cleavage sites from N-terminome studies were combined with semi-trptic peptide analyses to map precise cleavage sites in P1. Fragments between cleavage sites correlated with size fractionated fragments recovered from affinity 'Bait and Prey' experiments. A subset of these fragments were also identified as individual protein spots resolved by2D-SDS PAGE of *M. pneumoniae* whole cell lysates. Finally, western blots using serum raised against fifteen different regions of P1 (Sourced by Dr. Roger Dumke[93]) were performed to characterise fragments of P1. These independently acquired, but complimentary datasets enabled a rigorous assessment of cleavage events in the P1 adhesin.

6.5 Methods and Materials

6.5.1 Strains

M. pneumoniae (M129 strain, ATCC 29342) cells were cultured as described previously[250]. Cells were grown in modified Hayflick's medium in tissue culture flasks at 37°C. Human lung carcinoma (A549, ATCC CCL-185) cells were cultured in RPMI 1640 medium (Invitrogen) supplemented with 10% heat inactivated fetal bovine serum. Cells were grown in tissue culture flasks at 37°C with 5% CO_2.

6.5.2 Cell preparation for one dimensional- and two dimensional-SDS polyacrylamide gel electrophoresis

M. pneumoniae cells were harvested as described previously[187]. In brief, cells were lysed with sonication in 7 M urea, 2 M thiourea, 40 mM Tris-HCl, 1% (w/v) C7BzO detergent (Sigma) after washing with PBS. Proteins were reduced and alkylated with 5 mM tributylphosphine and 20 mM acrylamide monomers before precipitation with acetone. Protein was resuspended in 7 M urea, 2 M thiourea, 40 mM Tris-HCl, 1% (w/v) C7BzO for 1D- and 2D-SDS PAGE.

Gel electrophoresis was performed as described previously[221,224]. Approximately 80 µg and 250 µg of protein was used for 1D- and 2D-SDS PAGE, respectively. Gels were fixed and stained by either Flamingo fluorescent gel stain (Bio-Rad) or Coomassie Blue G-250 (Sigma).

In-gel trypsin digestion was performed as described previously[186] for mass spectrometry analysis. Gel pieces were excised, destained, dehydrated and then incubated with trypsin Gold MS grade (Promega) in 100 mM NH_4HCO_3. Tryptic peptides were extracted by sonication and stored in 4°C until needed for mass spectrometry.

6.5.3 Liquid chromatography tandem mass spectrometry (LC-MS/MS) and data analysis

LC-MS/MS was performed as described previously[224]. In brief, 5 µg of peptides in 15 µl was loaded into an Eksigent AS-1 autosampler connected to a Tempo nanoLC system (Eksigent, Livermore, CA, USA) and washed onto a PicoFrit column (75 µm × 150 mm) packed with Magic C18AQ resin (Michrom Biosciences, CA). Peptides were eluted from the column into the source of a QSTAR Elite hybrid quadrupole-time-of-flight mass spectrometer (Sciex, Redwood, CA, USA).

Files generated from LC-MS/MS were searched against the MSPnr100 database[251] with the following parameters: Fixed modifications: none; Variable modifications: propionamide, oxidized methionine, deamidation; Enzyme: semi-trypsin; Number of allowed missed cleavages: 3; Peptide mass tolerance: 100 ppm; MS/MS mass tolerance: 0.2 Da; Charge state: 2+, 3+ and 4+. For samples collected from the 'Surface proteome analysis of *M. pneumoniae* (Biotinylation)' and 'Affinity chromatography host binding *M. pneumoniae* complexes (A549)' listed below, variable modifications also included NHS-LC-Biotin (K)

and NHS-LC-Biotin (N-term). 'Affinity chromatography host binding *M. pneumoniae* complexes (A549)' was also searched against *'homo sapiens'* entries in MSPnr100 to identify biotinylated surface A549 proteins.

6.5.4 Surface proteome analysis of *M. pneumoniae*

Biotinylation of the *M. pneumoniae* cells was performed as described previously[230]. The biotinylation reaction was allowed to proceed for 30 seconds on ice. Biotinylated surface proteins were confirmed with western blots using ExtrAvidin-HRP (Sigma).

Trypsin shaving of *M. pneumoniae* cells was carried out as described previously[187] with modifications. Trypsin was added to adherent *M. pneumoniae* cells. Shaving was for 5 minutes at 37°C and released peptides were trypsin digested a second time before analysis by LC-MS/MS.

6.5.5 Affinity chromatography host binding *M. pneumoniae* complexes

'Bait' host proteins used for affinity chromatography include fibronectin (Code: 341635) and plasminogen (Code: 528175) from human plasma was supplied by Merck Millipore. Bovine actin (Code: A3653) and fetuin (Code: F3004) was supplied by Sigma.

Affinity chromatography using host proteins bound to Avidin Agarose (Pierce) as 'Bait' was performed as described previously[230]. *M. pneumoniae* cells were lysed in 1% (w/v) C7BzO (Sigma-Aldrich) in PBS (pH 7.8) to obtain native complexes. The native complex cell lysate was incubated with host proteins bound to Avidin Agarose ('Bait'). This mixture was washed with PBS and host protein binding complexes ('Prey') were eluted 7 M urea, 2 M thiourea, 40 mM Tris-HCl and 1% (w/v) C7BzO. Elutions were separated by 1D-SDS PAGE and proteins were identified by LC-MS/MS as described above.

Affinity chromatography using human lung carcinoma (A549) surface proteins as 'Bait' was performed as described previously[230]. A549 cells were biotinylated, lysed and bound to Avidin Agarose ('Bait'). As above, this mixture was incubated with native *M. pneumoniae* complexes followed by washes and eluents to obtain a fraction of A549 binding complexes ('Prey').

Affinity chromatography using heparin HiTrap columns (GE Healthcare) was performed as described previously[230]. *M. pneumoniae* cells were lysed in 10 mM sodium phosphate,

0.1% Triton TX-100 (pH 7.0) to obtain native complexes. Approximately 300 µg of soluble complexes was loaded onto a HiTrap Heparin HP column (GE Healthcare). The column was washed with 10 mM sodium phosphate (pH 7.0) and heparin binding complexes were sequentially eluted in increasing concentrations of sodium chloride (pH 7.0).

6.5.6 Affinity chromatography of P1 C-terminal tail binding complexes

The C-terminal sequence of the P1 adhesin (^{1613}PGAPKPPVQPPKKPA^{1627}) was synthesised with an N-terminal biotin tag by Chempeptide Limited (China). Affinity chromatography was performed similar to the section above. In brief, 1 mg of the peptide was added to Avidin Agarose beads for 16 h at 4 °C. The beads were washed four times (5 ml per wash) with PBS before being incubated with native A549 cell lysates (harvested in 1% w/v C7BzO in PBS) for 16 h at 4 °C. Non-binding proteins were washed from the column with four washes (5 ml per wash) of PBS and protein complexes with an affinity to the peptide were eluted from the column with 7 M urea, 2 M thiourea, 40 mM Tris-HCl and 1% (w/v) C7BzO (4 times of 2 ml). Eluents were concentrated with a Macrosep® 3 kDa cutoff centrifugal device (Pall), precipitated with acetone and separated by 1D-SDS PAGE. The whole lane was divided into sections, in-gel digested with trypsin and analysed by LC-MS/MS as described above.

6.5.7 Dimethyl labelling of *M. pneumoniae* and LC-MS/MS analysis

Dimethyl labelling of *M. pneumoniae* proteins was carried out as described previously[227,230]. 1 mg of *M. pneumoniae* protein was labelled in 40 mM formaldehyde (Ultrapure grade, Polysciences), 20 mM sodium cyanoborohydride, 100 mM Hepes (pH 6.7) for 4 hours at 37°C. The reaction was quenched with 100 mM ammonium bicarbonate, precipitated in acetone:methanol (8:1) and digested with trypsin.

Peptides were analysed using both the Sciex 5600 and Thermo Scientific Q Exactive™ mass spectrometers. The methods, protocols and parameters used are described in[230] (Chapter IV).

6.5.8 Immunoblot of *M. pneumoniae* cell lysates using Anti-P1 serum

60 µg of *M. pneumoniae* cell lysate proteins were separated on 1D-SDS PAGE as described above. Proteins were transferred to PVDF (polyvinylidene fluoride) membranes using a semidry method[432]. Membranes were blocked with 5% (w/v) skim milk powder in PBS, 0.1% (v/v) Tween 20 (PBS-Tween) for 1 hour at 25°C. Membranes were then separately probed with sera raised against different regions of the P1 adhesin (sourced by Dr. Roger Dumke[93]) for 1.5 hours at 25°C in PBS-Tween. Membranes were washed three times over 30 minutes before being probed a second time in peroxidase-conjugated anti-guinea pig antibodies (1:3000, Sigma) for 1 hour at 25°C in PBS-Tween. Membranes were washed again three times over 30 minutes and developed with DAB tablets (3,3'-Diaminobenzidine, Sigma).

6.5.9 Bioinformatic analysis of the P1 adhesin

Bioinformatic predictions and analysis was performed as described previously[230]. The bioinformatic tools used were: ProtParam[268], TMpred[258], PONDR® (VSL2 predictor)[313] and ScanProsite[315]. Predicted glycosaminoglycan binding sites searched in ScanProsite included binding sites for heparin (X-[HRK]-[HRK]-X-[HRK]-X motif)[276], heparin sulfate (X-[HRK]-X-[HRK]-[HRK]-X)[316], or clusters of basic amino acid residues (X-[HRK]-X(0,2)-[HRK]-X(0,2)-[HRK]-X and X-[HRK]-X(1,3)-[HRK]-X(1,3)-[HRK]-X).

6.6 Results

6.6.1 Bioinformatic analysis of the P1 adhesin

Using TmPred, P1 is predicted to contain six transmembrane regions (yellow 'TmD' boxes in Figure6.1). Using ScanProsite, thirteen putative glycosaminoglycan binding regions (blue 'Hep' boxes in Figure 6.1) were identified which consists of: heparin ('Hep'), heparan sulfate ('HepS'), and clusters of basic amino acid residues ('HepB'). Seven disordered regions that span at least 30 amino acids were identified in P1 based on output from PONDR® (purple boxes within a grey bar in Figure 6.1). Including the signal sequence, the P1 adhesin has a predicted mass of 176.3 kDa and a pI of 8.53. We have observed that adhesin molecules from *Mollicute* spp. are often modular in their design and contain regions enriched in acidic amino acids (D and E) while other regions are enriched in positively charged, basic residues (K, R, and H)[186,187,219-227,244]. Regions within P1 that were enriched in acidic and basic amino acids were identified using ProtParam.

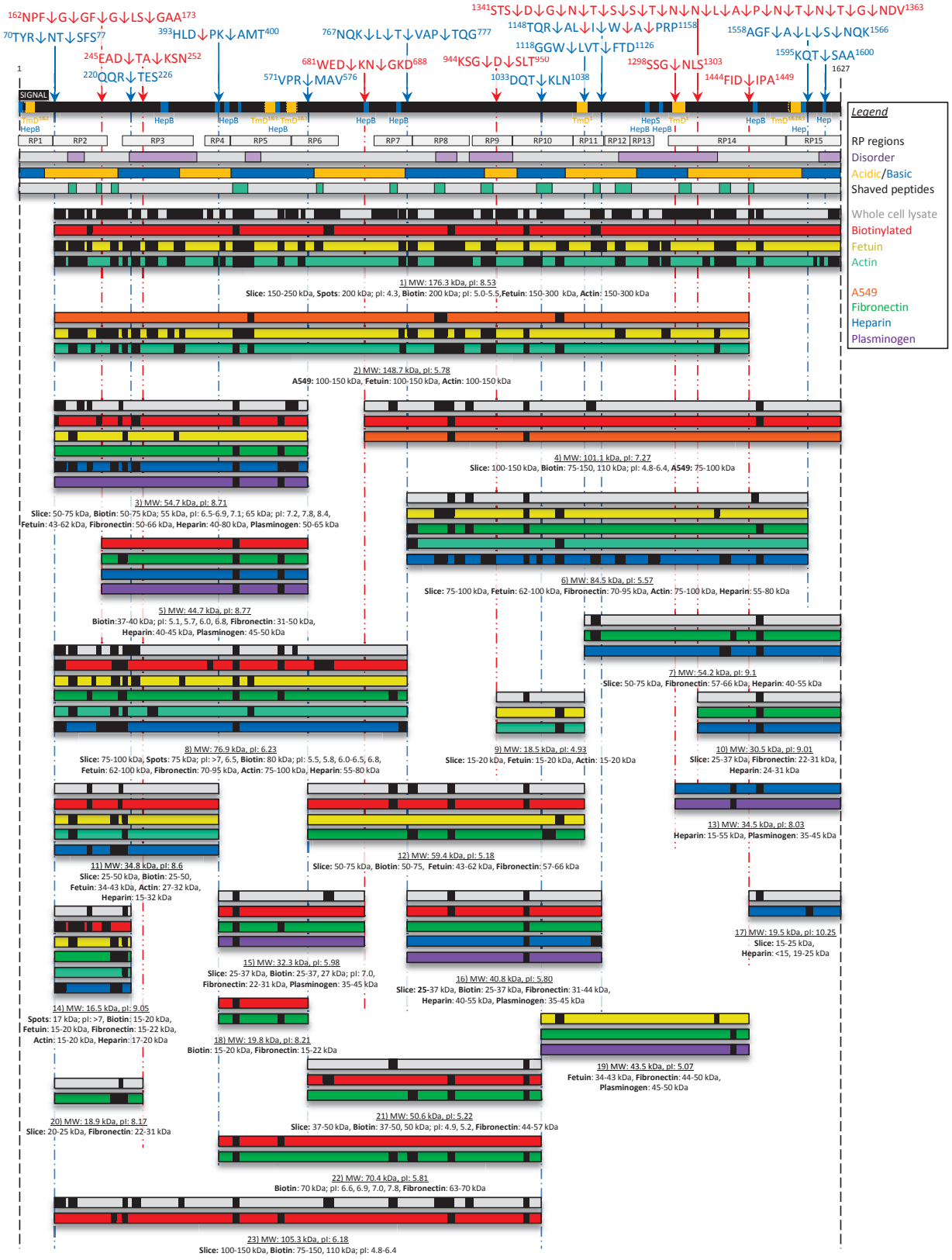

162NPF↓G↓GF↓G↓LS↓GAA173
70TYR↓NT↓SFS77
393HLD↓PK↓AMT400
245EAD↓TA↓KSN252
220QQR↓TES226
767NQK↓L↓T↓VAP↓TQG777
681WED↓KN↓GKD688
571VPR↓MAV576
1341STS↓D↓G↓N↓T↓S↓S↓T↓N↓N↓L↓A↓P↓N↓T↓N↓T↓G↓NDV1363
1148TQR↓AL↓I↓W↓A↓PRP1158
1118GGW↓LVT↓FTD1126
944KSG↓D↓SLT950
1033DQT↓KLN1038
1298SSG↓NLS1303
1444FID↓IPA1449
1558AGF↓A↓L↓S↓NQK1566
1595KQT↓SAA1600
1
1627
SIGNAL
RP1
RP2
RP3
RP4
RP5
RP6
RP7
RP8
RP9
RP10
RP11
RP12
RP13
RP14
RP15
Legend
RP regions
Disorder
Acidic/Basic
Shaved peptides
Whole cell lysate
Biotinylated
Fetuin
Actin
A549
Fibronectin
Heparin
Plasminogen
1) MW: 176.3 kDa, pI: 8.53
Slice: 150-250 kDa, Spots: 200 kDa; pI: 4.3, Biotin: 200 kDa; pI: 5.0-5.5,Fetuin: 150-300 kDa, Actin: 150-300 kDa
2) MW: 148.7 kDa, pI: 5.78
A549: 100-150 kDa, Fetuin: 100-150 kDa, Actin: 100-150 kDa
3) MW: 54.7 kDa, pI: 8.71
Slice: 50-75 kDa, Biotin: 50-75 kDa; 55 kDa, pI: 6.5-6.9, 7.1; 65 kDa; pI: 7.2, 7.8, 8.4, Fetuin: 43-62 kDa, Fibronectin: 50-66 kDa, Heparin: 40-80 kDa, Plasminogen: 50-65 kDa
4) MW: 101.1 kDa, pI: 7.27
Slice: 100-150 kDa, Biotin: 75-150, 110 kDa; pI: 4.8-6.4, A549: 75-100 kDa
5) MW: 44.7 kDa, pI: 8.77
Biotin:37-40 kDa; pI: 5.1, 5.7, 6.0, 6.8, Fibronectin: 31-50 kDa, Heparin: 40-45 kDa, Plasminogen: 45-50 kDa
6) MW: 84.5 kDa, pI: 5.57
Slice: 75-100 kDa, Fetuin: 62-100 kDa, Fibronectin: 70-95 kDa, Actin: 75-100 kDa, Heparin: 55-80 kDa
7) MW: 54.2 kDa, pI: 9.1
Slice: 50-75 kDa, Fibronectin: 57-66 kDa, Heparin: 40-55 kDa
8) MW: 76.9 kDa, pI: 6.23
Slice: 75-100 kDa, Spots: 75 kDa; pI: >7, 6.5, Biotin: 80 kDa; pI: 5.5, 5.8, 6.0-6.5, 6.8, Fetuin: 62-100 kDa, Fibronectin: 70-95 kDa, Actin: 75-100 kDa, Heparin: 55-80 kDa
9) MW: 18.5 kDa, pI: 4.93
Slice: 15-20 kDa, Fetuin: 15-20 kDa, Actin: 15-20 kDa
10) MW: 30.5 kDa, pI: 9.01
Slice: 25-37 kDa, Fibronectin: 22-31 kDa, Heparin: 24-31 kDa
11) MW: 34.8 kDa, pI: 8.6
Slice: 25-50 kDa, Biotin: 25-50, Fetuin: 34-43 kDa, Actin: 27-32 kDa, Heparin: 15-32 kDa
12) MW: 59.4 kDa, pI: 5.18
Slice: 50-75 kDa, Biotin: 50-75, Fetuin: 43-62 kDa, Fibronectin: 57-66 kDa
13) MW: 34.5 kDa, pI: 8.03
Heparin: 15-55 kDa, Plasminogen: 35-45 kDa
14) MW: 16.5 kDa, pI: 9.05
Spots: 17 kDa; pI: >7, Biotin: 15-20 kDa, Fetuin: 15-20 kDa, Fibronectin: 15-22 kDa, Actin: 15-20 kDa, Heparin: 17-20 kDa
15) MW: 32.3 kDa, pI: 5.98
Slice: 25-37 kDa, Biotin: 25-37, 27 kDa; pI: 7.0, Fibronectin: 22-31 kDa, Plasminogen: 35-45 kDa
16) MW: 40.8 kDa, pI: 5.80
Slice: 25-37 kDa, Biotin: 25-37 kDa, Fibronectin: 31-44 kDa, Heparin: 40-55 kDa, Plasminogen: 35-45 kDa
17) MW: 19.5 kDa, pI: 10.25
Slice: 15-25 kDa, Heparin: <15, 19-25 kDa
18) MW: 19.8 kDa, pI: 8.21
Biotin: 15-20 kDa, Fibronectin: 15-22 kDa
19) MW: 43.5 kDa, pI: 5.07
Fetuin: 34-43 kDa, Fibronectin: 44-50 kDa, Plasminogen: 45-50 kDa
20) MW: 18.9 kDa, pI: 8.17
Slice: 20-25 kDa, Fibronectin: 22-31 kDa
21) MW: 50.6 kDa, pI: 5.22
Slice: 37-50 kDa, Biotin: 37-50, 50 kDa; pI: 4.9, 5.2, Fibronectin: 44-57 kDa
22) MW: 70.4 kDa, pI: 5.81
Biotin: 70 kDa; pI: 6.6, 6.9, 7.0, 7.8, Fibronectin: 63-70 kDa
23) MW: 105.3 kDa, pI: 6.18
Slice: 100-150 kDa, Biotin: 75-150, 110 kDa; pI: 4.8-6.4

Figure 6.1: Cleavage map of the P1 adhesin. The black bar represents the full length P1 adhesin (1627 amino acids) with cleavage products underneath. Arrows above the black bar indicate main cleavage sites within P1 with red and blue indicating semi-tryptic and dimethyl labelled sites, respectively. The black 'SIGNAL' box is the signal peptide spanning amino acid positions 1 – 59 identified in[246]. The 15 recombinants spanning different regions of P1 that were generated in[93] are depicted below the black bar as grey 'RP' boxes. Transmembrane segments (empty box with broken yellow outline) predicted by previous studies (TmD^2 boxes[301], and TmD^3 boxes[159]) have been included within the black bar. Transmembrane domains (yellow 'TmD^1' boxes) were predicted with TMpred[258]. Motifs were searched using ScanProsite[315] to identify putative glycasomainoglycan binding sites such as for heparin (blue 'Hep' boxes), heparan sulfate (blue 'HepS' boxes), and clusters of basic amino acid residues (blue 'HepB' boxes). Regions of disorder (purple boxes) were predicted using PONDR® (VSL2 predictor)[313] and ProtParam[268] was used to predict acidic and basic domains within P1 (Yellow and blue bars under disordered regions), and the sizes (in kDa) and pI of all P1 fragments seen as text underneath the fragments. A legend has been included as a reference for the coloured bars. Peptides released and identified by LC-MS/MS in surface shaving experiments are seen as the light green bars indicate surface exposed regions. Tryptic peptides (black boxes in coloured bars) identified by LC-MS/MS of the full length P1 adhesin and fragments identified in this study are from experiments of 1D- and 2D-PAGE cell lysate digestion (grey bars); surface biotinylation (red bars); and 'Bait and Prey' affinity chromatography of A549 surface proteins (orange), fetuin (yellow), fibronectin (green), actin (light green), heparin (blue), and plasminogen (purple).

6.6.2 The P1 adhesin is extensively processed on the cell surface

By combining LC-MS/MS data derived from multiple sources where mass context was retained, including 1D- and 2D-SDS PAGE of *M. pneumoniae* cell lysates, surface biotinylation, and 'Bait and Prey' affinity chromatography experiments, the P1 adhesin was observed to be processed extensively generating fragments that may interact with different host molecules. From these experiments, a total of 22 fragments of P1 plus the full length protein were identified ranging from 17 to 149 kDa in size (Figure 6.1). The full length (labelled fragment 1) and an additional 16 fragments were observed from SDS-PAGE of *M. pneumoniae* whole cell lysates (grey bars in Figure 6.1) that match the molecular weight predicted by ProtParam[268]. To add to this, two fragments (fragments 8 and 14) observed in 2D-SDS PAGE gel spots of *M. pneumoniae* cell lysates also had the same pI predicted by ProtParam. Trypsin shaving of the *M. pneumoniae* cell surface released trypsin accessible peptides (green boxes within a grey bar in Figure 6.1) that span most y of the adhesin indicating that P1 is exposed on the cell surface. LC-MS/MS analysis of of size-fractionated biotinylated proteins that were first enriched using avidin chromatography, 13 fragments of the P1 adhesin plus the full length protein were identified (red bars in Figure 6.1; fragments: 1, 3, 4, 5, 8, 11, 12, 14, 15, 16, 18, 21, 22, and 23). In addition to these 13 fragments, several more fragments of P1 were identified when size-fractionated eluents from 'Bait and Prey' affinity columns were analysed by LC-MS/MS. Two fragments of P1 with masses of 149 kDa and 101 kDa (fragments 2 and 4, orange bars in Figure 6.1) were identified from columns coupled with biotinylated A549 surface protein complexes. Eluents derived from columns coupled with fetuin and actin were particularly useful for identifying fragments of P1. Fragments: 1, 2, 3, 6, 8, 9, 11, 12, 14, and 19 (yellow bars in Figure 6.1) were recovered from columns coupled with fetuin, and fragments: 1, 2, 6, 8, 9, 11, 14, and 19 were recovered from columns coupled with actin (light green bars in Figure 6.1). Fragments: 3, 5, 13, 15, 16, and 19 were identified from columns coupled with plasminogen (purple bars in Figure 6.1). For the eleven fragments identified from heparin chromatography (blue bars in Figure 6.1; fragments: 3, 5, 6, 7, 8, 10, 11, 13, 14, 16, and 17), two (fragment 14 and 16) did not contain any of the predicted glycosaminoglycan binding motifs identified with ScanProsite. Fragments: 3, 5, 6, 7, 8, 10, 12, 14, 15, 16, 18, 19, 20, 21, and 22 were identified in eluents from columns coupled with fibronectin (green bars in Figure 6.1).

A global *M. pneumoniae* dimethyl labelling approach was used to identify internal neo-N termini. Ten cleavage sites were identified in P1 (Table 6.1, blue arrows in Figure 6.1). Semi-tryptic peptides, defined as peptides with only one tryptic end (Table 6.1, red arrows in Figure 6.1) were also identified, implying seven additional cleavage sites in P1. Four distinct sites in P1 showed evidence of aminopeptidase activity (Figure 6.1 and Table 6.1): ^{162}NPF↓G↓GF↓G↓LS↓GAA^{173} (cleavage site 2), ^{767}NQK↓L↓T↓VAP↓TQG^{777} (cleavage site 8), ^{1148}TQR↓AL↓I↓W↓A↓PRP^{1158} (cleavage site 12), and ^{1558}AGF↓A↓L↓S↓NQK^{1566} (cleavage site 16). A large predicted disorder region spanning 196 amino acids near the carboxyl terminal of P1 represents a fifth site for high cleavage activity with 18 events residing between amino acid positions 1343 – 1361 (cleavage site 14 in Table 6.1; sequence: ^{1341}STS↓D↓G↓N↓T↓S↓S↓T↓N↓N↓L↓A↓P↓N↓T↓N↓T↓G↓NDV^{1363}).

Table 6.1: N-terminal peptides in the P1 adhesin identified by LC-MS/MS from dimethyl labelling *M. pneumoniae* cells.

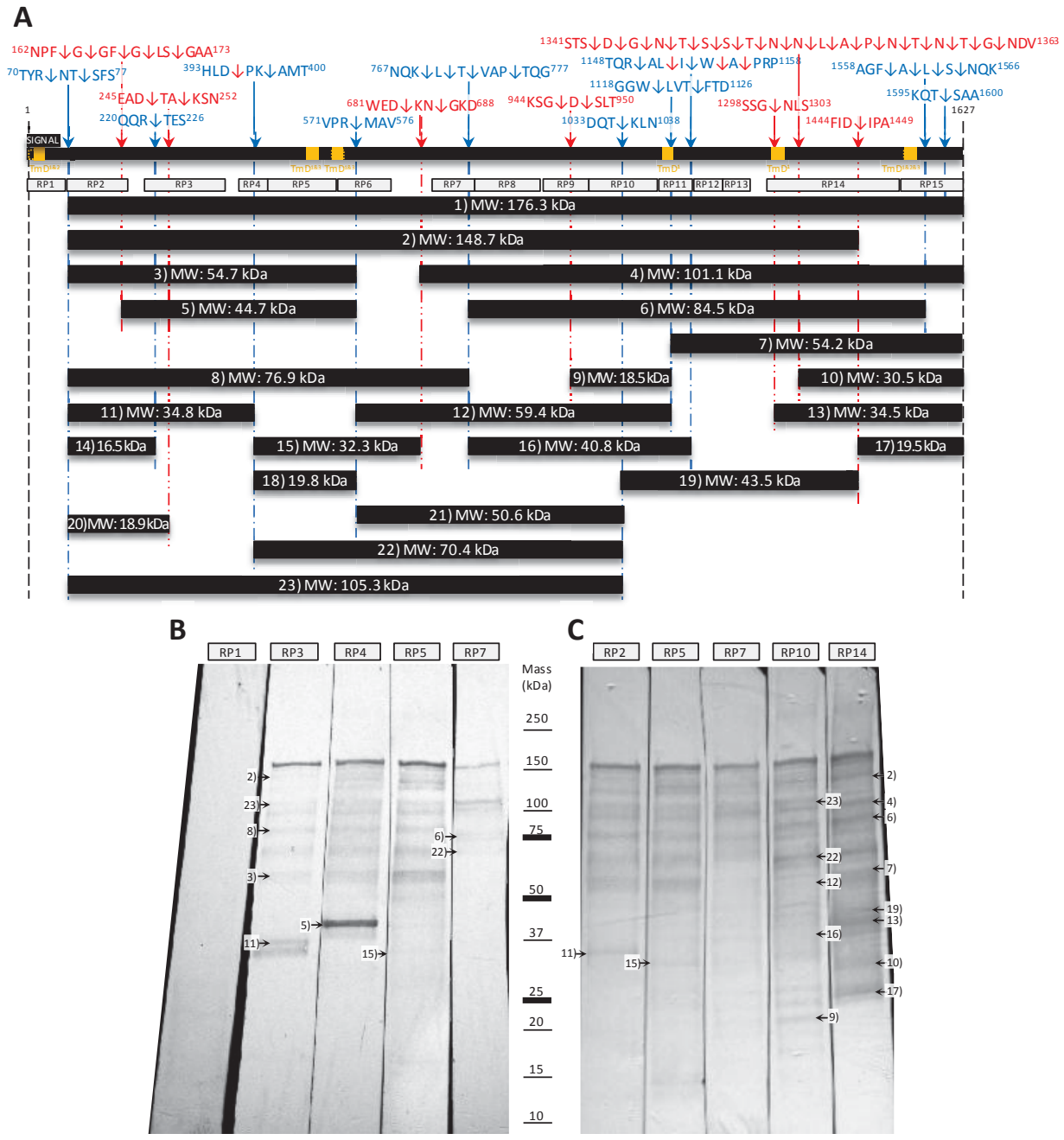
No.	ID	Peptide Sequence N-terminal dimethyl labelled peptides	Score	E-value
1	N1	R.^{73}**N**TSFSSLPLTGENPGAWALVR^{93}.D	107	6.20E^{-09}
	N2	T.^{75}**S**FSSLPLTGENPGAWALVR^{93}.D	26	0.039
3	N3	R.^{224}**T**ESGQNTSTTGAMFGLKVKNAEADTAKSNEKLQGAE ATGSSTTSGSGQSTQR^{275}.G	72	6.60E^{-07}
5	N4	K.^{398}**A**MTANYPPSWR^{408}.T	67	5.70E^{-05}
6	N5	R.^{574}**M**AVAGAKFVGR^{584}.E	62	5.00E^{-04}
8	N6	K.^{770}**L**TVAPTQGTNWSHFSPTLSR^{789}.F	121	7.00E^{-11}
	N7	L.^{771}**T**VAPTQGTNWSHFSPTLSR^{789}.F	77	8.10E^{-07}
	N8	T.^{772}**V**APTQGTNWSHFSPTLSR^{789}.F	64	1.00E^{-05}
	N9	P.^{775}**T**QGTNWSHFSPTLSR^{789}.F	35	5.30E^{-03}
10	N10	T.^{1036}**K**LNLPAYGEVNGLLNPALVETYFGNTR^{1062}.A	173	7.10E^{-16}
11	N11	W.^{1121}**L**VTFTDFVKPR^{1131}.A	58	7.20E^{-05}
	N12	T.^{1124}**F**TDFVKPR^{1131}.A	40	5.70E^{-03}
12	N13	R.^{1151}**A**LIWAPRPWAAFR^{1163}.G	36	1.20E^{-03}
	N14	I.^{1154}**W**APRPWAAFR^{1163}.G	27	2.10E^{-03}
16	N15	F.^{1561}**A**LSNQKVDVLTKAVGSVFKEIINR^{1584}.T	160	1.70E^{-14}

	N16	A.[1562]**L**SNQKVDVLTKAVGSVFKEIINR[1584].T	158	1.30E^{-13}
	N17	L.[1563]**S**LSNQKVDVLTKAVGSVFKEIINR[1584].T	184	3.90E^{-16}
	N18	S.[1564]**N**QKVDVLTKAVGSVFKEIINR[1584].T	102	4.60E^{-09}
17	N19	T.[1598]**S**AAKPGAPRPPVPPKPGAPKPPVQPPKKPA[1627]	58	4.50E^{-06}
		N-terminal semi-tryptic peptides		
2	S1	F.[165]**G**GFGLSGAAPQQWNEVKNKVPVEVAQDPSNPYR[197].F	39	2.10E^{-03}
	S2	G.[166]**G**FGLSGAAPQQWNEVKNKVPVEVAQDPSNPYR[197].F	34	3.20E^{-03}
	S3	F.[168]**G**LSGAAPQQWNEVKNKVPVEVAQDPSNPYR[197].F	42	0.037
	S4	G.[169]**L**SGAAPQQWNEVKNKVPVEVAQDPSNPYR[197].F	59	5.00E^{-06}
	S5	S.[171]**G**AAPQQWNEVKNKVPVEVAQDPSNPYR[197].F	46	3.0E^{-04}
4	S6	D.[248]**T**AKSNEKLQGAEATGSSTTSGSGQSTQR[275].G	126	1.00E^{-11}
	S7	A.[250]**A**KSNEKLQGAEATGSSTTSGSGQSTQR[275].G	80	4.20E^{-07}
5	S8	D.[396]**P**KAMTANYPPSWR[408].T	103	1.40E^{-08}
7	S9	D.[684]**K**NGKDDAKYIYPYR[694].Y	58	1.90E^{-04}
	S10	N.[686]**G**KDDAKYIYPYR[694].Y	61	2.00E^{-04}
12	S11	L.[1153]**I**WAPRPWAAFR[1163].G	56	2.00E^{-04}
	S12	I.[1156]**W**APRPWAAFR[1163].G	46	3.90E^{-03}
	S13	W.[1154]**A**PRPWAAFR[1163].G	49	3.30E^{-04}
	S14	A.[1155]**P**RPWAAFR[1163].G	46	3.10E^{-03}
14	S15	S.[1344]**D**GNTSSTNNLAPNTNTGNDVVGVGR[1368].L	90	1.20E^{-08}
	S16	D.[1345]**G**NTSSTNNLAPNTNTGNDVVGVGR[1368].L	192	8.00E^{-17}
	S17	G.[1346]**N**TSSTNNLAPNTNTGNDVVGVGR[1368].L	158	1.20E^{-13}
	S18	N.[1347]**T**SSTNNLAPNTNTGNDVVGVGR[1368].L	154	6.00E^{-13}
	S19	T.[1348]**S**STNNLAPNTNTGNDVVGVGR[1368].L	121	7.90E^{-10}
	S20	S.[1349]**S**TNNLAPNTNTGNDVVGVGR[1368].L	126	1.90E^{-10}
	S21	S.[1350]**T**NNLAPNTNTGNDVVGVGR[1368].L	121	4.60E^{-11}
	S22	T.[1351]**N**NLAPNTNTGNDVVGVGR[1368].L	117	4.20E^{-10}
	S23	N.[1352]**N**LAPNTNTGNDVVGVGR[1368].L	132	1.00E^{-11}
	S24	N.[1353]**L**APNTNTGNDVVGVGR[1368].L	118	6.30E^{-10}
	S25	L.[1354]**A**PNTNTGNDVVGVGR[1368].L	104	4.00E^{-09}
	S26	A.[1355]**P**NTNTGNDVVGVGR[1368].L	108	9.70E^{-10}

	S27	P.[1356]**N**TNTGNDVVGVGR[1368].L	74	9.20E^{-06}
	S28	N.[1357]**T**NTGNDVVGVGR[1368].L	85	3.40E^{-06}
	S29	T.[1358]**N**TGNDVVGVGR[1368].L	80	1.90E^{-06}
	S30	N.[1359]**T**GNDVVGVGR[1368].L	59	3.00E^{-04}
	S31	T.[1360]**G**NDVVGVGR[1368].L	46	2.10E^{-03}
	S32	G.[1361]**N**DVVGVGR[1368].L	52	4.40E^{-03}
15	S33	D.[1447]**I**PASVNPKMVR[1457].L	62	2.80E^{-05}
		C-terminal semi-tryptic peptides		
2		R.[137]ALYDLDFSKLNPQTPTRDQTGQITFNPF**G**[165].G	35	0.001
4		R.[224]TESGQNTSTTGAMFGLKVKNAEA**D**[247].T	102	3.70E^{-09}
5		R.[386]TAIDRVDHL**D**[395].P	38	6.40E^{-03}
9		R.[224]NDKASSGQSDENHTKFTSATGMDQQGQSGTSAGNPDSLKQDNISKS**G**[246].D	57	1.10E^{-05}
		R.[224]NDKASSGQSDENHTKFTSATGMDQQGQSGTSAGNPDSLKQDNISKSG**D**[247].S	68	3.50E^{-06}
13		R.[1273]QSFGTDHSTQPQPQSLKTTTPVFGTSS**G**[1300].N	27	6.20E^{-03}

Exact site within the peptide is indicated by the bold underlined amino acid, amino acid positions are included for the start and end of the peptides. The peptides listed are the highest scores identified from 4 biological replicates analysed separately using Sciex 5600 and Thermo Scientific Q Exactive™ mass spectrometers. All peptides have expectation values < 0.05.

Figure 6.2 (pictured right): Immunoblots of cell lysates of *M. pneumoniae* probed with sera raised against different regions against P1. Sera raised against 15 different regions ('RP' boxes) of P1 were sourced from Dr. Roger Dumke[93]. A) Simplified cleavage map depicting the P1 adhesin, cleavage sites, and the 15 different regions of P1. The cleavage sites, transmembrane domains are the same as Figure 6.1. The mass of the full length and fragments of P1 predicted by ProtParam[268] have been included. B) and C) Immunoblots depicting *M. pneumoniae* cell lysates probed with sera raised to the different P1 regions. C) Replicate immunoblots with the intensity adjusted to highlight low abundant bands. Many of the fragments from Figure 6.1 are mapped on the immunoblot.



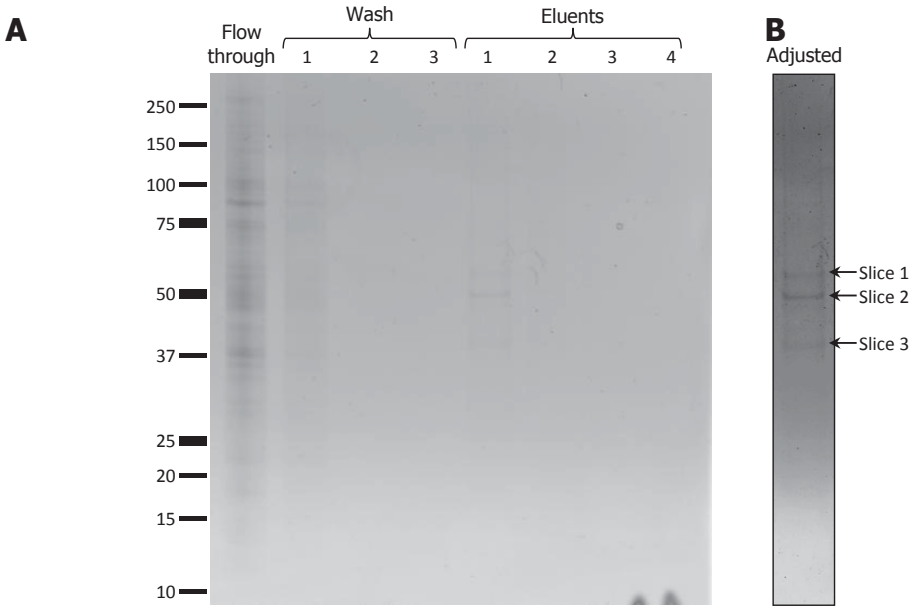
Immunoblots of *M. pneumoniae* cell lysates probed with sera raised against 15 different regions of P1, that were described previously[93]) showed complex banding profiles (Figure 6.2). The RP1 serum that targeted the signal sequence (first 59 amino acids) displayed no bands doubling as the secondary control (Figure 6.2B). RP3, RP4, RP5 and RP7 span the first half of P1 and the immunoblots detected the full length adhesin and fragments consistent with those representing regions (Figure 6.2B; fragments 2, 3, 5, 6, 8, 11, 15,

22, and 23). RP2, RP5, RP7 (higher antibody concentration), RP10, and RP14 sera revealed the full length adhesin and fragments 2, 4, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 19, 22, and 23 after transforming the image intensity (Figure 6.2C). The only fragments not identified confidently in the immunoblots were fragments 14, 18, 20 and 21. It can be seen from the immunoblots in Figure 6.2B and Figure 6.2C that there are more bands visible that were not identified by the approaches employed in this study. Immunoblotting is expected to be a more sensitive method to see lower abundance cleavage fragments.

6.6.3 Functional analysis of the C-terminal tail of P1

Dimethyl labelling data indicated that the carboxy-terminal 30 residues of P1 is released by a cleavage event at serine^{1598} (cleavage site 17 in Table 6.1, sequence: ^{1595}KQT↓SAA^{1600}). The C-terminal peptide has an unusual composition comprising five alanine, five lysine, and thirteen proline residues. It also shares a sequence identity of 54.8% with the carboxy-terminal 30 residues of Mpn142. On closer inspection, the final 15 residues of P1 share 73.3% sequence identity with the last 14 residues of Mpn142 (11 identical positions). The last 15 residues of P1 were synthesised (^{1613}PGAPKPPVQPPKKPA^{1627}) with an N-terminal biotin tag (Chempeptide Limited, China). To investigate binding affinities, interactions with the P1 peptide and a range of host molecules was examined by microscale thermophoresis. The peptide was observed to bind heparin (K_D = 82 ± 2 nM), fetuin (K_D = 322 ± 2 nM), and plasminogen (K_D = 504 ± 89 nM) with nanomolar affinity (from two replicates).

To identify potential binding partners, the biotinylated peptide was coupled to Avidin Agarose. As a control, proteins from a native A549 cell lysate were incubated only with Avidin Agarose beads. This profile was compared to a native A549 cell lysate incubated with Avidin Agarose coupled with the biotinylated peptide. Flow through, washes, and eluents were fractionated by SDS-PAGE (Figure 6.3A). Three protein bands were analysed by LC-MS/MS (Figure 6.3B). LC-MS/MS analysis of Slice 1 and 2 identified tryptic peptides that mapped to the intermediate filament cytoskeletal proteins cytokeratin 7 (Mascot score = 1157), cytokeratin 8 (Mascot score = 2737 & 1486), cytokeratin 18 (Mascot score = 2592), and vimentin (Mascot score = 617) (Figure 6.3C). Slice 3 was identified to be glyceraldehyde-3-phosphate dehydrogenase, which was also identified in the control eluents and therefore not included as a potential binding partner.



C

Slice 1

P05787|Keratin, type II cytoskeletal 8

```
  1 MSIRVTQKSY KVSTSGPRAF SSRSYTSGPG SRISSSSFSR VGSSNFRGGL
 51 GGGYGGASGM GGITAVTVNQ SLLSPLVLEV DPNIQAVRTQ EKEQIKTLNN
101 KFASFIDKVR FLEQQNKMLE TKWSLLQQQK TARSNMDNMF ESYINNLRRQ
151 LETLGQEKLK LEAELGNMQG LVEDFKNKYE DEINKRTEME NEFVLIKKDV
201 DEAYMNKVEL ESRLEGLTDE INFLRQLYEE EIRELQSQIS DTSVVLSMDN
251 SRSLDMDSII AEVKAQYEDI ANRSRAEAES MYQIKYEELQ SLAGKHGDDL
301 RRTKTEISEM NRNISRLQAE IEGLKGQRAS LEAAIADAEQ RGELAIKDAN
351 AKLSELEAAL QRAKQDMARQ LREYQELMNV KLALDIEIAT YRKLLEGEES
401 RLESGMQNMS IHTKTTSGYA GGLSSAYGGL TSPGLSYSLG SSFGSGAGSS
451 SFSRTSSSRA VVVKKIETRD GKLVSESSDV LPK
```

P08729|Keratin, type II cytoskeletal 7

```
  1 MSIHFSSPVF TSRSAAFSGR GAQVRLSSAR PGGLGSSSLY GLGASRPRVA
 51 VRSAYGGPVG AGIREVTINQ SLLAPLRLDA DPSLQRVRQE ESEQIKTLNN
101 KFASFIDKVR FLEQQNKLLE TKWTLLQEQK SAKSSRLPDI FEAQIAGLRG
151 QLEALQVDGG RLEAELRSMQ DVVEDFKNKY EDEINHRTAA ENEFVVLKKD
201 VDAAYMSKVE LEAKVDALND EINFLRTLNE TELTELQSQI SDTSVVLSMD
251 NSRSLDLDGI IAEVKAQYEE MAKGSRAEAE AWYQTKFETL QAQAGKHGDD
301 LRNTRNEISE MNRAIQRLQA EIDNIKNQRA KLEAAIAEAE ERGELALKDA
351 RAKQEELEAA LQRGKQDMAR QLREYQELMS VKLALDIEIA TYRKLLEGEE
401 SRLAGDGVGA VNISVMNSTG GSSSGGGIGL TLGGTMGSNA LSFSSSAGPG
451 LLKAYSIRTA SASRRSARD
```

P08670|Vimentin

```
  1 MSTRSVSSSS YRRMFGGPGT ASRPSSSRSY VTTSTRTYSL GSALRPSTSR
 51 SLYASSPGGV YATRSSAVRL RSSVPGVRLL QDSVDFSLAD AINTEFKNTR
101 TNEKVELQEL NDRFANYIDK VRFLEQQNKI LLAELEQLKG QGKSRLGDLY
151 EEEMRELRRQ VDQLTNDKAR VEVERDNLAE DIMRLREKLQ EEMLQREEAE
201 NTLQSFRQDV DNASLARLDL ERKVESLQEE IAFLKKLHEE EIQELQAQIQ
251 EQHVQIDVDV SKPDLTAALR DVRQQYESVA AKNLQEAEEW YKSKFADLSE
301 AANRNNDALR QAKQESTEYR RQVQSLTCEV DALKGTNESL ERQMREMEEN
351 FAVEAANYQD TIGRLQDEIQ NMKEEMARHL REYQDLLNVK MALDIEIATY
401 RKLLEGEESR ISLPLPNFSS LNLRETNLDS LPLVDTHSKR TLLIKTVETR
451 DGQVINETSQ HHDDLE
```

Slice 2

P05783|Keratin, type I cytoskeletal 18

```
  1 MSFTTRSTFS TNYRSLGSVQ APSYGARPVS SAASVYAGAG GSGSRISVSR
 51 STSFRGGMGS GGLATGIAGG LAGMGGIQNE KETMQSLNDR LASYLDRVRS
101 LETENRRLES KIREHLEKKG PQVRDWSHYF KIIEDLRAQI FANTVDNARI
151 VLQIDNARLA ADDFRVKYET ELAMRQSVEN DIHGLRKVID DTNITRLQLE
201 TEIEALKEEL LFMKKNHEEE VKGLQAQIAS SGLTVEVDAP KSQDLAKIMA
251 DIRAQYDELA RKNREELDKY WSQQIEESTT VVTTQSAEVG AAETTLTELR
301 RTVQSLEIDL DSMRNLKASL ENSLREVEAR YALQMEQLNG ILLHLESELA
351 QTRAEGQRQA QEYEALLNIK VKLEAEIATY RRLLEDGEDF NLGDALDSSN
401 SMQTIQKTTT RRIVDGKVVS ETNDTKVLRH
```

P05787|Keratin, type II cytoskeletal 8

```
  1 MSIRVTQKSY KVSTSGPRAF SSRSYTSGPG SRISSSSFSR VGSSNFRGGL
 51 GGGYGGASGM GGITAVTVNQ SLLSPLVLEV DPNIQAVRTQ EKEQIKTLNN
101 KFASFIDKVR FLEQQNKMLE TKWSLLQQQK TARSNMDNMF ESYINNLRRQ
151 LETLGQEKLK LEAELGNMQG LVEDFKNKYE DEINKRTEME NEFVLIKKDV
201 DEAYMNKVEL ESRLEGLTDE INFLRQLYEE EIRELQSQIS DTSVVLSMDN
251 SRSLDMDSII AEVKAQYEDI ANRSRAEAES MYQIKYEELQ SLAGKHGDDL
301 RRTKTEISEM NRNISRLQAE IEGLKGQRAS LEAAIADAEQ RGELAIKDAN
351 AKLSELEAAL QRAKQDMARQ LREYQELMNV KLALDIEIAT YRKLLEGEES
401 RLESGMQNMS IHTKTTSGYA GGLSSAYGGL TSPGLSYSLG SSFGSGAGSS
451 SFSRTSSSRA VVVKKIETRD GKLVSESSDV LPK
```

Figure 6.3: Affinity chromatography of the P1 peptide. A) SDS-PAGE of the chromatography flow through, washes, and elunts. B) Same eluent 1 as panel (A) but transformed with the three slices analysed by LC-MS/MS indicated. C) Peptides identified by LC-MS/MS of the two slices are underlined.

6.7 Discussion

This study shows that P1 is processed at multiple sites generating a complex array of cleavage fragments. Fragmentation of P1 was determined by LC-MS/MS analysis of size-fractionated eluents of columns that were coupled with a range of host molecules. P1, and cleavage fragments derived from it, are likely to bind to the host proteins because i) the cleavage fragments show a different binding profile compared with full length P1, ii) different fragments bind to different host molecules; iii) some fragments bind many host molecules suggesting that P1 and its cleavage fragmants are multifunctional adhesins. The peptide was observed to bind heparin, fetuin, and plasminogen with nanomolar affinity. Affinity chromatography using the C-terminal 30 residues of P1 as bait recovered intermediate filament cytoskeletal proteins cytokeratin 7, cytokeratin 8, cytokeratin 18, and vimentin. These data are consistent with the multifunctional adhesive roles of P1 and that the C-terminus of P1 binds to cytoskeletal proteins, a novel and groundbreaking observation that warrants further investigation.

2D-gel electrophoresis and affinity chromatography using a range of host molecules as bait identified 22 fragments of the P1 adhesin. The recovery of fragments from Avidin Agarose columns loaded with biotinylated *M. pneumoniae* surface proteins indicated that most of the cleavage events that generated the fragments occur on the cell surface (Figure 6.1). Dimethyl labelling experiments allowed the precise location of cleavage events to be mapped onto P1 (Table 6.1). P1 fragments were also recovered from affinity columns loaded with a range of host proteins. Furthermore, immunoblots containing whole cell lysates of *M. pneumoniae* probed with sera from 15 regions spanning the P1 molecule identified many fragments of P1 (Figure 6.2). From these blotting data, more fragments were identified, particularly between the 100 – 150 kDa range, that were not characterised during 2D-SDS PAGE or in eluents from 'Bait and Prey' chromatography experiments. This is expected as immunoblotting is more sensitive than LC-MS/MS for the identification of low abundant proteins (reviewed in[433]).

A cleavage event that removes the first 59 amino acids of P1 has been reported[246]; this sequence spans most of the 'RP1' recombinant[93]. No other cleavage events have been observed for P1 except for a ~40 kDa carboxyl terminal truncated fragment of P1 (potentially representing fragment 19 from this study) that forms a complex with full

length P1 protein and other accessory proteins[247]. A truncated form of DnaK was also found in this complex, but neither fragments were investigated further[247].

The P1 adhesin is a functionally versatile molecule executing essential functions in adherence to biotic and abiotic surfaces and a major structural protein in motility. Given that P1 is known to be localise to the tip of the attachment organelle, a structure that provides polarity and drives motility, P1 is likely to also play a role in cell invasion. Significant regions in P1 have been extensively characterised and mapped by Schurwanz et al.[93]. Highly immunogenic regions and adherence mediating regions were found distributed throughout P1 but the carboxyl half was prominent in this role[93]. Schurwanz et al. observed that patient sera bound to regions in P1 that were not responsible for adherence[93]. Schurwanz et al. also illustrated this from previous studies where all immunoreactive regions of P1 were localised outside of adherence-mediating regions[93]. Taken together, this suggests that P1 diverts the binding of host antibodies away from regions in P1 responsible for adherence. It can be hypothesised from the study presented here that post-translational processing releases some of these P1 fragments from the surface, which contain these immunogenic regions, into the extracellular milieu, a process that may represent an immune decoy mechanism that seeks to bind and block host antibodies as observed in Protein M of *Mycoplasma genitalium*; a close relative of *M. pneumoniae*[434]. After releasing fragments with immunoreactive regions, the fragments that contain adherence-mediating regions could continue to have a direct role in adherence without being blocked from host antibodies. However, further investigation is required to test this hypothesis.

We recently showed that Mpn142, a member of the same operon that houses the P1 gene (*mpn141*), is cleaved extensively[230]. Post-translational processing of adhesins has been well characterised in *M. hyopneumoniae* where cleavage fragments have been shown to adhere to porcine cilia, porcine kidney epithelial cells, and a range of host molecules such as the glycosaminoglycan mimic heparin, plasminogen and fibronectin[186,217-221,223-227]. Cleavage could be a mechanism to release functional protein modules, alter folding patterns in a way that exposes different protein:protein interaction sites that are hidden in the tertiary structure of the parent molecule (Widjaja et al., submitted, Chapter 5). All the P1 fragments that were recovered during heparin affinity chromatography contained

putative glycosaminoglycan binding motifs except an N-terminal and a central fragment (Figure 6.1, fragment 14 and 16). These motifs consists of clustered, positively charged amino acids that have been shown to have a role in binding to glycosaminoglycans[276], actin[401], and plasminogen[401]. Heparin mimics the glycosaminoglycans found in the extracellular matrix and on the surface of host cells[281]. Due to its negative charge[275], heparin affinity chromatography can also be employed to enrich for DNA binding proteins, though further work is required to determine if P1 fragments bind DNA. Pathogens such as *Staphylococcus* and *Neisseria* spp., *Helicobacter pylori,* and *S Streptococcus pyogenes* are able to recruit heparin to the bacterial cell surface and employ bound heparin to bind other host molecules[406]. *M. hyopneumoniae*, and *Mycoplasma gallisepticum* have been shown to bind heparin to aid in host adherence[217,272]. Finally, heparin has also been implicated in biofilm formation by increasing cell-cell interactions in the Gram-positive pathogens, *S. aureus*[283] and *Lactobacillus rhamnosus*[284]. Heparin affinity chromatography of *M. pneumoniae* has been performed previously[43] identifying only nine proteins, none of which was P1. Recently, we showed that Ef-Tu in *M. pneumoniae* displays a strong affinity to heparin (Widjaja et al., submitted, Chapter 5).

Cleavage sites within P1 were often clustered within a small sequence of amino acids. For example, 18 cleavage sites clustered between amino acids 1343 – 1361 in the C-terminus of P1. Sequential cleavage patterns similar to this was also reported in Mpn142[230] and in Mhp493 from *M. hyopneumoniae*[225]. The surface proteome of *M. pneumoniae* (Appendix 1) reveal two surface accessible aminopeptidases that may target a neo-N-terminal cleavage event and sequentially clip amino acids. The function of these clipping events remains unknown but be mechanism to alter the function and localisation of cleavage fragments on the cell surface or within membrane associated protein complexes or represent a mechanism to recycle amino acids[225]. Cleavage site 14 in P1 (Figure 6.1) occurs within a large predicted disordered region (amino acid range 1187 – 1382, PONDR®[313]). The inherent flexibility of disordered regions facilitatess protein:protein interaction events[435]. Disordered regions are hypothesised to be prone to protease activity[186,221-225].

Inaddition to the full length protein, biotinylation studies indicated 12 fragments of P1 were accessible on the surface of *M. pneumoniae*. P1 only has one signal sequence and

six predicted transmembrane domains that span the adhesin; four of which have been predicted by previous studies[159,301]. Cleavage of Mpn142 in *M. pneumoniae*[167], occurs after the molecule translocates to the cell surface. It is conceivable that processing of P1 occurs on the surface after translocation and the fragments remain anchored to the surface via the predicted transmembrane domains. Consistent with this view, tryptic peptides identified in cell shaving experiments did not map to putative transmembrane domains except in the putative transmembrane domain located around residue position 1294. The fragments that were not identified on the surface may have potentially been released in the extracellular milieu during washes before surface labelling. Electron micrographs depicting *M. pneumoniae* immunostained with ferratin anti-P1 antibodies showed labelling at sites distant from the *M. pneumoniae* membrane[27].

The carboxyl tail of P1 contains a transmembrane domain (residues 1528 – 1545) that had the highest TMpred prediction score (score of 2045) which The C-terminal amino acid sequence of P1 tail is homologous to the carboxyl terminus of Mpn142 (43.1% sequence identity) which also contains a predicted transmembrane domain (TMpred score of 2518)[230]. The final 15 amino acids of P1 (^{1613}PGAPKPPVQPPKKPA^{1627}) has 73.3% sequence identity with the same region in Mpn142. Almost half of this sequence consists of proline residues, and proline-rich regions in proteins have been implicated in protein:protein interactions[436-438]. It has been suggested that proline residues could anchor the C-terminus of P1 in the cell membrane[425]. Lysine-rich regions are associated with binding plasminogen in *M. hyopneumoniae*[221], group A[439] and group B Streptococcus[401]. Lysine-rich regions have also been shown to be necessary for binding actin in group B Streptococcus[401], and binding DNA in *Brevibacillus thermoruber*[440]. It could be the presence of proline and lysine residues in this sequence that defines the high affinity to the host molecules tested by microscale thermophoresis in this study. Fragments 7, 10, 13, and 17 are all C-terminal fragments of P1 that bind heparin Due to the presence of thirteen putative heparin binding regions within P1, it is likely that other regions of P1 also bind these glycosaminoglycans.

Though plasminogen activation assays were not conducted in this study, the capacity for the P1 peptide to bind plasminogen may indicate that the P1 adhesin can promote the activation of bound plasminogen into plasmin. This plasmin could then be used to degrade

extracellular matrix proteins as seen in other *M. pneumoniae* proteins[209,210,213,215] and possibly assist in cell invasion[441,442].

Previous work suggests that sialic acid is the dominant host receptor for the P1 adhesin[144-146,271]. Consistent with these earlier studies the P1 tail has a strong affinity to the sialic acid rich protein, fetuin. However, it is very plausible that this is not the only region of P1 that binds sialic acid as other adherence-mediating sites have been discovered (covered in[93]). It has previously been shown that both sialic acid residues and the P1 adhesin are required for the formation of *M. pneumoniae* biofilms[160]. By establishing biofilms, bacteria are more resilient to antibiotics and avoid removal by the host immune system[421] adding to the chronic persistence of *M. pneumoniae*. Schurwanz et al. reported that the C-terminus of P1 (amino acid range 1521 – 1627) had reactivity to all fourteen *M. pneumoniae* infected patient sera tested. The C-terminal 30 amino acids may be released into the extracellular milieu following cleavage and function as an immunogenic decoy protein.

The cytoskeletal proteins, vimentin, cytokeratin 7, cytokeratin 8, and cytokeratin 18 were selectively recovered from A549 cell lysates when the C-terminal peptide was coupled to Avidin Agarose (Table 6.2). These preliminary observations will require further studies to determine if there is a direct interaction between the C-terminus of P1 and these host molecules. Cytokeratin 7 is found in simple epithelia of lungs and other tissues[443], and has been shown to be partly responsible for stabilising cytokeratin 18 in mouse models[444]. Both cytokeratin 8 and 18 are major structural proteins of epithelial cells[445] and are found in the intermediate filaments of A549 cells[446]. Cytokeratin 8 has been identified to reside on the cellular surface of carcinogenic keratinocyte cells (HaCat)[447], carcinogenic mammary cells[448], and carcinogenic hepatocytes[449]. Cytokeratin 8 and 18 are co-expressed and frequently associated together[450,451], whereas vimentin forms filaments completely separate to both cytokeratins[452]. Vimentin is primarily expressed when epithelial cells transition into mesenchymal cells and function to induce changes in cell shape, motility and adhesin during this transition[453]. Vimentin has also been observed to be secreted to the extracellular matrix and on the surface of activated macrophages[454]. Cytokeratin 8, 18, and vimentin are suggested to be targeted by different pathogens after successfully

invading host cells[447,455-458] or cytoskeletal rearrangement[459-463]. Additionally pathogens have also manipulated these cytoskeletal proteins for immune regulation[458,464,465].

6.8 Conclusion

The P1 adhesin is a major protein in adherence, motility, and biofilm formation in the *M. pneumoniae* attachment organelle. This study reports that the P1 adhesin is subject to extensive post-translational processing forming twenty-two fragments from seventeen cleavage sites. P1 has already been shown to have a role in immune evasion through genomic recombination of its RepMP elements. Cleavage fragments may also contribute to immune evasion by binding with adherence-blocking antibodies before they target regions in P1 that are localised to the tip structure and along the length of *M. pneumoniae* cells[425]. Processing also leads to greater protein disorder and exposes sites in molecules that previously were inaccessible. Processing may also generate functionally-important cleavage fragments that target sites distant from where *M. pneumoniae* binds to ciliated epithelium. These cleaved fragments may be peptide hormones or antibacterial peptides or function as nucleation sites during the early stages of biofilm formation either by promoting protein:DNA (for extracellular DNA) or protein:protein interactions. Protein processing has been described in *M. hyopneumoniae*, *M. gallisepticum*, and *Spiroplasma citri* and is likely to be a widespread mecahism to promote protein modification in the Mollicutes. Further investigation is required to determine the degree of protein processing in different protein species that reside on the surface of *M. pneumoniae*. Moreover, shown here is that the very carboxyl tail of P1 plays a role in adhering to glycosaminoglycans, sialic acid residues, and plasminogen and potentially host cytoskeletal proteins. The cleaved C-terminal tail could also further expand the role P1 already has in immune evasion as an immune decoy protein fragment. This study expands on our knowledge of the role that the P1 adhesin plays during the pathogenesis of *M. pneumoniae* and hopefully will assist in the generation of a successful vaccine against this respiratory pathogen.

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Chapter 7.
Final discussion

i3 ithreeinstitute

The overarching aim of this thesis was to identify the full gamut of proteins that reside on the external cell surface of *M. pneumoniae* (Aim I) and determine which of these interact with molecules that reside on the surface of host cells (Aims II and III). The presence of a cytoskeletal extension known as the attachment organelle has been the subject of many previous studies. The role of different proteins in determining the structure of the organelle as well as efforts to understand how they are positioned in the adhesive tip has been a primary research focus over the past 30 years. The work presented in this thesis was performed compliment these studies by taking an empirical approach to studying the cell surface protein topography of *M. pneumoniae*, and to then selectively study several surface proteins in greater detail.

Two approaches were used to define the list of surface accessible proteins. While both methods were complimentary in that they both identified a large proportion (117 proteins or 73%) of the 160 proteins that were identified, some proteins were only identified using one of the two protocols. This dual method approach not only provided a level of confidence in the final list of proteins but each protocol offered distinct yet complimentary advantages. For example, trypsin shaving was useful for identifying regions of membrane proteins that were surface exposed and other regions that were inaccessible to trypsin, providing insight into how proteins might position themselves in the bilipid layer (Figure 2.4). Biotinylation proved to be extremely useful because it was a more sensitive approach due to the ability to enrich for labelled intact proteins with avidin chromatography. LC-MS/MS analysis of avidin eluents that were fractionated by size using SDS-PAGE was insightful for the following reasons: i) it became clear that protein processing occurs frequently in a wide variety of proteins that localise on the surface of *M. pneumoniae*; ii) one could characterise the cleavage fragments using 2D-SDS-PAGE; iii) it provided a platform to identify functional domains in surface proteins and confirmed that processed fragments reside on the cell surface

The data described in this thesis shows that adhesins, as seen in other Mollicutes, are not the only target of protein processing events. Cleavage events of *M. pneumoniae* proteins were identified in a wide range of functionally diverse molecules including proteins that are retained by being anchored to the lipid membrane (lipoproteins and ABC transporters with multiple transmembrane domains), metabolic enzymes including most members of

the glycolytic pathway, and ribosomal proteins. Native preparations of whole cell lysates of *M. pneumoniae* proteins were applied to affinity resins coupled with a diverse array of host molecules, including: actin, fibronectin, fetuin, heparin, and plasminogen. These molecules are likely targets for adhesion by *M. pneumoniae* surface proteins and thus play an important role in colonising epithelial and other cell surfaces. Surface protein complexes derived from A549 epithelial cells labelled with biotin were also coupled to Avidin Agarose to recover *M. pneumoniae* binding proteins in affinity chromatography experiments. These complimentary approaches each generated a list of proteins that suggest they interact directly with the target host bait (Figure 3.2) (Aim II).

An underlying feature of all these studies (surface proteome and 'Bait and Prey' affinity chromatography) was that protein mass context was retained using SDS-PAGE. The 'slice n dice' enhanced the detection of low abundance proteins through fractionation, thus reducing the complexity of the peptide sample introduced to the mass spectrometer. Secondly, it also highlighted the scale of endoproteolytic processing events that occur in *M. pneumoniae* in a way that 'shotgun' LC/MS/MS based proteomics cannot, particularly the subset of proteins that localise to the cell surface (Aim III). This observation was ground-breaking because it indicated that Mollicutes that belong to diverse Clades post-translationally process surface accessible proteins. Up until that point, extensive proteolysis of surface accessible proteins had been described in *M. hyopneumoniae*[186,187,216-227,244], an organism that i) specifically colonises swine, ii) belongs to Clade Hominis, iii) lacks an attachment organelle, and iv) other than having evolved an ability to colonise ciliated epithelial cells in the respiratory tract, shares little in common with *M. pneumoniae*. In this regard, *S. citri* was recently shown to process adhesins extensively by an unconventional protease[236]. The frequency of cleavage of these adhesins was postulated to be influenced by conformational changes (induced by environmental factors such as mechanical or chemical stress) of cell membrane proteins[236]. *S. citri* (of the Spiroplasma Clade) is a sophisticated pathogen that displays the ability to colonise both host plant cells and the insect vector, *Circulifer haematoceps*[236]. *S. citri* is able to adhere to the salivary gland and mid-intestinal cells of the *C. haematoceps* and is transferred from saliva to the plant[466]. These observations indicate that post-translational proteolysis is a central theme and novel mechanism to expand protein function in genome reduced bacteria belonging to the Mollicutes Class.

Proteolysis of key surface proteins increases the surface diversity and expands the functional proteome. This may assist genome reduced bacteria to colonise a range of eukaryote host surfaces ranging from insects to humans. This thesis expands further on this theme for *M. pneumoniae*. A systems wide N-terminomics method was developed by Iain Berry in the Djordjevic laboratory. This method was applied to characterise the N-terminome of *M. pneumoniae*, and a subset of this data is integrated into this thesis. This is part of a larger effort to map cleavage sites (neo-N-termini) in representative *M. pneumoniae* proteins of interest. Sites from the N-terminome assisted with validating cleavage fragments identified in 'Bait and Prey' and surface proteome experiments. Identification of cleavage sites provides information on the exact size of the fragments, how the fragment may re-fold, and what motifs are present. Collectively, the pioneering studies presented in this thesis have helped to uncover important post-translational strategies used by *M. pneumoniae* to colonise host cells. By systemically combining the surface proteome with 'Bait and Prey' chromatography, potential interaction partners (Appendix 3) can be identified and highlight the importance of how cell surface proteins are processed. Post-translational proteolysis increases cell surface protein diversity; though the exact function of these cleavage fragments remains to be experimentally verified. In Chapter 6, a peptide was identified to promote binding to host proteins. This provided proof of concept that cleavage fragments are likely to be functionally significant molecules that play an important role in the biology of *M. pneumoniae*. It is conceivable that cleavage fragments derived from surface accessible proteins could: i) compete for host molecule binding sites; ii) refold exposing novel binding sites and new functions; iii) influence biofilm formation; or iv) act as immune decoys, antimicrobial peptides or peptide hormones.

Indeed, almost all of the attachment organelle proteins were identified on the cell surface (Appendix 1) and were processed (Appendix 2). This in itself is a major finding because it suggests that the attachment organelle proteins are multifunctional. They are able to form important interactions with other organelle proteins to generate a functioning organelle; yet are also localised on the cell surface and subject to complex processing events to perform other functions. Because of the previously published work and knowledge of their importance, an in-depth analysis of processing events was performed and is presented for the major *M. pneumoniae* adhesin P1, and its accessory proteins P40 and P90 (Aim IV).

Data is also presented for elongation factor Tu which has been reported as being an important moonlighting adhesin not only in *M. pneumoniae*, but also other bacteria (Aim IV). Taken together, this thesis adds insight to the identities of proteins that localise to the cell surface of *M. pneumoniae* that may play a key role in the biology of this pathogen.

7.1 The surface proteome of *Mycoplasma pneumoniae*

By employing two independent methods, enzymatic shaving and surface labelling, 160 proteins were identified on the surface of *M. pneumoniae* (Appendix 1). These complimentary approaches identified 117 proteins common to both methodologies.

Enzymatic shaving is essential to demonstrate how proteins are presented on the cell surface. By mapping peptides back to the protein, surface exposed regions of membrane associated proteins were identified. This was particularly useful for proteins that contained multiple transmembrane domains, which can be difficult to solubilise, such as zinc metalloprotease (Figure 2.4). Three peptides were identified from shaving experiments that span the last 500 amino acids. The molecule contains two predicted transmembrane domains in the first 200 amino acids. This suggests that the first 200 amino acids of the zinc metalloprotease loops inwards, and the remainder of the protein is surface exposed. Enzymatic shaving allows these surface exposed regions to be mapped, and identifies regions that are better targets for vaccine development. The identification of DNA gyrase B and ribosomal protein L4 on the *M. pneumoniae* cell surface (Appendix 1) is interesting as both are targeted by fluoroquinolone and macrolide antibiotics, respectively[467]. Though further work is required, it is possible that these proteins could be binding antibiotics on the cell surface, thus preventing a subset of antibiotics from entering and killing the cell. Further work is required to investigate this. It is also feasible that the use of some antibiotics may influence how surface accessible proteoforms of DNA gyrase B and ribosomal protein L4 execute their extracellular function(s).

Cell surface shaving also identified unusual proteins in *M. pneumoniae* that contain long stretches of amino acids that are bereft of lysine or arginine residues and these molecules cannot be characterised using trypsin. The attachment organelle proteins, P30 (Figure 2.7) and HMW1 (Figure 2.8), are examples of this. PeptideCutter[268] was employed to predict peptides that are generated from using different enzymes digests. Based on in silico models, formic acid (which cleaves after aspartic acid[468]) and endoproteinase GluC

(which cleaves after glutamic acid[469]) seem to be the appropriate strategies to study these unusual proteins but significant optimisation would be necessary. HMW1 is rich in glutamic acid residues and so future work would need to reduce the efficiency of the GluC enzyme so that peptides generated are of suitable size for detection. Despite repeated efforts, surface accessible regions could not be reliably mapped for HMW1 (Figure 2.8). Nonetheless, trypsin shaving and surface biotinylation data demonstrated that HMW1, as well HMW2 and HMW3, are presented on the surface of *M. pneumoniae*. This is the first time that HMW2 and HMW3 have been shown to be surface exposed. As both are well known for their significant role in attachment organelle formation and stability[118,165], knowledge of surface exposed regions in either (or both) of these proteins could be desirable to understand how these complex proteins function.

Biotinylation provides another layer of sophistication in efforts to understand how proteins are presented on the cell surface that cannot be ascertained from shaving experiments. The method is very robust because it uses a membrane impermeable reagent to label accessible lysine residues on proteins that are exposed on the cell surface and takes advantage of the strong non-covalent interaction between biotin and avidin. Size fractionation afforded by 1D- and 2D-SDS-PAGE provided an added degree of resolution, because it showed that cleavage fragments were located on the cell surface and highlighted the modular nature of proteins residing within the attachment organelle. As such a protein encoded from an ORF sequence may have an overall predicted pI of 8. However, surface proteins often have a modular structure. Cleavage fragments not only have different masses, but their pI can vary. Thus, multiple proteoforms can be generated from a single ORF. As such it was not uncommon to find protein spots on 2-D gels with different masses and isoelectric points (as seen in lipoprotein MPN_284, Figure 3.4 and Table 3.4). These observations provided the first clue that different proteoforms may execute different functions. This has significant ramification for understanding the functional proteome of a pathogen.

Another notable observation was the identification of 55 proteins (25% of the surface proteome) (Figure 2.5) that reside on the surface of *M. pneumoniae* that lack evidence of signal sequence motifs (SignalP 4.1 and SecretomeP 2.0 predictions) or transmembrane domains (TMpred predictions). Additionally, the predicted cellular locations of ~76% (121

of the 160 proteins, Figure 2.6) of these proteins were either cytoplasmic or unknown according to UniProt and PSORT analyses. These predictions highlight the limitations of relying on protein location algorithms and poorly annotated databases to hunt for candidate vaccine antigens. Wang and Jeffery reported more than a third of the surface proteome identified in previous studies did not have any bioinformatically predicted transmembrane domains or secretion signals and suggested novel secretion pathways[195]. Reverse vaccinology has been touted widely as an important strategy for identifying vaccine targets[470-472]. The data presented here suggest that surface proteome studies should be used in conjunction with reverse vaccinology to identify targets for vaccine development.

Cell lysis is often quoted as a reason why proteins that have canonical functions in the cytosol may be detected on the surfaces of bacteria. For many years these proteins were ignored in surface proteome studies and regarded as artefacts[179,181,363]. Cell lysis may indeed prove to be an important mechanism needed to release these proteins, and genomic DNA, as "public goods"[193,264,473] particularly within biofilms. As such it is perhaps ironic that the main reason offered to dismiss molecules with canonical functions in the cytosol as valid surface proteins may indeed provide a valid explanation for their retention on the cell surface. Membrane vesicles have also been shown to be responsible for the delivery of these "public goods" to the cell surface[191,192].

Extracellular DNA is a major constituent of the extrapolymeric matrix and has been shown to aid in biofilm formation by promoting cell:cell interactions, adhesions to surfaces (biotic and abiotic), reinforcing the biofilm structure, and increasing resistance to antibiotics[261,264]. The negative charge of heparin is similar to the charge of DNA[275] and it is likely that extracellular DNA is important in *M. pneumoniae* biofilms. In Chapter 3, 111 of the 160 *M. pneumoniae* surface proteome was identified in heparin affinity chromatography eluents. However, further work is required to determine the role of these heparin binding proteins in *M. pneumoniae* biofilms.

Pseudomonas aeruginosa was recently identified to undergo explosive cell lysis, releasing cytosolic components such as DNA and cytosolic proteins[193]. In the same process, membrane membrane vesicles are createad that package some of these cytosolic components[193]. These membrane vesicles serve to provide public goods for the bacterial

biofilm, where cytosolic proteins are able to execute moonlighting functions in biofilm formation[261-263]. The presence of moonlighting proteins on the surface that can bind extracellular DNA could be a prerequisite for initiating biofilm formation. As such there is opportunity for these cytosolic proteins to evolve alternate functions as adhesins that promote protein:protein, and protein:DNA interactions. Two aminopeptidases have been shown to moonlight on the surface of *M. hyopneumoniae* and bind plasminogen and heparin[243,244]. Both are also able to promote the activation of plasminogen into plasmin in the presence of host plasminogen activators[243,244]. The leucine aminopeptidase is also able to bind foreign DNA[244] which may assist in biofilm formation in *M. hyopneumoniae*.

Previous studies have independently shown that several of these "predicted cytosolic" proteins reside on the surface of *M. pneumoniae*[205,208-210,213-215,270]. These studies found that several enzymes (8 out of the 19 tested) in the glycolytic pathway were exposed on the surface of *M. pneumoniae*[209]. These glycolytic enzymes were demonstrated to display an affinity for a range of host molecules including plasminogen, with some able to promote the activation of plasminogen (in the presence of plasminogen activators) to cleave key host extracellular matrix proteins[205,208-210,214,215]. Using the approaches in this thesis, 16 of the 19 glycolytic enzymes tested by Gründel et al., were identified in the surface proteome presented in Chapter 2, highlighting the difference between methodologies used. Surface accessibility and different targets for the enzymes, antibodies, and tags could account for differences in results between previous surface studies and Chapter 2.

7.2 Characterising potential adhesins of *Mycoplasma pneumoniae*

A systems wide approach, where host molecules were used as bait in affinity chromatography experiments to capture *M. pneumoniae* proteins and their cleavage fragments provided the simplest and most efficient mechanism to determine functional roles for processed fragments. There are clear limitations in this approach. For example, most of the binding interactions identified in these affinity studies were not confirmed to be due to direct interactions between *M. pneumoniae* proteins and the host molecules used as bait. Some proteins identified in the chromatography eluents may be part of a protein complex where another protein binds to the bait. Direct binding interactions were only confirmed for a small number of protein:bait interactions using microscale thermophoresis. Nonetheless, the studies provided sufficient information to enable

hypothesis driven experiments to be carried out. However, there is evidence that some of these proteins were reproducibly isolated from columns coupled to different bait host molecules. These proteins could be prioritised for future studies.

A range of host molecules were immobilised separately onto affinity columns as part of an overall strategy to identify potential *M. pneumoniae* proteins that interact with these host molecules. These experiments were performed under native conditions (physiological salt and pH with mild detergent) and *M. pneumoniae* protein complexes are likely to have been retained. The host molecules included: surface protein complexes derived from human lung carcinoma (A549) cells, cytoskeletal actin, fetuin (sialic acid rich), the extracellular matrix protein fibronectin, the glycosaminoglycan mimic heparin, and the protease proenzyme plasminogen. LC-MS/MS analysis of tryptic fragments generated from eluents derived from all the 'Bait and Prey' experiments identified 337 *M. pneumoniae* proteins, nearly 49% of predicted *M. pneumoniae* proteome (Appendix 2). Approximately a third of these proteins were identified in all the eluents which suggests that either these proteins are a common feature of multiple complexes, or the proteins are multifunctional and bind a range of different host molecules. Each of the 337 proteins will need to be cloned, expressed and purified as purified recombinant proteins to determine if these are direct interactions with host molecule. Due to time and resource constraints, interactions were not verified for all 337 proteins. Rather, interactions were investigated for model proteins of interest such as Ef-Tu and the P1 adhesin.

261 of the 337 proteins (77%) were recovered in eluents from heparin chromatography (Figure 3.1). Heparin is a molecular mimic of the glycosaminoglycans (and proteoglycans) found on host cell surfaces and extracellular matrices[281]. Many bacterial pathogens and commensals[281,282,406] including *Mycoplasma spp.*[217,218,224,272] target glycosaminoglycans for adherence to host cell surfaces. Many proteins in biology interact with glycosaminoglycans and the unique sulfate patterns found in heparin and heparan sulfate underpin many of the important biological interactions that occur between bacterial adhesins and these glycosaminoglycans. The proteins identified in the heparin eluents interact: i) due to electrostatic interactions arising from the negative charges of the glycosaminoglycans, or ii) directly to sugar residues in heparin, or iii) as part of a complex where another protein has bound to heparin. Further work would need to generate recombinants of each of these

proteins and measure direct protein-heparin binding by ELISA, surface plasmon resonance, or microscale thermophoresis. Pathogens use heparin as a bridging molecule to bind other host molecules, strengthening adherence to host cells[406]. Additionally, heparin has been shown to increase cell-cell interactions such as those required for biofilm formation in *S. aureus*[283] and *L. rhamnosus*[284].

The 'Bait and Prey' data was combined with the *M. pneumoniae* surface proteome presented in Chapter 2 to yield a list of 149 proteins exposed on the cell surface which may potentially bind one of the host molecules tested (Appendix 3). Though further work is needed to verify direct binding, this list suggests that *M. pneumoniae* has a much larger population of moonlighting adhesins than previously thought. Previously reported moonlighting adhesins such as Ef-Tu, Pdh-B, GAPDH, Pdh-A, Pdh-C, lactate dehydrogenase, phosphoglycerate mutase, pyruvate kinase, GroEL, DnaK, and the uncharacterised Mpn474 protein[205,208-210,213-215,269] are also present in this list of 149 proteins. Enolase, and possibly other moonlighting proteins, in *M. pulmonis* are tethered to the cell membrane via the sugar rhamnose[206,207]. *M. pneumoniae* has been shown to also produce rhamnose along with galactose[207], and an N-acetylglucosamine containing polysaccharide[114]. However, there were no subsequent studies that investigated the roles these carbohydrates might have in anchoring proteins to the cell surface. It would be interesting to see if *M. pneumoniae* moonlighting proteins are also tethered in the same way or if there is another molecule(s) responsible for anchoring moonlighting proteins. How moonlighting proteins associate with the surfaces of the organism that released them remains an active area of research.

'Bait and Prey' chromatography is also useful to enrich for protein cleavage fragments as it allows the identification of low abundant fragments that would normally be masked in more complex samples. One of the most novel findings of the work presented in this thesis is that surface accessible moonlighting proteins are targets of post-translational proteolysis. Moonlighting proteins are a class of multifunctional proteins that have been encoded from a single ORF[189]. Proteins arising from gene fusion, alternative splicing, or those that display low secondary functions (promiscuous enzymes) are not classified as moonlighting[189,190]. 108 of the 337 proteins (32%) of the proteins identified by 'Bait and Prey' were also identified as smaller fragments (Appendix 2). Since almost all proteolytic

events are expected to destroy canonical function, proteolysis processing provides a mechanism to release functional modules from within larger multifunctional proteins. Protein cleavage represents a new dimension in protein moonlighting and hypothesise the idea of 'Proteolysis Induced Moonlighting' or PIM. In PIM, cleaved fragments may present novel interaction sites or motifs that are inaccessible due to the structural constraints of the parent full length molecule (Figure 7.1). PIM may release conformationally inaccessible interaction sites, therefore expanding the functional repertoire of ORFs in the genome. These data suggest the functional proteome is much larger than originally envisaged.

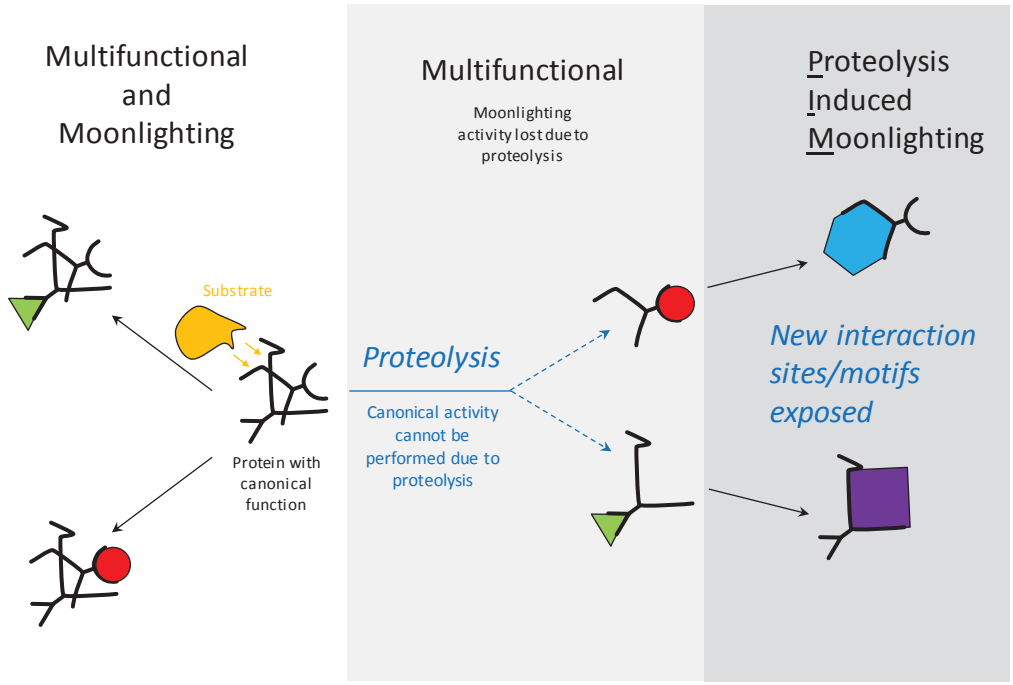


Figure 7.1: Schematic of the concept of Proteolysis Induced Moonlighting (PIM). Substrates are shown as different coloured shapes. A typical protein with a canonical function will interact with one substrate. If the same protein can interact with other substrates, it gains 'Multifunctional and Moonlighting' status as shown on the left. Post-translational proteolysis will create protein fragments, thereby losing the canonical function of the original protein (middle panel). However, proteolysis may expose new interaction sites that were hidden or obscured due to the conformation of the original protein. Surface accessible cleavage fragments can display novel interaction sites.

The precise cleavage sites that generate cleavage fragments were identified and mapped by N-terminal dimethyl labelling (performed by Berry, I.J). Fragments generated by proteolytic processing were also identified by LC-MS/MS of 2D-SDS PAGE spots of *M. pneumoniae* cell lysates, highlighting the authenticity of processing. The proteins identified in this thesis that undergo proteolysis span a wide range of protein groups such as ribosomal proteins, glycolytic enzymes, elongation factors, ATP synthases, chaperones, lipoproteins, and uncharacterised proteins. This diversity of proteins that undergo proteolytic processing suggest that a wide range of proteins have the potential to become multifunctional. Additional analysis is required to determine the extent of endoproteolytic processing for these proteins and to quantify the strength of binding with target molecules. It is unknown which protease(s) is responsible for cleaving surface accessible proteins, but six putative proteases were identified on the *M. pneumoniae* cell surface (Appendix 1).

Proteolytic processing of tip structure proteins has previously been described. Cleavage events shown to occur in *M. pneumoniae* proteins remove the signal sequence in the *M. pneumoniae* attachment organelle protein P1[163,245,246] and P40[167,168]. Processing also cleaves Mpn142, generating P40 and P90[166,167]. More recently Mpn142 was shown to be a target of extensive processing on the *M. pneumoniae* cell surface[230] (see Chapter 4). Lastly, processing has also been shown to generate immunomodulatory peptides from surface exposed lipoproteins[228,229], cleave DnaK and P1[247], and generate fragments of ten lipoproteins[248]. Functionally important cleavage events have also been reported in other Mycoplasmas[231-236] and have been well documented in a variety of bacterial pathogens[237-241]. Extensive cleavage of key surface proteins has also been well documented in *M. hyopneumoniae* generating fragments; some of which bind heparin, fibronectin, actin, and plasminogen[186,187,216-227,244]. Three proteins identified in the surface proteome study presented in Chapter 2 were shown to be cleaved extensively. Post-translational processing is unlikely to be limited to these three proteins. Data mining of the *M. pneumoniae* N-terminome is underway to identify cleavage sites within proteins of interest.

7.3 The multifunctional and moonlighting activity of Elongation factor Tu

Ef-Tu was selected as a model moonlighting protein for this thesis as it was one of the first characterised moonlighting proteins of *M. pneumoniae*. Dallo et al. reported that 17% of total Ef-Tu content in *M. pneumoniae* was surface localised and bound fibronectin in a dose-dependent manner[205]. The data presented in Chapter 5 expands on the number of host targets that Ef-Tu can bind on the surface of *M. pneumoniae*. Using a combination of microscale thermophoresis and ELISA, Ef-Tu was shown to bind a wide range of host molecules including fetuin, actin, heparin, plasminogen, fibronectin, fibrinogen, vitronectin, lactoferrin, and laminin. Other *M. pneumoniae* proteins have been shown to bind these molecules previously highlighting the diversity of *M. pneumoniae* host targets[205,208-210,213-215,271]. Additionally, bound plasminogen was converted to plasmin in the presence of plasminogen activators and this Ef-Tu-plasmin complex was able to degrade fibronectin and vitronectin (Figure 5.7). By binding and activating plasminogen, pathogenic bacteria gain the capacity to penetrate host extracellular matrix and potentially gain access to membrane receptors that promote bacterial uptake (reviewed in[211] and[212]), this could enable *M. pneumoniae* to invade the epithelium. *M. pneumoniae* has been shown to invade human A549 cells[441] and is hypothesised to disseminate to extrapulmonary sights such as the brain, resulting in central nervous system complications[442]. Though further work is needed, it is possible that *M. pneumoniae* invades host cells by first degrading the host cellular matrix with plasmin bound to Ef-Tu.

A central and C-terminal fibronectin binding region has previously been identified in Ef-Tu[389,390]. In this work, bioinformatic analyses identified 12 putative heparin binding motifs that were located along the entire protein. Using thermophoresis, a dissociation constant (K_D) of 48.3 ± 1.3 nM indicates that Ef-Tu binds strongly to heparin (Figure 5.2). The predicted 3D structure of Ef-Tu reinforces the surface accessibility of the putative heparin binding sites (Figure 5.3). Peptide libraries which span the putative heparin binding domains or mutation studies that replace positively charged residues with alanine are required to determine if these putative binding motifs in Ef-Tu are functional.

The sequences of Ef-Tu from *M. hyopneumoniae* and *S. aureus* were aligned with the sequence from *M. pneumoniae* to determine if these putative heparin binding sites are

conserved across members of the low G + C Firmicutes (Widjaja et al., submitted, Chapter 5). All three Ef-Tu proteins were shown to be a target of processing events. Fifteen fragments of *M. pneumoniae* Ef-Tu were characterised in Chapter 5 and many of these were independently isolated from 'Bait and Prey' affinity chromatography of different host molecules (Figure 5.2). Since cleavage fragments are likely to fold in a manner that is different to the parent proteoforms, they may have lost the ability to form complexes with other proteins. All the identified fragments of Ef-Tu contain putative heparin binding motifs underscoring the potential of all these fragments to bind heparin. However, the full length protein and only ten fragments were recovered from heparin chromatography. This implies that either the confirmation of the remaining four fragments of Ef-Tu block the heparin binding sites, or that these fragments are in too low in abundance to be detected by LC-MS/MS. Further work such as expressing and testing direct heparin interactions are needed. It would be interesting to see whether the number of heparin binding domains in the fragments, determine the binding affinity of these fragments to heparin.

Moonlighting proteins can evolve by minor evolutionary base mutations within genes, resulting in amino acid changes which adds an extra function(s) without affecting the canonical role of the protein[474-476]. These amino acid mutations are likely to occur within loops and disordered regions within proteins so as not to affect canonical function[476-480]. A prime example of these amino acid changes is in the difference between Ef-Tu proteins of *M. pneumoniae* and *M. genitalium* which share 96% sequence identity[390]. Six amino acid differences in the sequences between the two organisms result in the inability of *M. genitalium* Ef-Tu to bind fibronectin[390]. Three of these amino acid differences are localised to loops in the molecular model generated by Balasubramanian et al.[390].

Disordered regions in proteins are known to promote protein:protein interactions[435] and thus multifunctional behaviour. However, enzymes and ribosomal proteins typically lack disordered regions and have strict structural constraints imposed on them to execute canonical function efficiently. As such there are only limited sites in a protein sequence that can tolerate mutations that change amino acid sequence. Bioinformatic tools were employed to predict putative protein:protein interaction regions mediated by SLiMs (Short Linear Motifs) for Ef-Tu. Due to a low GC-content in the genome, *M. pneumoniae* tends to encode for a higher proportion of positively charged lysine residues[148]. The hypothesis

presented in Chapter 5 is that moonlighting proteins, like Ef-Tu, have evolved SLiMs in structurally non-conserved, variable regions enriched in positively charged amino acids. SLiMs allow moonlighting proteins to display novel protein:protein interactions with host molecules, expanding protein function.

These evolutionary base changes are tolerated in the genomes of organisms that have duplicated genes, one retaining original canonical function while the other is mutating to evolve a different function[481]. As Mollicutes are genome reduced, moonlighting functions are not likely to arise in this manner as there is no evidence that any of the cell surface moonlighting proteins are products of gene duplication events. It is possible that genetic mutations in *Mollicute* spp. are accelerated to generate moonlighting activity compared to non-*Mollicute* spp. This is seen in accumulation of positively charged residues (underlined) in three SLiMs that are predicted to promote binding to heparin in $Mpn_{Ef\text{-}Tu}$ (^{37}A<u>K</u>EG<u>K</u>SAAT<u>R</u>Y^{47}, ^{183}P<u>K</u>WEA<u>K</u>I<u>H</u>D^{191}, and ^{248}L<u>R</u>PI<u>RK</u>A^{254}). Both ^{37}AKEGKSAATRY^{47} and ^{183}PKWEAKIHD^{191} in *M. pneumoniae* have lysine or arginine residues not seen in $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$ (Figure 5.1) resulting in novel putative heparin binding domains only in $Mpn_{Ef\text{-}Tu}$. The third sequence, ^{248}LRPIRKA^{254}, contains two arginine residues (R^{249} and R^{252}) not seen in $Sa_{Ef\text{-}Tu}$ ($Sa_{Ef\text{-}Tu}$ sequence: ^{248}LHDTSKT^{254}) and thus does not have a heparin binding domain (Figure 5.1). These three residue changes are predicted to occur within evolutionarily non-conserved, variable regions (Appendix 5: Supplementary Bioinformatics). Collectively, this supports the idea that single evolutionary amino acid changes results in the generation of SLiMs for moonlighting activity.

The bioinformatic tool ISIS, predicted only seven protein:protein interaction regions in full length Ef-Tu (Figure 5.2). However, many of the cleavage fragments of Ef-Tu contained predicted proteins-protein interaction sites that were absent in the full length parent protein (Figure 5.2). These data are consistent with the hypothesis of PIM where post-translational proteolysis results in newly folded protein fragments that expose new binding sites originally hidden in the conformation of the full length protein. To verify this, recombinants of all the fragments would need to be crystallised and modelled to determine how their structures compare to the parent protein.

7.4 Proteolysis Induced Moonlighting activity of P1, and potentially P90 of *Mycoplasma pneumoniae*

P40 and P90 from Mpn142 (Chapter 4), and P1 (Chapter 6) were shown to undergo post-translational proteolysis events. By combining 1D- and 2D SDS PAGE of *M. pneumoniae* cell lysates; avidin enrichment of labelled surface proteins; 'Bait and Prey' affinity chromatography; and N-terminal dimethyl labelling (performed by Berry, I.J), both Mpn142 and P1 was found to be cleaved to form 15 (Figure 4.2) and 22 (Figure 6.1) fragments, respectively.

There is one large predicted disordered region within P40 and one region at the very C-terminal tail of P90 (Figure 4.2). Both these regions had at least six sequential cleavage sites across nine amino acids (Table 4.1, sequences: ^{299}GAA↓GS↓A↓S↓S↓LQ↓GNG^{311} and ^{1187}ALK↓A↓A↓N↓N↓A↓AP↓KAP^{1199}). There is also a large predicted disordered region spanning 196 residues in P1 (position 1187 – 1382, Figure 6.1) that has 18 sequential cleavage sites across 19 amino acids (Table 6.1, sequence: ^{1341}STS↓D↓G↓N↓T↓S↓S↓T↓N↓N↓L↓A↓P↓N↓T↓N↓T↓G↓NDV^{1363}). Sequential removal of N- or C-terminal residues from cleavage sites within these predicted disordered regions was observed. It is unknown exactly why this occurs, but it has been suggested to allow the cleaved amino acids to be recycled for the synthesis of new proteins[225,227]; echoing the absence of amino acid biosynthesis pathways in *Mycoplasma* spp.. If this hypothesis has merit then it suggests that processing creates new proteoforms and a mechanism to provide amino acids to nourish the pathogen. Further studies are required to interrogate this hypothesis.

Both the C-terminus of P1 and P90 contain predicted disordered regions (Figure 6.1 and Figure 4.2), and within these disordered regions are cleavage sites which release the last 30 amino acids from both protein. The C-terminal 30 amino acids of these proteins share sequence identity of 54.8% (Figure 4.3), and are enriched in positively charged amino acids (lysine and arginine) and bulky amino acids (proline). These residues are known to promote protein:protein interactions. Regions enriched in positive residues are known to interact with plasminogen and actin[221,401,436-439]. The cleavage event and the sequence similarity provoked interest because it suggests that the C-terminal tail of both these molecules have evolved to interact with a similar target. On closer inspection, the final 15

amino acids of both proteins have a higher sequence identity (68.8%). To examine the function of the C-terminal tail of P1 (sequence: ^{1613}PGAPKPPVQPPKKPA^{1627}) was synthesized.

Remarkably, host molecules that interact with the P1 adhesin have not been characterised in detail. Microscale thermophoresis revealed a strong affinity between the synthesised P1 peptide and the host molecules tested (heparin, fetuin, and plasminogen). It is likely the proline and lysine residues present in this peptide are responsible for the strong affinity. The last 15 amino acids of P90 (sequence: ^{1204}APTAPRPPVQPPKKA^{1218}) contains the same number of positively charged residues and six proline residues (oppose to seven in P1). It is likely that the C-terminal tail of P90 shares the same affinity to heparin, fetuin, and plasminogen as does P1 but that was not evaluated in this work. The P1 peptide was coupled to Avidin Agarose to identify host proteins that interact with it. The observation that this peptide could recover cytokeratin 7, cytokeratin 8, cytokeratin 18, and vimentin from native lysates of human lung epithelial (A549) cells is both a novel and ground-breaking observation. Vimentin has been implicated as a target for *E. coli* during invasion into brain endothelial cells[482]. Further work is required to verify binding between peptide and these cytoskeletal proteins. It is possible that the P1 adhesin targets these cytoskeletal proteins during or after invasion of host cells.

Overall, post-translational proteolytic processing of dominant adhesins remains poorly understood. However, this work suggests that the adhesins and cleavage fragments generated from proteolysis can interact with multiple host molecules, expanding their function. Cleavage fragments may also function as immune decoy molecules. The P1 adhesin is known to be highly immunogenic and is often detected in *M. pneumoniae* infected patient sera[159,422,423]. Schurwanz et al. mapped immunoreactive sites within P1 and showed that they resided in regions that were separate to those required for adherence[93]. It is conceivable that P1 fragments function by binding host immunoglobulins and promote bacterial adherence. Schurwanz et al. also demonstrated that the carboxyl half of P1 was highly immunogenic particularly the C-terminal 341 amino acids (recombinant proteins RP14 and RP15 from[93])[93]. It is possible that the C-terminal tail of P1, which lacks any predicted transmembrane domains, is cleaved off and released into the extracellular milieu where it may bind to and block host antibodies. A similar scenario

has been observed in *M. genitalium*, a close relative of *M. pneumoniae*, where Protein M has been shown to display affinity to host antibodies[434]. This is also seen in a range of Gram-positive bacteria (reviewed in[483]). *S.aureus*, *Streptococcus pyogenes*, and group B streptococci have been shown to bind to the tail (Fc region) of antibodies thereby blocking the interaction between antibodies and host immune cells[483]. Additional experiments are required to determine the affinity of cleavage fragments and the C-terminal tail of P1 to host antibodies.

7.5 Concluding remarks

This thesis investigated the proteins involved during *M. pneumoniae* colonisation of the human respiratory tract. *M. pneumoniae* is one of the major causes of bacterial pneumonia in close contact settings such as schools and hospitals. A successful vaccine against this respiratory pathogen is yet to be developed and treatment options are limited. Additionally, children are limited to one class of antibiotics due to the adverse long-term side effects of other agents. The work presented within this thesis investigated the functional proteome of *M. pneumoniae*, with the goal of discovering potential novel therapeutic or vaccine targets. Data presented in this thesis has provided novel insight for vaccine development. By mapping proteolytic processing events, one can gain insight into how different regions within surface molecules may be critical for adherence while avoiding other regions of the same molecule that might promote adverse immune responses.

Investigating the surface proteome of *M. pneumoniae* identified proteins that were originally thought to solely occupy an intracellular location. These essential proteins appear to function as auxiliary adhesins and may be useful in vaccine and other therapeutic formulations. Is the efficacy of current antibiotics reduced due to inhibiting surface exposed targets, as opposed to intracellular essential proteins? Many microbial vaccines suffer because they are serotype or lineage-specific. The incorporation of functionally conserved proteins may assist in generating cross-protection. Processing of dedicated adhesins such as P1, and key accessory proteins such as Mpn142 has assisted with understanding the structure and function of these molecules. The studies presented in this thesis suggest that different regions of these molecules may be most suitable as therapeutic targets, while others may promote excessive inflammation and detract from

their suitability as a vaccine antigen. Cleavage fragment analysis also assists with the functional identification of binding motifs.

By combining the analysis of the surface proteome with the analysis of host binding proteins, a list of 149 potential proteins that function as adhesins on the *M. pneumoniae* cell surface was generated (Appendix 3). Proteolysis of surface proteins in *M. pneumoniae* provides a mechanism to generate functioning fragments by exposing new interaction sites and motifs, a phenomenon I would like to introduce as Proteolysis Induced Moonlighting. These new sites or motifs may expand on the moonlighting capabilities of surface accessible proteins. The data generated in this thesis strongly suggests that post-translational proteolysis is not limited to the three proteins that were studied intensively here, and is a widespread phenomenon.

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Chapter 8.

Appendices

ithreeinstitute

8.1 Appendix 1: Surface proteome of *M. pneumoniae*

Appendix 1: Table 1: Proteins identified in Biotinylation and surface shaving experiments of *M. pneumoniae*.

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Shave	Biotin
P0CJ81	Proline-rich P65 protein	*p65*	405	47.0	4.3	Y	Y
P11311	Adhesin P1	*mgpA*	1635	176.2	8.5	Y	Y
P22447	DNA gyrase subunit B	*gyrB*	650	73.8	5.7	Y	Y
P23568	Elongation factor Tu	*tuf*	394	43.1	6.1	Y	Y
P41204	50S ribosomal protein L16	*rplP*	139	15.6	10.2	Y	Y
P46775	30S ribosomal protein S4	*rpsD*	205	23.8	10.7	Y	Y
P53527	Probable ribose-5-phosphate isomerase B	*rpiB*	150	16.7	5.7	-	Y
P54125	Oligoendopeptidase F homolog	*pepF*	611	70.9	8.8	Y	Y
P75033	Uncharacterized ATP-dependent helicase MG140 homolog	*MPN_153*	1113	130.3	8.2	Y	-
P75044	Ribose-phosphate pyrophosphokinase	*prs*	328	36.7	8.2	Y	Y
P75049	Protein MG054 homolog	*MPN_067*	320	36.1	5.2	Y	Y
P75053	Purine nucleoside phosphorylase deoD-type	*deoD*	238	26.3	6.7	Y	Y
P75054	Signal recognition particle protein	*ffh*	450	50.1	9.3	Y	-
P75061	Phosphocarrier protein HPr	*ptsH*	88	9.5	8.0	Y	Y
P75062	Uncharacterized lipoprotein MG040 homolog	*MPN_052*	657	71.6	9.3	-	Y
P75063	Uncharacterized protein MG039 homolog	*MPN_051*	384	42.7	9.0	Y	Y
P75064	Glycerol kinase	*glpK*	508	56.6	8.7	Y	Y
P75070	Thymidine kinase	*tdk*	191	21.5	6.4	Y	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Shave	Biotin
P75085	Elongation factor P	*efp*	190	21.8	5.1	Y	Y
P75089	Fructose-bisphosphate aldolase	*fba*	288	31.0	6.4	Y	Y
P75090	Probable DNA-directed RNA polymerase subunit delta	*rpoE*	144	17.2	4.3	-	Y
P75092	Putative proline iminopeptidase	*pip*	309	34.7	7.7	Y	Y
P75109	Uncharacterized ABC transporter permease MG468 homolog	*MPN_684*	1882	209.3	8.2	Y	Y
P75114	Glutamyl-tRNA synthetase	*gltX*	484	55.6	8.6	Y	Y
P75120	ATP-dependent zinc metalloprotease	*ftsH*	709	77.7	9.1	Y	Y
P75129	Peptide methionine sulfoxide reductase	*msrB*	151	17.3	7.0	-	Y
P75131	30S ribosomal protein S16	*rpsP*	88	10.0	10.2	-	Y
P75159	Putative type I restriction enzyme specificity protein	*MPN_638*	375	42.6	8.5	-	Y
P75161	Ribosome-recycling factor	*frr*	184	21.7	7.8	-	Y
P75167	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	*gpmI*	508	56.3	6.1	Y	Y
P75168	Phosphoenolpyruvate-protein phosphotransferase	*ptsI*	572	63.9	7.7	Y	Y
P75170	Uncharacterized protein MG427 homolog	*MPN_625*	141	15.5	6.5	Y	Y
P75179	30S ribosomal protein S9	*rpsI*	132	15.1	10.5	Y	Y
P75184	Uncharacterized lipoprotein MG412 homolog	*MPN_611*	385	41.0	9.4	-	Y
P75189	Enolase	*eno*	456	49.2	6.1	Y	Y
P75205	10 kDa chaperonin	*groS*	116	12.6	9.0	Y	Y
P75206	Probable cytosol aminopeptidase	*pepA*	445	48.8	8.5	Y	Y
P75223	Uncharacterized protein MG377 homolog	*MPN_555*	193	22.4	5.6	Y	Y
P75231	Uncharacterized protein MG369 homolog	*MPN_547*	558	62.4	8.4	Y	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Shave	Biotin
P75237	30S ribosomal protein S20	*rpsT*	56	10.0	11.6	-	Y
P75239	50S ribosomal protein L7/L12	*rplL*	122	13.1	4.9	Y	Y
P75240	50S ribosomal protein L10	*rplJ*	161	17.6	9.7	Y	Y
P75245	Acetate kinase	*ackA*	390	43.7	8.1	Y	Y
P75247	Chaperone protein ClpB	*clpB*	715	81.3	8.1	Y	Y
P75248	Uncharacterized protein MG354 homolog	*MPN_530*	136	15.6	5.2	Y	Y
P75249	Uncharacterized protein MG353 homolog	*MPN_529*	109	12.4	10.1	Y	Y
P75250	Inorganic pyrophosphatase	*ppa*	184	21.4	5.5	-	Y
P75254	UPF0134 protein MPN_524	*MPN_524*	168	20.1	8.8	-	Y
P75255	Uncharacterized lipoprotein MG348 homolog	*MPN_523*	305	33.5	9.2	Y	Y
P75259	UPF0134 protein MPN_139	*MPN_139*	163	19.1	9.2	-	Y
P75265	Uncharacterized lipoprotein MG186 homolog	*MPN_133*	286	33.2	6.5	Y	Y
P75269	Uncharacterized protein MG343 homolog	*MPN_518*	348	40.7	8.9	Y	-
P75271	DNA-directed RNA polymerase subunit beta	*rpoC*	1290	144.8	9.3	Y	Y
P75277	Uncharacterized protein MPN_509	*MPN_509*	427	49.8	9.1	Y	-
P75280	Uncharacterized lipoprotein MPN_506	*MPN_506*	793	87.4	9.2	Y	Y
P75286	UPF0134 protein MPN_501	*MPN_501*	202	23.2	9.1	-	Y
P75295	Uncharacterized protein MPN_491	*MPN_491*	474	52.9	9.1	Y	Y
P75296	Uncharacterized lipoprotein MG338 homolog	*MPN_489*	1300	143.0	9.1	Y	Y
P75305	FMN-dependent NADH-azoreductase	*azoR*	197	21.5	8.7	Y	Y
P75310	Uncharacterized protein MG328 homolog	*MPN_474*	1033	118.1	4.9	Y	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Shave	Biotin
P75313	Putative Xaa-Pro aminopeptidase	*pepP*	354	39.6	8.8	Y	Y
P75327	Uncharacterized lipoprotein MG321 homolog	*MPN_456*	1005	110.4	6.3	-	Y
P75329	Uncharacterized protein MG319 homolog	*MPN_454*	193	21.4	10.2	Y	Y
P75334	Uncharacterized lipoprotein MG309 homolog	*MPN_444*	1289	146.2	7.9	Y	Y
P75342	Uncharacterized lipoprotein MG307 homolog	*MPN_436*	1244	139.0	9.1	Y	Y
P75344	Chaperone protein dnaK	*dnaK*	595	65.1	5.4	Y	Y
P75353	Uncharacterized protein MG202 homolog	*MPN_121*	121	13.3	4.9	Y	Y
P75354	DnaJ-like protein MG200 homolog (TopJ accessory protein)	*MPN_119*	910	100.1	4.2	Y	Y
P75358	Glyceraldehyde-3-phosphate dehydrogenase	*gapA*	337	36.8	8.9	Y	Y
P75359	Phosphate acetyltransferase	*pta*	320	35.2	6.9	Y	Y
P75360	Uncharacterized protein MPN_427	*MPN_427*	290	33.2	6.3	Y	Y
P75374	UPF0134 protein MPN_410	*MPN_409*	148	17.4	9.0	-	Y
P75376	Uncharacterized lipoprotein MPN_408	*MPN_408*	760	83.3	9.2	Y	Y
P75383	Uncharacterized protein MG281 homolog	*MPN_400*	582	65.9	7.1	Y	Y
P75385	Uncharacterized protein MG279 homolog	*MPN_398*	218	24.9	9.0	Y	-
P75389	Probable NADH oxidase	*nox*	479	52.8	6.6	Y	Y
P75390	Pyruvate dehydrogenase E1 component subunit alpha	*pdhA*	358	40.6	6.2	Y	Y
P75391	Pyruvate dehydrogenase E1 component subunit beta	*pdhB*	327	35.9	6.5	Y	Y
P75392	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	*pdhC*	402	42.4	5.5	Y	Y
P75393	Dihydrolipoyl dehydrogenase	*pdhD*	457	49.4	8.4	Y	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Shave	Biotin
P75394	Probable lipoate-protein ligase A	*lplA*	339	39.2	6.2	Y	Y
P75395	Uncharacterized protein MG269 homolog	*MPN_387*	358	42.6	5.7	Y	Y
P75396	Uncharacterized protein MG268 homolog	*MPN_386*	229	27.0	8.9	Y	Y
P75399	Uncharacterized protein MG265 homolog	*MPN_383*	282	31.9	6.0	Y	Y
P75405	Uncharacterized protein MPN_376	*MPN_376*	1140	130.3	9.4	Y	Y
P75409	ADP-ribosylating toxin CARDS	*MPN_372*	591	68.0	5.6	Y	Y
P75454	Trigger factor	*tig*	444	51.3	6.1	Y	Y
P75460	Protein nrdI	*nrdI*	153	17.1	7.8	Y	Y
P75461	Ribonucleoside-diphosphate reductase subunit beta	*nrdF*	339	39.4	5.5	Y	Y
P75466	Ribosomal RNA small subunit methyltransferase H	*rsmH*	308	35.2	9.2	Y	Y
P75467	Protein MraZ	*mraZ*	141	16.3	5.9	Y	Y
P75470	Uncharacterized protein MG218.1 homolog (P41 accessory protein)	*MPN_311*	357	40.6	4.8	Y	Y
P75471	Cytadherence high molecular weight protein 2	*hmw2*	1818	215.5	9.0	Y	Y
P75477	Segregation and condensation protein B	*scpB*	208	23.7	8.5	-	Y
P75482	Uncharacterized protein MG210.1 homolog	*MPN_295*	220	25.7	7.7	-	Y
P75489	Uncharacterized lipoprotein MPN_288	*MPN_288*	787	86.8	9.1	Y	Y
P75493	Uncharacterized lipoprotein MPN_284	*MPN_284*	794	87.1	9.2	-	Y
P75497	Ribonuclease J	*MPN_280*	569	64.0	7.6	Y	Y
P75504	Uncharacterized 16.1 kDa HIT-like protein	*MPN_273*	144	16.1	6.6	Y	Y
P75509	Uncharacterized protein MG127 homolog	*MPN_266*	145	16.8	9.8	-	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Shave	Biotin
P75512	Thioredoxin	*trxA*	102	11.2	7.7	Y	Y
P75518	Uncharacterized protein MG117 homolog	*MPN_256*	223	25.7	4.4	Y	Y
P75526	Guanylate kinase	*gmk*	189	27.0	7.7	Y	Y
P75529	Ribonuclease R	*rnr*	726	83.2	5.6	Y	Y
P75538	Uncharacterized lipoprotein MG095 homolog	*MPN_233*	454	49.8	9.1	-	Y
P75539	Replicative DNA helicase	*dnaB*	473	54.5	5.6	Y	Y
P75542	Single-stranded DNA-binding protein	*ssb*	166	18.4	4.3	-	Y
P75544	Elongation factor G	*fusA*	688	76.5	5.5	Y	Y
P75545	30S ribosomal protein S7	*rpsG*	155	17.9	10.5	Y	Y
P75550	50S ribosomal protein L11	*rplK*	137	14.8	9.7	Y	Y
P75551	Oligopeptide transport ATP-binding protein OppF	*oppF*	851	98.6	9.2	Y	-
P75556	Uncharacterized protein MG075 homolog (P116 adhesin)	*MPN_213*	1030	115.9	5.9	Y	Y
P75558	UvrABC system protein B	*uvrB*	657	75.1	8.5	Y	Y
P75559	Protein translocase subunit SecA	*secA*	808	91.8	7.0	Y	-
P75560	30S ribosomal protein S2	*rpsB*	294	33.4	9.2	Y	Y
P75563	Phenylalanyl-tRNA synthetase beta chain	*pheT*	805	91.7	8.4	Y	Y
P75569	PTS system glucose-specific EIICBA component	*ptsG*	940	101.6	9.4	Y	Y
P75576	30S ribosomal protein S19	*rpsS*	87	10.0	10.5	-	Y
P75577	50S ribosomal protein L2	*rplB*	287	31.9	10.4	Y	Y
P75578	50S ribosomal protein L23	*rplW*	237	25.8	10.1	-	Y
P75579	50S ribosomal protein L4	*rplD*	212	23.6	10.1	-	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Shave	Biotin
P75580	50S ribosomal protein L3	*rplC*	287	31.2	9.9	Y	Y
P75581	30S ribosomal protein S10	*rpsJ*	108	12.2	10.0	-	Y
P75583	Uncharacterized lipoprotein MG149 homolog	*MPN_162*	320	36.1	6.2	-	Y
P75588	Uncharacterized protein MG144 homolog	*MPN_157*	402	44.3	10.0	Y	Y
P75590	Translation initiation factor IF-2	*infB*	617	67.9	8.0	Y	-
P75591	Transcription termination/antitermination protein NusA	*nusA*	540	60.2	5.9	Y	Y
P75610	Uncharacterized lipoprotein MPN_083	*MPN_083*	533	59.9	7.7	Y	-
P78007	L-lactate dehydrogenase	*ldh*	312	33.9	8.4	Y	Y
P78009	Elongation factor Ts	*tsf*	298	33.6	7.7	Y	Y
P78010	Triosephosphate isomerase	*tpiA*	244	26.9	7.0	Y	-
P78012	60 kDa chaperonin	*groL*	543	58.1	5.6	Y	Y
P78013	DNA-directed RNA polymerase subunit beta	*rpoB*	1391	155.5	6.1	Y	Y
P78017	Protein grpE	*grpE*	217	24.7	7.8	Y	Y
P78018	Phosphoglycerate kinase	*pgk*	409	44.2	7.1	Y	Y
P78019	Transcription elongation factor greA	*greA*	160	18.1	4.9	Y	Y
P78022	RNA polymerase sigma factor	*rpoD*	499	57.8	8.9	Y	Y
P78025	Lon protease	*lon*	795	90.1	7.6	Y	-
P78027	Ribonucleoside-diphosphate reductase subunit alpha	*nrdE*	721	82.3	6.6	Y	Y
P78031	Pyruvate kinase	*pyk*	508	57.2	9.2	Y	Y
P78032	DNA topoisomerase 1	*topA*	711	81.9	9.3	Y	Y
P78035	50S ribosomal protein L1	*rplA*	226	24.3	9.6	Y	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	Shave	Biotin
Q50288	Uncharacterized lipoprotein MPN_200	*MPN_200*	798	87.6	9.4	Y	Y
Q50289	Uncharacterized lipoprotein MPN_199	*MPN_199*	760	84.2	9.3	Y	Y
Q50293	Energy-coupling factor transporter ATP-binding protein EcfA 2	*ecfA2*	303	34.3	8.2	Y	Y
Q50295	DNA-directed RNA polymerase subunit alpha	*rpoA*	327	36.6	7.0	Y	Y
Q50296	30S ribosomal protein S11	*rpsK*	121	12.7	10.3	-	Y
Q50298	Translation initiation factor IF-1	*infA*	78	9.0	9.9	-	Y
Q50301	30S ribosomal protein S5	*rpsE*	219	24.1	11.0	Y	Y
Q50303	50S ribosomal protein L6	*rplF*	184	20.6	9.9	Y	Y
Q50313	DNA polymerase III subunit beta	*dnaN*	380	43.8	5.8	Y	Y
Q50314	ParA family protein	*MPN_688*	270	30.1	7.0	-	Y
Q50329	ATP synthase subunit alpha	*atpA*	518	57.3	6.2	Y	Y
Q50331	ATP synthase subunit beta	*atpD*	475	52.2	5.5	Y	Y
Q50335	Uncharacterized lipoprotein MPN_592	*MPN_592*	521	59.5	7.2	Y	-
Q50341	Mgp-operon protein 3	*MPN_142*	1218	130.4	8.0	Y	Y
Q50360	Cytadherence high molecular weight protein 3	*hmw3*	672	73.7	4.7	Y	Y
Q50365	Cytadherence high molecular weight protein 1	*hmw1*	1018	112.1	4.0	Y	Y
Q9EXD3	Uncharacterized protein MPN_377	*MPN_377*	74	9.1	6.4	Y	Y
Q9EXD8	Ascorbate-specific phosphotransferase enzyme IIB component	*ulaB*	95	10.3	7.9	-	Y

Y under 'Shave' and 'Biotin' columns indicates that the protein was identified by that methodology. Proteins listed had a minimum MASCOT protein score of 75 and at least identified twice from six replicates (three from shaving, three from biotinylation).

Appendix 1: Table 2: Bioinformatic predictions of the proteins identified on the surface of *M. pneumoniae*.

UniProt #	Name	SignalP	SecretomeP	TMpred	UniProt Localisation	PSORT Localisation
P0CJ81	Proline-rich P65 protein	No	Yes	0	Cell membrane	Cytoplasmic Membrane
P11311	Adhesin P1	Yes	Yes	6	Cell membrane	Unknown
P22447	DNA gyrase subunit B	No	No	1	Unknown	Cytoplasmic
P23568	Elongation factor Tu	No	No	0	Cytoplasm	Cytoplasmic
P41204	50S ribosomal protein L16	No	Yes	0	Unknown	Cytoplasmic
P46775	30S ribosomal protein S4	No	No	0	Unknown	Unknown
P53527	Probable ribose-5-phosphate isomerase B	No	No	1	Unknown	Cytoplasmic
P54125	Oligoendopeptidase F homolog	No	Yes	0	Unknown	Cytoplasmic
P75033	Uncharacterized ATP-dependent helicase MG140 homolog	No	No	0	Unknown	Cytoplasmic
P75044	Ribose-phosphate pyrophosphokinase	No	No	0	Cytoplasm	Cytoplasmic
P75049	Protein MG054 homolog	No	No	0	Unknown	Cytoplasmic
P75053	Purine nucleoside phosphorylase deoD-type	No	No	1	Unknown	Cytoplasmic
P75054	Signal recognition particle protein	No	No	0	Unknown	Cytoplasmic Membrane
P75061	Phosphocarrier protein HPr	No	No	0	Cytoplasm	Cytoplasmic
P75062	Uncharacterized lipoprotein MG040 homolog	Yes	Yes	3	Cell membrane	Unknown
P75063	Uncharacterized protein MG039 homolog	No	No	1	Unknown	Cytoplasmic
P75064	Glycerol kinase	No	No	1	Unknown	Cytoplasmic
P75070	Thymidine kinase	No	No	0	Cytoplasm	Cytoplasmic

UniProt #	Name	SignalP	SecretomeP	TMpred	UniProt Localisation	PSORT Localisation
P75085	Elongation factor P	No	No	0	Cytoplasm	Cytoplasmic
P75089	Fructose-bisphosphate aldolase	No	No	0	Unknown	Cytoplasmic
P75090	Probable DNA-directed RNA polymerase subunit delta	No	No	0	Unknown	Cytoplasmic
P75092	Putative proline iminopeptidase	No	No	1	Cytoplasm	Cytoplasmic
P75109	Uncharacterized ABC transporter permease MG468 homolog	No	Yes	8	Cell membrane	Cytoplasmic Membrane
P75114	Glutamyl-tRNA synthetase	No	No	0	Cytoplasm	Cytoplasmic
P75120	ATP-dependent zinc metalloprotease	No	No	2	Cell membrane	Cytoplasmic Membrane
P75129	Peptide methionine sulfoxide reductase	No	Yes	0	Unknown	Unknown
P75131	30S ribosomal protein S16	No	Yes	0	Unknown	Cytoplasmic
P75159	Putative type I restriction enzyme specificity protein	No	No	1	Unknown	Unknown
P75161	Ribosome-recycling factor	No	No	0	Cytoplasm	Cytoplasmic
P75167	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	No	No	0	Unknown	Cytoplasmic
P75168	Phosphoenolpyruvate-protein phosphotransferase	No	No	0	Cytoplasm	Cytoplasmic
P75170	Uncharacterized protein MG427 homolog	No	No	1	Unknown	Unknown
P75179	30S ribosomal protein S9	No	Yes	0	Unknown	Cytoplasmic
P75184	Uncharacterized lipoprotein MG412 homolog	Yes	Yes	1	Cell membrane	Unknown
P75189	Enolase	No	No	1	Cell surface, Cytoplasm, Secreted	Cytoplasmic

UniProt #	Name	SignalP	SecretomeP	TMpred	UniProt Localisation	PSORT Localisation
P75205	10 kDa chaperonin	No	Yes	0	Cytoplasm	Unknown
P75206	Probable cytosol aminopeptidase	No	No	1	Cytoplasm	Cytoplasmic
P75223	Uncharacterized protein MG377 homolog	No	No	0	Unknown	Cytoplasmic
P75231	Uncharacterized protein MG369 homolog	No	No	1	Unknown	Cytoplasmic
P75237	30S ribosomal protein S20	No	Yes	0	Unknown	Unknown
P75239	50S ribosomal protein L7/L12	No	No	1	Unknown	Cytoplasmic
P75240	50S ribosomal protein L10	No	No	1	Unknown	Cytoplasmic
P75245	Acetate kinase	No	No	0	Cytoplasm	Cytoplasmic
P75247	Chaperone protein ClpB	No	No	1	Cytoplasm	Cytoplasmic
P75248	Uncharacterized protein MG354 homolog	No	No	1	Unknown	Cytoplasmic
P75249	Uncharacterized protein MG353 homolog	No	Yes	0	Unknown	Cytoplasmic
P75250	Inorganic pyrophosphatase	No	No	1	Cytoplasm	Cytoplasmic
P75254	UPF0134 protein MPN_524	No	No	0	Unknown	Cytoplasmic
P75255	Uncharacterized lipoprotein MG348 homolog	Yes	Yes	1	Cell membrane	Unknown
P75259	UPF0134 protein MPN_139	No	Yes	0	Unknown	Unknown
P75265	Uncharacterized lipoprotein MG186 homolog	No	Yes	1	Cell membrane	Extracellular
P75269	Uncharacterized protein MG343 homolog	No	No	0	Unknown	Cytoplasmic Membrane
P75271	DNA-directed RNA polymerase subunit beta	No	No	0	Unknown	Cytoplasmic
P75277	Uncharacterized protein MPN_509	No	No	1	Unknown	Unknown
P75280	Uncharacterized lipoprotein MPN_506	Yes	Yes	2	Cell membrane	Unknown
P75286	UPF0134 protein MPN_501	No	Yes	0	Unknown	Unknown

UniProt #	Name	SignalP	SecretomeP	TMpred	UniProt Localisation	PSORT Localisation
P75295	Uncharacterized protein MPN_491	No	Yes	1	Cell membrane	Unknown
P75296	Uncharacterized lipoprotein MG338 homolog	No	Yes	2	Cell membrane	Unknown
P75305	FMN-dependent NADH-azoreductase	No	Yes	3	Unknown	Unknown
P75310	Uncharacterized protein MG328 homolog	No	Yes	0	Unknown	Cytoplasmic
P75313	Putative Xaa-Pro aminopeptidase	No	No	0	Unknown	Cytoplasmic
P75327	Uncharacterized lipoprotein MG321 homolog	Yes	Yes	4	Cell membrane	Unknown
P75329	Uncharacterized protein MG319 homolog	No	Yes	2	Cell membrane	Cytoplasmic Membrane
P75334	Uncharacterized lipoprotein MG309 homolog	Yes	Yes	1	Cell membrane	Unknown
P75342	Uncharacterized lipoprotein MG307 homolog	Yes	Yes	2	Cell membrane	Extracellular
P75344	Chaperone protein dnaK	No	No	0	Unknown	Cytoplasmic
P75353	Uncharacterized protein MG202 homolog	No	Yes	0	Unknown	Unknown
P75354	DnaJ-like protein MG200 homolog (TopJ accessory protein)	No	Yes	0	Unknown	Cytoplasmic
P75358	Glyceraldehyde-3-phosphate dehydrogenase	No	No	0	Cytoplasm	Cytoplasmic
P75359	Phosphate acetyltransferase	No	No	1	Cytoplasm	Unknown
P75360	Uncharacterized protein MPN_427	No	Yes	0	Unknown	Cytoplasmic
P75374	UPF0134 protein MPN_410	No	Yes	0	Unknown	Unknown
P75376	Uncharacterized lipoprotein MPN_408	Yes	Yes	3	Cell membrane	Unknown
P75383	Uncharacterized protein MG281 homolog	No	Yes	1	Cell membrane	Cytoplasmic
P75385	Uncharacterized protein MG279 homolog	Yes	Yes	2	Cell membrane	Unknown
P75389	Probable NADH oxidase	No	No	3	Unknown	Cytoplasmic

UniProt #	Name	SignalP	SecretomeP	TMpred	UniProt Localisation	PSORT Localisation
P75390	Pyruvate dehydrogenase E1 component subunit alpha	No	No	1	Unknown	Cytoplasmic
P75391	Pyruvate dehydrogenase E1 component subunit beta	No	No	2	Unknown	Cytoplasmic
P75392	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	No	No	2	Unknown	Cytoplasmic
P75393	Dihydrolipoyl dehydrogenase	No	No	3	Cytoplasm	Cytoplasmic
P75394	Probable lipoate-protein ligase A	No	No	0	Cytoplasm	Cytoplasmic
P75395	Uncharacterized protein MG269 homolog	No	Yes	0	Unknown	Cytoplasmic
P75396	Uncharacterized protein MG268 homolog	No	No	1	Unknown	Cytoplasmic
P75399	Uncharacterized protein MG265 homolog	No	No	0	Unknown	Cytoplasmic
P75405	Uncharacterized protein MPN_376	Yes	Yes	2	Cell membrane	Cytoplasmic Membrane
P75409	ADP-ribosylating toxin CARDS	No	Yes	0	Cell membrane	Cytoplasmic Membrane
P75454	Trigger factor	No	No	0	Unknown	Cytoplasmic
P75460	Protein nrdI	No	No	0	Unknown	Unknown
P75461	Ribonucleoside-diphosphate reductase subunit beta	No	No	1	Unknown	Cytoplasmic
P75466	Ribosomal RNA small subunit methyltransferase H	No	No	0	Cytoplasm	Cytoplasmic
P75467	Protein MraZ	No	No	0	Unknown	Unknown
P75470	Uncharacterized protein MG218.1 homolog (P41 accessory protein)	No	Yes	0	Unknown	Cytoplasmic

UniProt #	Name	SignalP	SecretomeP	TMpred	UniProt Localisation	PSORT Localisation
P75471	Cytadherence high molecular weight protein 2	No	No	0	Unknown	Cytoplasmic
P75477	Segregation and condensation protein B	No	No	1	Cytoplasm	Cytoplasmic
P75482	Uncharacterized protein MG210.1 homolog	No	Yes	0	Unknown	Cytoplasmic
P75489	Uncharacterized lipoprotein MPN_288	Yes	Yes	2	Cell membrane	Unknown
P75493	Uncharacterized lipoprotein MPN_284	Yes	Yes	2	Cell membrane	Unknown
P75497	Ribonuclease J	No	No	0	Cytoplasm	Cytoplasmic
P75504	Uncharacterized 16.1 kDa HIT-like protein	No	No	0	Unknown	Cytoplasmic
P75509	Uncharacterized protein MG127 homolog	No	Yes	0	Unknown	Cytoplasmic
P75512	Thioredoxin	No	No	0	Unknown	Cytoplasmic
P75518	Uncharacterized protein MG117 homolog	No	Yes	0	Unknown	Unknown
P75526	Guanylate kinase	No	No	1	Cytoplasm	Cytoplasmic
P75529	Ribonuclease R	No	No	3	Unknown	Cytoplasmic
P75538	Uncharacterized lipoprotein MG095 homolog	Yes	Yes	1	Cell membrane	Unknown
P75539	Replicative DNA helicase	No	No	0	Unknown	Cytoplasmic
P75542	Single-stranded DNA-binding protein	No	Yes	0	Unknown	Unknown
P75544	Elongation factor G	No	No	1	Cytoplasm	Cytoplasmic
P75545	30S ribosomal protein S7	No	No	0	Unknown	Unknown
P75550	50S ribosomal protein L11	No	No	0	Unknown	Cytoplasmic
P75551	Oligopeptide transport ATP-binding protein OppF	No	No	0	Cell membrane	Cytoplasmic Membrane
P75556	Uncharacterized protein MG075 homolog (P116 adhesin)	Yes	Yes	2	Cell membrane	Unknown

UniProt #	Name	SignalP	SecretomeP	TMpred	UniProt Localisation	PSORT Localisation
P75558	UvrABC system protein B	No	No	0	Cytoplasm	Cytoplasmic
P75559	Protein translocase subunit SecA	No	No	1	Cell membrane; Cytoplasmic side	Cytoplasmic
P75560	30S ribosomal protein S2	No	No	0	Unknown	Cytoplasmic
P75563	Phenylalanyl-tRNA synthetase beta chain	No	No	0	Cytoplasm	Cytoplasmic
P75569	PTS system glucose-specific EIICBA component	No	Yes	10	Cell membrane	Cytoplasmic Membrane
P75576	30S ribosomal protein S19	No	Yes	0	Unknown	Cytoplasmic
P75577	50S ribosomal protein L2	No	No	0	Unknown	Unknown
P75578	50S ribosomal protein L23	No	Yes	0	Unknown	Unknown
P75579	50S ribosomal protein L4	No	Yes	0	Unknown	Unknown
P75580	50S ribosomal protein L3	No	Yes	1	Unknown	Cytoplasmic
P75581	30S ribosomal protein S10	No	No	0	Unknown	Cytoplasmic
P75583	Uncharacterized lipoprotein MG149 homolog	Yes	Yes	1	Cell membrane	Unknown
P75588	Uncharacterized protein MG144 homolog	No	Yes	3	Cell membrane	Cytoplasmic Membrane
P75590	Translation initiation factor IF-2	No	No	1	Cytoplasm	Cytoplasmic
P75591	Transcription termination/antitermination protein NusA	No	Yes	0	Unknown	Cytoplasmic
P75610	Uncharacterized lipoprotein MPN_083	No	Yes	2	Cell membrane	Unknown
P78007	L-lactate dehydrogenase	No	No	1	Cytoplasm	Cytoplasmic
P78009	Elongation factor Ts	No	No	0	Cytoplasm	Cytoplasmic
P78010	Triosephosphate isomerase	No	No	1	Cytoplasm	Cytoplasmic

UniProt #	Name	SignalP	SecretomeP	TMpred	UniProt Localisation	PSORT Localisation
P78012	60 kDa chaperonin	No	No	2	Cytoplasm	Cytoplasmic
P78013	DNA-directed RNA polymerase subunit beta	No	No	1	Unknown	Cytoplasmic
P78017	Protein grpE	No	No	0	Cytoplasm	Cytoplasmic
P78018	Phosphoglycerate kinase	No	No	2	Cytoplasm	Cytoplasmic
P78019	Transcription elongation factor greA	No	No	0	Unknown	Cytoplasmic
P78022	RNA polymerase sigma factor	No	No	0	Unknown	Cytoplasmic
P78025	Lon protease	No	No	0	Cytoplasm	Cytoplasmic
P78027	Ribonucleoside-diphosphate reductase subunit alpha	No	No	1	Unknown	Cytoplasmic
P78031	Pyruvate kinase	No	No	1	Unknown	Unknown
P78032	DNA topoisomerase 1	No	Yes	0	Unknown	Cytoplasmic
P78035	50S ribosomal protein L1	No	No	0	Unknown	Cytoplasmic
Q50288	Uncharacterized lipoprotein MPN_200	Yes	Yes	3	Cell membrane	Unknown
Q50289	Uncharacterized lipoprotein MPN_199	Yes	Yes	1	Cell membrane	Unknown
Q50293	Energy-coupling factor transporter ATP-binding protein EcfA 2	No	No	0	Cell membrane	Cytoplasmic Membrane
Q50295	DNA-directed RNA polymerase subunit alpha	No	No	2	Unknown	Cytoplasmic
Q50296	30S ribosomal protein S11	Yes	Yes	0	Unknown	Cytoplasmic
Q50298	Translation initiation factor IF-1	No	No	0	Cytoplasm	Cytoplasmic
Q50301	30S ribosomal protein S5	No	No	1	Unknown	Cytoplasmic
Q50303	50S ribosomal protein L6	No	No	0	Unknown	Cytoplasmic
Q50313	DNA polymerase III subunit beta	No	No	0	Unknown	Cytoplasmic

UniProt #	Name	SignalP	SecretomeP	TMpred	UniProt Localisation	PSORT Localisation
Q50314	ParA family protein	No	No	1	Cytoplasm	Unknown
Q50329	ATP synthase subunit alpha	No	No	1	Cell membrane	Cytoplasmic
Q50331	ATP synthase subunit beta	No	No	0	Cell membrane	Cytoplasmic Membrane
Q50335	Uncharacterized lipoprotein MPN_592	Yes	Yes	1	Cell membrane	Unknown
Q50341	Mgp-operon protein 3	Yes	Yes	5	Cell membrane	Unknown
Q50360	Cytadherence high molecular weight protein 3	No	Yes	0	Cell membrane	Unknown
Q50365	Cytadherence high molecular weight protein 1	No	Yes	0	Cell membrane	Unknown
Q9EXD3	Uncharacterized protein MPN_377	No	Yes	0	Unknown	Cytoplasmic
Q9EXD8	Ascorbate-specific phosphotransferase enzyme IIB component	No	No	0	Cytoplasm	Cytoplasmic

The cut off scores for a positive result for the bioinformatics tool used are: 0.450 for SignalP 4.1; 0.5 for SecretomeP 2.0; 500 for TMpred; and 7.5 for PSORTb 3.0.2.

8.2 Appendix 2: 'Bait and Prey' affinity chromatography

Appendix 2: Table 1: *M. pneumoniae* proteins identified in elutions from 'Bait and Prey' affinity chromatography.

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P09924	Deoxyribose-phosphate aldolase	*deoC*	224	24.9	8.6	-	Y	Y	Y	Y	Y	-
P0CJ81	Proline-rich P65 protein	*p65*	405	47.0	4.3	-	Y	Y	Y	Y	Y	-
P11311	Adhesin P1	*mgpA*	1627	176.2	8.5	Y	Y	Y	Y	Y	Y	Y
P22446	DNA gyrase subunit A	*gyrA*	839	93.3	8.5	Y	Y	Y	Y	Y	Y	Y
P22447	DNA gyrase subunit B	*gyrB*	650	73.8	5.7	Y	Y	Y	Y	Y	Y	Y
P23568	Elongation factor Tu	*tuf*	394	43.1	6.1	Y	Y	Y	Y	Y	Y	Y
P41205	30S ribosomal protein S3	*rpsC*	273	30.6	10.2	-	Y	-	Y	Y	Y	-
P46775	30S ribosomal protein S4	*rpsD*	205	23.8	10.7	Y	Y	-	Y	Y	-	-
P53527	Probable ribose-5-phosphate isomerase B	*rpiB*	150	16.7	5.7	Y	-	-	Y	Y	-	-
P54125	Oligoendopeptidase F homolog	*pepF*	611	70.9	8.8	Y	Y	-	Y	Y	Y	Y
P75033	Uncharacterized ATP-dependent helicase MG140 homolog	*MPN_153*	1113	130.3	8.2	-	Y	-	Y	Y	Y	-
P75035	UPF0134 protein MPN_151	*MPN_151*	133	15.5	5.9	-	-	-	-	Y	-	-
P75038	Putative 1-phosphofructokinase	*fruK*	300	33.6	5.6	-	Y	-	Y	-	-	-
P75041	Uncharacterized protein MG061 homolog 1	*MPN_076*	564	61.6	9.4	-	-	-	-	Y	-	-
P75044	Ribose-phosphate pyrophosphokinase	*prs*	328	36.7	8.2	Y	Y	-	Y	Y	-	-
P75046	Ribosomal RNA small subunit methyltransferase I	*rsmI*	276	31.0	6.9	-	-	-	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75049	Transcription termination/antitermination protein NusG	*nusG*	320	36.1	5.2	Y	Y	Y	Y	Y	Y	Y
P75050	Phosphomannomutase (CpsG accessory protein)	*manB*	554	63.2	8.1	Y	Y	-	Y	Y	-	Y
P75052	Thymidine phosphorylase	*deoA*	421	46.6	9.0	-	Y	-	-	Y	Y	-
P75053	Purine nucleoside phosphorylase DeoD-type	*deoD*	238	26.3	6.7	Y	Y	Y	Y	Y	Y	-
P75054	Signal recognition particle protein	*ffh*	450	50.1	9.3	-	Y	-	Y	Y	-	-
P75056	Uncharacterized lipoprotein MG045 homolog	*MPN_058*	470	55.0	5.2	-	Y	-	Y	-	Y	-
P75059	Spermidine/putrescine import ATP-binding protein PotA	*potA*	560	65.1	9.1	-	Y	-	-	-	-	Y
P75061	Phosphocarrier protein HPr	*ptsH*	88	9.5	8.0	Y	-	-	-	Y	-	-
P75062	Uncharacterized lipoprotein MG040 homolog	*MPN_052*	657	71.6	9.3	Y	Y	Y	Y	Y	Y	Y
P75063	Uncharacterized protein MG039 homolog	*MPN_051*	384	42.7	9.0	-	-	Y	-	Y	-	-
P75064	Glycerol kinase	*glpK*	508	56.6	8.7	Y	Y	Y	Y	Y	Y	Y
P75068	Aspartate--tRNA ligase	*aspS*	557	64.1	8.7	-	Y	-	-	-	-	-
P75069	Histidine--tRNA ligase	*hisS*	414	47.2	6.6	-	-	-	-	Y	-	Y
P75070	Thymidine kinase	*tdk*	224	21.5	6.4	-	-	-	-	Y	-	-
P75078	Uncharacterized protein MPN_036	*MPN_036*	673	76.6	8.6	-	-	-	Y	-	-	-
P75080	DNA polymerase III PolC-type	*polC*	1443	165.0	7.0	-	-	-	-	Y	-	Y
P75081	Uracil phosphoribosyltransferase	*upp*	206	22.7	9.1	-	Y	-	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75088	Probable GTP-binding protein MG024 homolog	*MPN_026*	366	40.6	6.5	-	-	-	-	Y	-	-
P75089	Fructose-bisphosphate aldolase	*fba*	288	31.0	6.4	Y	Y	Y	Y	Y	Y	-
P75090	Probable DNA-directed RNA polymerase subunit delta	*rpoE*	144	17.2	4.3	-	Y	-	Y	Y	Y	-
P75091	Methionine--tRNA ligase	*metG*	512	59.2	6.6	-	-	-	Y	Y	-	-
P75092	Putative proline iminopeptidase	*pip*	309	34.7	7.7	Y	-	-	Y	Y	-	-
P75093	Uncharacterized ATP-dependent helicase MPN_020	*MPN_020*	1030	119.5	8.4	-	Y	-	-	Y	-	Y
P75100	UPF0134 protein MPN_013	*MPN_013*	257	29.1	5.4	-	-	-	-	Y	-	-
P75106	Thymidylate kinase	*tmk*	210	23.7	6.8	Y	Y	Y	-	Y	-	-
P75107	Serine--tRNA ligase	*serS*	420	47.9	8.7	Y	Y	Y	Y	Y	-	Y
P75109	Uncharacterized ABC transporter permease MG468 homolog	*MPN_684*	1882	209.3	8.2	-	Y	-	-	Y	-	Y
P75113	Ribosomal RNA small subunit methyltransferase A	*rsmA*	263	29.8	9.5	-	Y	-	-	Y	-	Y
P75114	Glutamate--tRNA ligase	*gltX*	484	55.6	8.6	-	-	-	Y	Y	-	Y
P75115	Uncharacterized protein MG461 homolog	*MPN_677*	425	50.6	8.7	-	-	-	-	Y	-	-
P75117	UPF0134 protein MPN_675	*MPN_675*	138	11.9	7.9	-	-	-	-	Y	-	-
P75118	Uncharacterized protein MG459 homolog	*MPN_673*	169	19.5	9.2	Y	-	-	-	Y	Y	-
P75119	Hypoxanthine-guanine phosphoribosyltransferase	*hpt*	175	19.6	6.1	-	-	-	Y	-	-	-
P75120	ATP-dependent zinc metalloprotease	*ftsH*	709	77.7	9.1	Y	Y	Y	Y	Y	Y	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75121	Uncharacterized protein MG456 homolog	*MPN_670*	345	39.2	9.1	-	Y	-	Y	-	-	-
P75122	Tyrosine--tRNA ligase	*tyrS*	399	44.6	9.5	-	Y	-	-	Y	-	-
P75123	Organic hydroperoxide resistance protein-like	*MPN_668*	140	14.9	6.6	Y	-	Y	Y	-	-	-
P75124	UTP--glucose-1-phosphate uridylyltransferase	*galU*	291	32.1	6.3	-	-	-	Y	-	-	-
P75127	DegV domain-containing protein MG450 homolog	*MPN_664*	237	26.8	9.1	-	-	Y	-	Y	-	Y
P75129	Peptide methionine sulfoxide reductase	*msrB*	151	17.3	7.0	-	-	-	-	Y	-	-
P75131	30S ribosomal protein S16	*rpsP*	88	10.0	10.2	Y	-	-	-	Y	-	-
P75132	tRNA (guanine-N(1)-)-methyltransferase	*trmD*	231	26.3	7.0	-	-	-	-	Y	-	-
P75133	50S ribosomal protein L19	*rplS*	119	13.8	10.5	-	Y	-	-	Y	-	-
P75144	Bifunctional oligoribonuclease and PAP phosphatase NrnA	*nrnA*	324	37.1	8.8	-	Y	-	Y	-	Y	-
P75159	Putative type I restriction enzyme specificity protein	*MPN_638*	375	42.6	8.5	-	Y	-	Y	Y	-	Y
P75161	Ribosome-recycling factor	*frr*	185	21.7	7.8	Y	Y	Y	-	Y	-	-
P75165	Uridylate kinase	*pyrH*	235	25.4	8.7	-	Y	-	-	Y	-	-
P75167	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	*gpmI*	508	56.3	6.1	Y	-	-	-	Y	-	-
P75168	Phosphoenolpyruvate-protein phosphotransferase	*ptsI*	572	63.9	7.7	-	Y	Y	Y	Y	-	-
P75170	Hydroperoxide reductase	*MPN_625*	141	15.5	6.5	Y	-	-	Y	Y	Y	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75173	30S ribosomal protein S15	*rpsO*	86	9.9	10.2	-	-	-	-	Y	-	-
P75174	Uncharacterized protein MG423 homolog	*MPN_621*	561	63.3	8.8	Y	Y	Y	Y	Y	-	Y
P75176	UvrABC system protein A	*uvrA*	948	105.4	8.2	-	Y	-	-	Y	-	-
P75177	DNA polymerase III subunit gamma/tau	*dnaX*	681	76.2	6.9	-	Y	-	Y	Y	Y	Y
P75178	50S ribosomal protein L13	*rplM*	141	16.8	9.9	-	Y	-	-	-	-	-
P75179	30S ribosomal protein S9	*rpsI*	132	15.1	10.5	Y	Y	-	-	-	-	-
P75184	Uncharacterized lipoprotein MG412 homolog	*MPN_611*	385	41.0	9.4	-	Y	Y	Y	-	Y	-
P75187	Uncharacterized protein MG409 homolog	*MPN_608*	225	26.8	8.2	-	-	-	-	Y	-	-
P75188	Peptide methionine sulfoxide reductase MsrA	*msrA*	157	18.4	5.6	-	-	-	-	Y	-	-
P75189	Enolase	*eno*	456	49.2	6.1	Y	Y	Y	Y	Y	Y	Y
P75205	10 kDa chaperonin	*groS*	116	12.6	9.0	Y	Y	Y	Y	Y	Y	-
P75206	Probable cytosol aminopeptidase	*pepA*	445	48.8	8.5	Y	Y	Y	Y	Y	Y	-
P75210	GTPase Era	*era*	291	33.5	8.9	-	Y	-	-	-	-	-
P75211	Protein P200	*p200*	1036	116.8	4.1	-	Y	-	-	-	-	-
P75212	Uncharacterized protein MG385 homolog	*MPN_566*	237	27.7	9.2	-	Y	-	-	Y	-	-
P75218	Arginine deiminase-like protein	*MPN_560*	438	49.4	7.8	Y	Y	Y	Y	-	Y	Y
P75220	Ribosomal RNA small subunit methyltransferase G	*rsmG*	191	22.1	7.0	-	-	-	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75221	tRNA uridine 5-carboxymethylaminomethyl modification enzyme	*mnmG*	612	68.0	9.1	-	Y	-	-	Y	-	Y
P75222	Arginine--tRNA ligase	*argS*	537	62.1	7.4	Y	Y	-	-	Y	-	-
P75223	Uncharacterized protein MG377 homolog	*MPN_555*	193	22.4	5.6	Y	Y	Y	Y	Y	Y	Y
P75224	Uncharacterized protein MG376 homolog	*MPN_554*	104	12.5	8.9	-	-	-	-	Y	-	-
P75225	Threonine--tRNA ligase	*thrS*	564	64.9	9.0	Y	Y	-	-	-	Y	-
P75227	Uncharacterized protein MG373 homolog	*MPN_551*	281	32.5	7.6	-	-	-	-	Y	-	Y
P75228	Probable tRNA sulfurtransferase	*thiI*	387	43.3	9.3	-	-	-	-	Y	-	-
P75229	Uncharacterized protein MG371 homolog	*MPN_549*	325	36.7	9.1	-	Y	-	-	Y	Y	-
P75231	Uncharacterized protein MG369 homolog	*MPN_547*	558	62.4	8.4	-	Y	-	Y	Y	-	Y
P75232	Phosphate acyltransferase	*plsX*	328	36.6	8.7	-	Y	-	Y	-	Y	Y
P75233	Ribonuclease 3	*rnc*	282	32.6	9.3	-	-	-	-	Y	-	-
P75234	Uncharacterized protein MG366 homolog	*MPN_544*	664	76.7	7.5	-	-	-	-	Y	-	-
P75236	Uncharacterized protein MG364 homolog	*MPN_542*	218	25.6	10.1	-	Y	-	-	-	-	-
P75237	30S ribosomal protein S20	*rpsT*	87	10.0	11.6	-	-	-	-	Y	-	-
P75239	50S ribosomal protein L7/L12	*rplL*	122	13.1	4.9	Y	Y	Y	Y	Y	Y	-
P75240	50S ribosomal protein L10	*rplJ*	161	17.6	9.7	Y	-	-	Y	Y	-	-
P75245	Acetate kinase	*ackA*	390	43.7	8.1	Y	-	Y	-	Y	-	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75246	Uncharacterized protein MG356 homolog	*MPN_532*	282	33.3	8.7	-	-	-	-	Y	-	-
P75247	Chaperone protein ClpB	*clpB*	715	81.3	8.1	Y	Y	Y	Y	Y	Y	Y
P75248	Uncharacterized protein MG354 homolog	*MPN_530*	136	15.6	5.2	Y	Y	-	Y	Y	Y	-
P75249	Uncharacterized protein MG353 homolog	*MPN_529*	109	12.4	10.1	-	Y	-	-	-	-	-
P75250	Inorganic pyrophosphatase	*ppa*	184	21.4	5.5	-	-	Y	Y	-	Y	-
P75252	Uncharacterized protein MG350 homolog	*MPN_526*	328	39.4	8.4	-	Y	-	-	-	-	-
P75254	UPF0134 protein MPN_524	*MPN_524*	168	20.1	8.8	-	-	-	-	Y	-	Y
P75255	Uncharacterized lipoprotein MG348 homolog	*MPN_523*	305	33.5	9.2	-	Y	Y	Y	Y	Y	-
P75258	Isoleucine--tRNA ligase	*ileS*	861	99.5	9.3	-	Y	-	-	Y	-	Y
P75259	UPF0134 protein MPN_139	*MPN_139*	163	19.1	9.2	-	-	-	-	Y	-	-
P75264	Putative ABC transporter ATP-binding protein MG187 homolog	*MPN_134*	586	66.4	9.1	Y	Y	-	Y	-	Y	Y
P75269	Uncharacterized protein MG343 homolog	*MPN_518*	348	40.7	8.9	-	Y	-	-	-	Y	-
P75270	Uncharacterized protein MG342 homolog	*MPN_517*	166	18.9	7.0	Y	Y	-	Y	Y	Y	Y
P75271	DNA-directed RNA polymerase subunit beta	*rpoC*	1290	144.8	9.3	Y	Y	Y	Y	Y	Y	Y
P75282	UPF0134 protein MPN_504	*MPN_504*	126	15.0	9.1	-	-	-	-	Y	-	-
P75293	Probable 3-keto-L-gulonate-6-phosphate decarboxylase	*ulaD*	218	24.1	5.8	Y	-	-	-	-	-	-
P75294	Probable L-ribulose-5-phosphate 3-epimerase	*ulaE*	292	34.9	8.7	-	-	-	-	Y	-	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75295	Uncharacterized protein MPN_491	*MPN_491*	474	52.9	9.1	-	Y	-	-	Y	-	Y
P75296	Uncharacterized lipoprotein MG338 homolog	*MPN_489*	1300	143.0	9.1	-	-	Y	-	-	-	Y
P75301	UPF0134 protein MPN_484	*MPN_484*	125	11.9	9.3	-	-	-	-	Y	-	-
P75303	Probable GTP-binding protein	*engB*	193	21.8	9.7	-	-	-	-	Y	-	-
P75304	Valine--tRNA ligase	*valS*	838	95.7	9.6	-	Y	-	-	-	-	-
P75305	FMN-dependent NADH-azoreductase	*azoR*	197	21.5	8.7	Y	Y	Y	Y	Y	Y	-
P75306	Probable transcriptional regulatory protein MPN_478	*MPN_478*	235	26.0	5.9	-	Y	-	-	Y	-	-
P75308	Cytidylate kinase	*cmk*	217	24.6	7.9	-	Y	Y	Y	Y	-	-
P75309	GTPase	*der*	449	50.1	9.7	-	Y	-	-	-	-	-
P75310	Uncharacterized protein MG328 homolog	*MPN_474*	1140	118.0	5.0	Y	Y	Y	Y	Y	Y	Y
P75312	DegV domain-containing protein MG326 homolog	*MPN_472*	293	32.8	9.3	-	-	-	-	Y	-	Y
P75313	Putative Xaa-Pro aminopeptidase	*pepP*	354	39.6	8.8	Y	Y	-	Y	Y	Y	-
P75324	Uncharacterized lipoprotein MPN_459	*MPN_459*	569	63.1	5.9	Y	-	-	-	-	-	Y
P75327	Uncharacterized lipoprotein MG321 homolog	*MPN_456*	1005	110.4	6.3	Y	Y	Y	Y	Y	Y	Y
P75329	Uncharacterized protein MG319 homolog	*MPN_454*	193	21.4	10.2	Y	Y	Y	-	Y	-	-
P75330	P30 adhesin	*p30*	274	29.7	9.8	Y	Y	Y	Y	-	Y	-
P75333	Putative esterase/lipase 1	*MPN_445*	269	30.9	8.4	-	-	-	-	Y	-	-
P75334	Uncharacterized lipoprotein MG309 homolog	*MPN_444*	1325	146.2	7.9	Y	Y	Y	Y	Y	-	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75342	Uncharacterized lipoprotein MG307 homolog	*MPN_436*	1244	139.0	9.1	-	Y	Y	Y	Y	-	Y
P75344	Chaperone protein DnaK	*dnaK*	595	65.1	5.4	Y	Y	Y	Y	Y	Y	Y
P75351	Heat-inducible transcription repressor HrcA	*hrcA*	351	40.4	8.7	-	Y	-	-	-	-	-
P75352	DNA topoisomerase 4 subunit A	*parC*	789	88.6	9.0	-	Y	-	-	-	-	-
P75353	Uncharacterized protein MG202 homolog	*MPN_121*	121	13.3	4.9	Y	-	-	-	-	-	-
P75354	DnaJ-like protein MG200 homolog (TopJ accessory protein)	*MPN_119*	910	100.1	4.2	-	Y	Y	-	Y	-	Y
P75358	Glyceraldehyde-3-phosphate dehydrogenase	*gapA*	337	36.8	8.9	Y	Y	Y	Y	Y	Y	Y
P75359	Phosphate acetyltransferase	*pta*	320	35.2	6.9	Y	Y	Y	Y	Y	-	Y
P75360	Uncharacterized protein MPN_427	*MPN_427*	290	33.2	6.3	Y	Y	Y	Y	-	-	-
P75361	Chromosome partition protein Smc	*smc*	982	110.8	5.9	-	Y	Y	Y	-	Y	-
P75362	Signal recognition particle receptor FtsY	*ftsY*	348	38.8	6.9	-	Y	Y	Y	Y	Y	-
P75365	tRNA-specific 2-thiouridylase	*mnmA*	370	41.7	8.8	-	Y	-	Y	Y	-	-
P75367	Uncharacterized protein MG293 homolog	*MPN_420*	241	28.4	6.3	Y	Y	-	Y	Y	-	-
P75368	Alanine--tRNA ligase	*alaS*	900	103.6	6.4	-	Y	-	Y	Y	-	Y
P75376	Uncharacterized lipoprotein MPN_408	*MPN_408*	760	83.3	9.2	-	Y	-	-	-	-	Y
P75378	Acyl carrier protein homolog	*MPN_406*	84	9.8	5.0	-	-	-	Y	Y	-	-
P75382	Proline--tRNA ligase	*proS*	483	55.4	8.7	-	-	-	Y	Y	Y	-
P75383	Uncharacterized protein MG281 homolog	*MPN_400*	582	65.9	7.1	-	Y	Y	Y	-	Y	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75384	Uncharacterized protein MG280 homolog	*MPN_399*	287	31.8	8.7	-	Y	Y	Y	Y	-	Y
P75385	Uncharacterized protein MG279 homolog	*MPN_398*	218	24.9	9.0	-	Y	-	-	Y	-	-
P75387	Uncharacterized protein MG277 homolog	*MPN_396*	971	108.2	9.1	-	Y	-	-	-	-	-
P75388	Adenine phosphoribosyltransferase	*apt*	177	19.9	8.6	Y	Y	-	-	-	Y	-
P75389	Probable NADH oxidase	*nox*	479	52.8	6.6	Y	Y	Y	Y	Y	Y	Y
P75390	Pyruvate dehydrogenase E1 component subunit alpha	*pdhA*	358	40.6	6.2	Y	Y	Y	Y	Y	Y	Y
P75391	Pyruvate dehydrogenase E1 component subunit beta	*pdhB*	327	35.9	6.5	Y	Y	Y	Y	Y	Y	Y
P75392	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	*pdhC*	402	42.4	5.5	Y	-	Y	Y	Y	Y	Y
P75393	Dihydrolipoyl dehydrogenase	*pdhD*	457	49.4	8.4	Y	Y	Y	Y	Y	Y	Y
P75394	Probable lipoate-protein ligase A	*lplA*	339	39.2	6.2	Y	Y	Y	Y	Y	Y	-
P75395	Uncharacterized protein MG269 homolog	*MPN_387*	358	42.6	5.7	Y	Y	-	Y	Y	Y	Y
P75396	Uncharacterized protein MG268 homolog	*MPN_386*	229	27.0	8.9	Y	Y	-	Y	Y	-	-
P75398	Leucine--tRNA ligase	*leuS*	793	91.0	8.1	-	-	-	Y	-	-	-
P75399	Putative phosphatase MPN_383	*MPN_383*	282	31.9	6.0	Y	Y	-	Y	Y	-	-
P75401	Uncharacterized protein MG263 homolog	*MPN_381*	292	33.8	6.2	-	-	-	Y	Y	-	-
P75402	Formamidopyrimidine-DNA glycosylase	*mutM*	277	31.8	9.4	-	-	-	-	Y	-	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75403	5~-3~ exonuclease	*polA*	291	33.0	7.6	-	-	-	Y	Y	-	-
P75404	DNA polymerase III subunit alpha	*dnaE*	872	99.2	6.7	-	Y	-	Y	Y	Y	-
P75405	Uncharacterized protein MPN_376	*MPN_376*	1140	130.3	9.4	-	Y	-	Y	Y	-	Y
P75409	ADP-ribosylating toxin CARDS	*MPN_372*	591	68.0	5.6	Y	Y	Y	Y	-	Y	Y
P75420	Peptide chain release factor 1	*prfA*	359	40.8	6.1	-	Y	Y	Y	Y	-	-
P75424	Uncharacterized tRNA/rRNA methyltransferase MG252 homolog	*MPN_355*	239	27.8	9.7	-	Y	-	-	-	-	-
P75425	Glycine--tRNA ligase	*glyQS*	449	52.3	6.9	-	-	-	Y	Y	Y	-
P75429	Uncharacterized protein MG246 homolog	*MPN_349*	281	31.4	8.8	-	-	-	-	Y	-	-
P75430	Probable 5-formyltetrahydrofolate cyclo-ligase	*MPN_348*	164	19.3	6.6	-	-	-	-	Y	-	-
P75442	Uncharacterized protein MG240 homolog	*MPN_336*	349	40.0	9.8	-	Y	-	-	-	-	-
P75446	Uncharacterized protein MG199 homolog	*MPN_118*	236	26.3	9.4	-	-	-	-	Y	-	-
P75453	Uncharacterized protein MPN_109	*MPN_109*	165	19.8	5.7	-	-	-	-	Y	-	-
P75454	Trigger factor	*tig*	444	51.3	6.1	-	Y	Y	Y	Y	Y	Y
P75456	Uncharacterized protein MG236 homolog	*MPN_329*	158	18.8	9.0	-	Y	-	-	-	Y	-
P75457	Probable endonuclease 4	*nfo*	286	31.9	6.7	-	-	-	-	Y	-	Y
P75459	Uncharacterized protein MG233 homolog	*MPN_326*	100	10.9	6.4	Y	-	-	Y	Y	-	-
P75460	Protein NrdI	*nrdI*	153	17.1	7.8	-	Y	-	Y	Y	-	-
P75461	Ribonucleoside-diphosphate reductase subunit beta	*nrdF*	339	39.4	5.5	-	Y	-	Y	Y	Y	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75464	Cell division protein FtsZ	*ftsZ*	380	42.8	7.2	-	-	-	Y	-	Y	-
P75466	Ribosomal RNA small subunit methyltransferase H	*rsmH*	308	35.2	9.2	Y	Y	Y	Y	-	Y	-
P75467	Protein MraZ	*mraZ*	141	16.3	5.9	Y	-	Y	-	Y	-	-
P75469	Uncharacterized protein (P24 accessory protein)	*MPN_312*	218	24.2	5.9	-	Y	-	-	Y	-	-
P75470	Uncharacterized protein MG218.1 homolog (P41 accessory protein)	*MPN_311*	357	40.6	4.8	Y	Y	Y	Y	-	Y	-
P75471	Cytadherence high molecular weight protein 2	*hmw2*	1818	215.5	9.0	-	Y	Y	-	Y	Y	Y
P75476	6-phosphofructokinase	*pfkA*	328	36.0	7.1	Y	-	-	-	Y	-	Y
P75477	Segregation and condensation protein B	*scpB*	208	23.7	8.5	-	Y	Y	Y	Y	-	-
P75481	Uncharacterized protein MG211 homolog	*MPN_297*	149	17.8	5.0	Y	Y	-	Y	Y	-	-
P75482	Uncharacterized protein MG210.1 homolog	*MPN_295*	220	25.7	7.7	Y	Y	-	Y	Y	Y	Y
P75483	Uncharacterized protein MPN_294	*MPN_294*	206	23.2	9.5	-	Y	-	-	Y	-	-
P75485	Uncharacterized RNA pseudouridine synthase MG209 homolog	*MPN_292*	309	35.2	8.7	-	-	-	-	Y	-	-
P75489	Uncharacterized lipoprotein MPN_288	*MPN_288*	787	86.8	9.1	-	Y	Y	Y	Y	-	Y
P75490	UPF0134 protein MPN_287	*MPN_287*	145	13.4	9.3	-	-	-	-	Y	-	-
P75493	Uncharacterized lipoprotein MPN_284	*MPN_284*	794	87.1	9.2	Y	Y	Y	Y	Y	Y	Y
P75497	Ribonuclease J	*MPN_280*	569	64.0	7.6	Y	Y	-	Y	Y	Y	Y
P75498	Elongation factor 4	*lepA*	598	68.0	7.6	-	-	-	Y	Y	Y	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75500	Lysine--tRNA ligase	*lysS*	489	56.0	6.6	-	Y	-	Y	Y	Y	-
P75502	Uncharacterized protein MG134 homolog	*MPN_275*	100	11.5	4.9	-	-	-	Y	Y	-	-
P75504	Uncharacterized 16.1 kDa HIT-like protein	*MPN_273*	144	16.1	6.6	Y	Y	-	Y	Y	-	-
P75507	Putative phosphotransferase enzyme IIB component	*MPN_268*	117	13.3	10.3	Y	-	-	-	Y	-	-
P75508	NAD kinase	*nadK*	259	29.0	8.2	-	Y	-	Y	Y	Y	Y
P75509	Uncharacterized protein MG127 homolog	*MPN_266*	145	16.8	9.8	Y	Y	-	-	Y	-	-
P75510	Tryptophan--tRNA ligase	*trpS*	345	39.1	9.6	Y	Y	-	Y	Y	-	-
P75511	Uncharacterized protein MG125 homolog	*MPN_264*	281	32.6	9.5	-	Y	-	-	-	-	-
P75512	Thioredoxin	*trxA*	102	11.2	7.7	Y	Y	Y	Y	Y	-	-
P75516	Putative carbohydrate transport ATP-binding protein	*MPN_258*	572	64.8	9.6	Y	Y	-	Y	Y	-	Y
P75517	UDP-glucose 4-epimerase	*galE*	338	38.1	6.8	-	Y	-	Y	-	-	-
P75518	Uncharacterized protein MG117 homolog	*MPN_256*	223	25.7	4.4	-	-	-	Y	Y	-	-
P75519	Uncharacterized protein MG116 homolog	*MPN_255*	251	29.0	9.0	-	Y	-	Y	Y	Y	Y
P75521	Asparagine--tRNA ligase	*asnS*	455	51.9	8.8	-	-	-	Y	Y	-	-
P75522	Probable ribulose-phosphate 3-epimerase	*rpe*	215	24.8	6.6	Y	-	-	-	Y	-	-
P75525	Putative protein phosphatase	*MPN_247*	259	29.7	8.6	-	Y	-	Y	Y	-	-
P75526	Guanylate kinase	*gmk*	189	27.0	7.7	-	Y	Y	Y	Y	Y	Y
P75527	Peptide deformylase	*def*	193	24.6	6.4	-	-	-	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75528	Uncharacterized protein MG105 homolog	*MPN_244*	202	22.7	9.4	-	Y	-	Y	Y	-	-
P75529	Ribonuclease R	*rnr*	726	83.2	5.6	-	Y	Y	Y	Y	-	-
P75531	Thioredoxin reductase	*trxB*	315	34.5	6.7	-	-	-	-	Y	-	-
P75532	Uncharacterized HTH-type transcriptional regulator MG101 homolog	*MPN_239*	222	25.8	9.5	Y	Y	-	-	-	-	-
P75533	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	*gatB*	477	54.4	7.7	-	Y	-	Y	Y	-	-
P75534	Glutamyl-tRNA(Gln) amidotransferase subunit A	*gatA*	478	53.2	7.7	-	Y	-	Y	Y	Y	-
P75535	Uncharacterized protein MG098 homolog	*MPN_236*	479	54.1	9.6	-	-	-	-	Y	-	-
P75538	Uncharacterized lipoprotein MG095 homolog	*MPN_233*	454	49.8	9.1	-	Y	Y	Y	Y	-	Y
P75539	Replicative DNA helicase	*dnaB*	473	54.5	5.6	Y	Y	Y	Y	Y	Y	Y
P75540	50S ribosomal protein L9	*rplI*	149	17.1	9.5	-	Y	-	-	Y	-	-
P75541	30S ribosomal protein S18	*rpsR*	103	12.4	10.0	-	-	-	-	Y	-	-
P75542	Single-stranded DNA-binding protein	*ssb*	166	18.4	4.3	-	-	-	-	Y	-	-
P75543	30S ribosomal protein S6	*rpsF*	215	25.4	9.8	Y	Y	Y	Y	Y	Y	Y
P75544	Elongation factor G	*fusA*	688	76.5	5.5	Y	Y	Y	Y	Y	-	Y
P75545	30S ribosomal protein S7	*rpsG*	155	17.9	10.5	Y	Y	Y	-	Y	Y	-
P75546	30S ribosomal protein S12	*rpsL*	139	15.6	11.0	Y	Y	-	-	-	-	-
P75548	HPr kinase/phosphorylase	*hprK*	312	35.2	8.9	-	-	-	-	Y	-	-
P75550	50S ribosomal protein L11	*rplK*	137	14.8	9.7	Y	Y	-	Y	Y	Y	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75551	Oligopeptide transport ATP-binding protein OppF	*oppF*	851	98.6	9.2	-	Y	-	Y	-	-	Y
P75552	Oligopeptide transport ATP-binding protein OppD	*oppD*	423	47.4	9.4	-	Y	-	-	-	-	-
P75553	Oligopeptide transport system permease protein OppC	*oppC*	376	41.2	9.7	-	Y	Y	-	Y	Y	Y
P75556	Uncharacterized protein MG075 homolog (P116 adhesin)	*MPN_213*	1030	115.9	5.9	Y	Y	Y	Y	Y	Y	Y
P75558	UvrABC system protein B	*uvrB*	657	75.1	8.5	-	Y	-	-	Y	-	-
P75559	Protein translocase subunit SecA	*secA*	808	91.8	7.0	Y	Y	Y	Y	-	Y	Y
P75560	30S ribosomal protein S2	*rpsB*	294	33.4	9.2	-	Y	Y	Y	Y	Y	Y
P75563	Phenylalanine--tRNA ligase beta subunit	*pheT*	805	91.7	8.4	Y	Y	-	Y	-	-	-
P75564	Phenylalanine--tRNA ligase alpha subunit	*pheS*	341	39.2	6.6	-	-	-	Y	-	-	-
P75568	Uncharacterized protein MPN_101	*MPN_101*	425	46.9	6.6	-	-	-	-	Y	-	-
P75569	PTS system glucose-specific EIICBA component	*ptsG*	940	101.6	9.4	Y	Y	Y	Y	Y	Y	Y
P75571	Uncharacterized protein MPN_205	*MPN_205*	438	47.9	9.1	-	-	-	-	Y	-	-
P75572	UPF0134 protein MPN_204	*MPN_204*	148	17.3	9.3	-	-	-	-	Y	-	-
P75575	50S ribosomal protein L22	*rplV*	159	17.3	10.9	Y	Y	-	Y	Y	Y	-
P75576	30S ribosomal protein S19	*rpsS*	87	10.0	10.5	-	-	-	-	Y	-	-
P75577	50S ribosomal protein L2	*rplB*	287	31.9	10.4	-	Y	-	-	Y	-	-
P75578	50S ribosomal protein L23	*rplW*	237	25.8	10.1	-	Y	-	-	Y	-	-
P75579	50S ribosomal protein L4	*rplD*	212	23.6	10.1	-	Y	-	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75580	50S ribosomal protein L3	*rplC*	287	31.2	9.9	Y	Y	-	Y	Y	Y	Y
P75581	30S ribosomal protein S10	*rpsJ*	108	12.2	10.0	Y	Y	-	Y	Y	Y	-
P75583	Uncharacterized lipoprotein MG149 homolog	*MPN_162*	320	36.1	6.2	-	-	-	-	-	Y	Y
P75584	Uncharacterized protein MG148 homolog	*MPN_161*	445	52.8	5.9	-	Y	-	Y	Y	-	-
P75587	Putative riboflavin biosynthesis protein RibF	*ribF*	269	30.4	9.2	-	Y	-	-	Y	-	-
P75588	Uncharacterized protein MG144 homolog	*MPN_157*	402	44.3	10.0	-	-	-	-	Y	-	-
P75589	Ribosome-binding factor A	*rbfA*	116	13.4	9.4	-	Y	-	Y	Y	-	-
P75590	Translation initiation factor IF-2	*infB*	617	67.9	8.0	-	Y	-	Y	Y	-	-
P75591	Transcription termination/antitermination protein NusA	*nusA*	540	60.2	5.9	Y	Y	Y	Y	Y	Y	Y
P75592	UPF0134 protein MPN_100	*MPN_100*	183	21.0	9.5	-	-	-	-	Y	-	-
P75598	UPF0134 protein MPN_094	*MPN_094*	140	16.6	9.5	-	-	-	-	Y	-	-
P75611	Transketolase	*tkt*	648	72.3	7.1	Y	-	Y	-	Y	-	-
P75612	Putative ABC transporter ATP-binding protein MG065 homolog	*MPN_081*	465	53.5	9.3	-	Y	-	-	-	-	-
P75613	Uncharacterized ABC transporter permease MG064 homolog	*MPN_080*	1386	154.8	9.3	-	-	-	-	Y	-	-
P78003	S-adenosylmethionine synthase	*metK*	383	42.5	6.1	-	-	-	-	-	Y	-
P78004	Chaperone protein DnaJ	*dnaJ*	390	43.1	8.9	-	Y	-	Y	-	-	-
P78005	Uncharacterized deoxyribonuclease MG009 homolog	*MPN_009*	261	29.8	8.2	-	-	-	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P78007	L-lactate dehydrogenase	*ldh*	312	33.9	8.4	Y	Y	Y	Y	Y	Y	Y
P78009	Elongation factor Ts	*tsf*	298	33.6	7.7	Y	Y	Y	Y	Y	-	Y
P78010	Triosephosphate isomerase	*tpiA*	244	26.9	7.0	-	-	Y	-	Y	-	-
P78011	Serine hydroxymethyltransferase	*glyA*	406	45.2	9.1	-	Y	-	-	Y	Y	-
P78012	60 kDa chaperonin	*groL*	543	58.1	5.6	Y	Y	Y	Y	Y	-	Y
P78013	DNA-directed RNA polymerase subunit beta	*rpoB*	1391	155.5	6.1	Y	Y	Y	Y	Y	-	Y
P78017	Protein GrpE	*grpE*	217	24.7	7.8	Y	Y	Y	Y	Y	Y	Y
P78018	Phosphoglycerate kinase	*pgk*	409	44.2	7.1	Y	Y	Y	Y	Y	Y	Y
P78019	Transcription elongation factor greA	*greA*	160	18.1	4.9	Y	Y	Y	Y	Y	-	-
P78021	DNA ligase	*ligA*	658	73.9	8.4	-	Y	-	Y	Y	-	-
P78022	RNA polymerase sigma factor SigA	*sigA*	499	57.8	8.9	-	Y	-	Y	Y	Y	Y
P78023	50S ribosomal protein L20	*rplT*	127	14.7	11.3	Y	Y	-	-	Y	-	-
P78024	Translation initiation factor IF-3	*infC*	201	23.1	9.9	-	Y	-	-	Y	-	-
P78025	Lon protease	*lon*	795	90.1	7.6	Y	Y	-	Y	Y	Y	Y
P78026	50S ribosomal protein L21	*rplU*	100	11.6	9.5	Y	Y	-	-	-	-	-
P78027	Ribonucleoside-diphosphate reductase subunit alpha	*nrdE*	721	82.3	6.6	Y	Y	Y	Y	Y	Y	Y
P78028	Dihydrofolate reductase	*folA*	160	18.5	8.4	-	-	-	-	-	Y	-
P78029	Thymidylate synthase	*thyA*	287	33.6	8.6	-	-	-	Y	Y	-	Y
P78030	Carbamate kinase-like protein	*MPN_307*	309	32.8	8.7	-	-	-	-	Y	-	-
P78031	Pyruvate kinase	*pyk*	508	57.2	9.2	Y	Y	Y	Y	Y	Y	Y

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
P78032	DNA topoisomerase 1	*topA*	711	81.9	9.3	-	Y	-	-	Y	-	Y
P78033	Glucose-6-phosphate isomerase	*pgi*	430	48.8	9.1	Y	-	-	-	Y	-	-
P78034	Peptidyl-tRNA hydrolase	*pth*	188	21.4	9.5	Y	-	-	-	Y	-	-
P78035	50S ribosomal protein L1	*rplA*	226	24.3	9.6	Y	Y	Y	Y	Y	-	-
P78036	Probable cation-transporting P-type ATPase	*pacL*	872	94.9	6.4	-	Y	-	-	-	-	-
Q50288	Uncharacterized lipoprotein MPN_200	*MPN_200*	798	87.6	9.4	-	-	-	Y	-	-	-
Q50291	tRNA pseudouridine synthase A	*truA*	243	27.7	9.8	-	Y	-	-	-	-	-
Q50292	Uncharacterized protein MG181 homolog	*MPN_195*	434	49.1	9.9	-	-	-	-	-	Y	-
Q50293	Energy-coupling factor transporter ATP-binding protein EcfA2	*ecfA2*	303	34.3	8.2	Y	-	-	-	-	-	-
Q50294	Energy-coupling factor transporter ATP-binding protein EcfA1	*ecfA1*	274	30.5	9.1	Y	Y	Y	Y	Y	-	-
Q50295	DNA-directed RNA polymerase subunit alpha	*rpoA*	327	36.6	7.0	Y	Y	Y	Y	Y	Y	-
Q50296	30S ribosomal protein S11	*rpsK*	121	12.7	10.3	Y	Y	-	-	Y	-	-
Q50298	Translation initiation factor IF-1	*infA*	78	9.0	9.9	-	-	-	-	Y	-	-
Q50299	Adenylate kinase	*adk*	213	24.2	7.1	-	Y	Y	-	Y	-	-
Q50300	50S ribosomal protein L15	*rplO*	151	16.7	10.9	-	Y	-	-	-	-	-
Q50301	30S ribosomal protein S5	*rpsE*	219	24.1	11.0	Y	Y	-	-	-	-	-
Q50302	50S ribosomal protein L18	*rplR*	116	13.0	10.9	Y	Y	-	-	-	-	-
Q50303	50S ribosomal protein L6	*rplF*	184	20.6	9.9	-	Y	-	Y	Y	Y	Y
Q50304	30S ribosomal protein S8	*rpsH*	142	15.9	10.3	-	Y	-	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
Q50306	50S ribosomal protein L5	*rplE*	180	20.2	9.9	Y	Y	-	Y	Y	-	-
Q50308	50S ribosomal protein L14	*rplN*	122	13.4	10.4	Y	-	-	-	-	-	-
Q50310	50S ribosomal protein L29	*rpmC*	111	13.0	10.0	Y	Y	-	-	Y	-	-
Q50313	DNA polymerase III subunit beta	*dnaN*	380	43.8	5.8	Y	Y	Y	-	Y	Y	-
Q50314	ParA family protein	*MPN_688*	270	30.1	7.0	Y	Y	Y	Y	-	Y	Y
Q50315	Uncharacterized protein MPN_687	*MPN_687*	250	28.6	5.7	-	-	Y	Y	Y	-	-
Q50327	ATP synthase subunit b	*atpF*	207	24.0	5.6	Y	Y	Y	Y	Y	Y	-
Q50328	ATP synthase subunit delta	*atpH*	178	20.7	9.4	-	-	-	-	Y	-	-
Q50329	ATP synthase subunit alpha	*atpA*	518	57.3	6.2	Y	Y	Y	Y	Y	Y	Y
Q50330	ATP synthase gamma chain	*atpG*	279	32.4	9.0	-	-	-	-	-	Y	Y
Q50331	ATP synthase subunit beta	*atpD*	475	52.2	5.5	Y	Y	Y	Y	Y	Y	Y
Q50332	ATP synthase epsilon chain	*atpC*	133	15.2	9.7	-	Y	-	Y	-	Y	-
Q50333	Uncharacterized protein MG397 homolog	*MPN_596*	569	66.3	8.8	-	-	-	-	Y	-	-
Q50341	Mgp-operon protein 3	*MPN_142*	1218	130.4	8.0	Y	Y	Y	Y	Y	Y	Y
Q50360	Cytadherence high molecular weight protein 3	*hmw3*	672	73.7	4.7	Y	Y	Y	Y	Y	Y	Y
Q50365	Cytadherence high molecular weight protein 1	*hmw1*	1018	112.1	4.0	Y	Y	Y	Y	Y	Y	Y
Q59547	50S ribosomal protein L17	*rplQ*	124	14.2	10.2	-	Y	-	-	-	-	-
Q59548	Preprotein translocase subunit secY	*secY*	477	51.9	9.7	-	Y	-	-	-	-	-
Q59549	Chromosomal replication initiator protein DnaA	*dnaA*	439	50.6	8.9	-	Y	-	Y	Y	Y	-

UniProt #	Name	Gene	Length	Mass (kDa)	pI	A549	Act	Fet	Fn	Hep	Plg	Cleav
Q9EXC9	Protein MG115 homolog	*MPN_254*	157	16.9	8.8	Y	-	-	Y	Y	-	-
Q9EXD2	Uncharacterized protein MPN_272	*MPN_272*	93	10.6	9.0	-	-	-	-	Y	-	-
Q9EXD3	Uncharacterized protein MPN_377	*MPN_377*	74	9.1	6.4	-	-	-	-	Y	-	-
Q9EXD4	Uncharacterized protein MG269.1 homolog	*MPN_388*	128	15.2	4.9	Y	-	-	Y	Y	Y	-

Affinity chromatography columns were shortened to: A549 for human A549 epithelial cells, Act for actin, Fet for fetuin, Fn for fibronectin, Hep for heparin, and Plg for plasminogen. Cleav stands for whether or not the same protein was identified at a lower mass on SDS-PAGE from either of the affinity chromatography experiments. Y under any of these columns in the table indicates whether or not the protein was identified by the respective methodologies. Proteins listed had a minimum MASCOT protein score of 43.

8.3 Appendix 3: Potential *M. pneumoniae* adhesins

Appendix 3: Table 1: *M. pneumoniae* surface proteins also identified in 'Bait and Prey' affinity chromatography.

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
P0CJ81	Proline-rich P65 protein	*p65*	405	47.0	Non	0	-	Y	Y	Y	Y	Y	-
P11311	Adhesin P1	*mgpA*	1627	176.2	Class.	>2	Y	Y	Y	Y	Y	Y	Y
P22447	DNA gyrase subunit B	*gyrB*	650	73.8	-	1	Y	Y	Y	Y	Y	Y	Y
P23568	Elongation factor Tu	*tuf*	394	43.1	-	0	Y	Y	Y	Y	Y	Y	Y
P46775	30S ribosomal protein S4	*rpsD*	205	23.8	-	0	Y	Y	-	Y	Y	-	-
P53527	Probable ribose-5-phosphate isomerase B	*rpiB*	150	16.7	-	1	Y	-	-	Y	Y	-	-
P54125	Oligoendopeptidase F homolog	*pepF*	611	70.9	Non	0	Y	Y	-	Y	Y	Y	Y
P75033	Uncharacterized ATP-dependent helicase MG140 homolog	*MPN_153*	1113	130.3	-	0	-	Y	-	Y	Y	Y	-
P75044	Ribose-phosphate pyrophosphokinase	*prs*	328	36.7	-	0	Y	Y	-	Y	Y	-	-
P75049	Transcription termination/antitermination protein NusG	*nusG*	320	36.1	-	0	-	Y	Y	Y	Y	Y	Y
P75053	Purine nucleoside phosphorylase DeoD-type	*deoD*	238	26.3	-	1	Y	Y	Y	Y	Y	Y	-
P75054	Signal recognition particle protein	*ffh*	450	50.1	-	0	-	Y	-	Y	Y	-	-
P75061	Phosphocarrier protein HPr	*ptsH*	88	9.5	-	0	Y	-	-	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75062	Uncharacterized lipoprotein MG040 homolog	*MPN_052*	657	71.6	Class.	>2	Y	Y	Y	Y	Y	Y	Y
P75063	Uncharacterized protein MG039 homolog	*MPN_051*	384	42.7	-	1	-	-	Y	-	Y	-	-
P75064	Glycerol kinase	*glpK*	508	56.6	-	1	Y	Y	Y	Y	Y	Y	Y
P75070	Thymidine kinase	*tdk*	224	21.5	-	0	-	-	-	-	Y	-	-
P75089	Fructose-bisphosphate aldolase	*fba*	288	31.0	-	0	Y	Y	Y	Y	Y	Y	-
P75090	Probable DNA-directed RNA polymerase subunit delta	*rpoE*	144	17.2	-	0	-	Y	-	Y	Y	Y	-
P75092	Putative proline iminopeptidase	*pip*	309	34.7	-	1	Y	-	-	Y	Y	-	-
P75109	Uncharacterized ABC transporter permease MG468 homolog	*MPN_684*	1882	209.3	Non	>2	-	Y	-	-	Y	-	Y
P75114	Glutamate--tRNA ligase	*gltX*	484	55.6	-	0	-	-	-	Y	Y	-	Y
P75120	ATP-dependent zinc metalloprotease	*ftsH*	709	77.7	-	2	Y	Y	Y	Y	Y	Y	Y
P75129	Peptide methionine sulfoxide reductase	*msrB*	151	17.3	Non	0	-	-	-	-	Y	-	-
P75131	30S ribosomal protein S16	*rpsP*	88	10.0	Non	0	Y	-	-	-	Y	-	-
P75159	Putative type I restriction enzyme specificity protein	*MPN_638*	375	42.6	-	1	-	Y	-	Y	Y	-	Y
P75161	Ribosome-recycling factor	*frr*	185	21.7	-	0	Y	Y	Y	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75167	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	*gpmI*	508	56.3	-	0	Y	-	-	-	Y	-	-
P75168	Phosphoenolpyruvate-protein phosphotransferase	*ptsI*	572	63.9	-	0	-	Y	Y	Y	Y	-	-
P75170	Hydroperoxide reductase	*MPN_625*	141	15.5	-	1	Y	-	-	Y	Y	Y	-
P75179	30S ribosomal protein S9	*rpsI*	132	15.1	Non	0	Y	Y	-	-	-	-	-
P75184	Uncharacterized lipoprotein MG412 homolog	*MPN_611*	385	41.0	Class.	1	-	Y	Y	Y	-	Y	-
P75189	Enolase	*eno*	456	49.2	-	1	Y	Y	Y	Y	Y	Y	Y
P75205	10 kDa chaperonin	*groS*	116	12.6	Non	0	Y	Y	Y	Y	Y	Y	-
P75206	Probable cytosol aminopeptidase	*pepA*	445	48.8	-	1	Y	Y	Y	Y	Y	Y	-
P75223	Uncharacterized protein MG377 homolog	*MPN_555*	193	22.4	-	0	Y	Y	Y	Y	Y	Y	Y
P75231	Uncharacterized protein MG369 homolog	*MPN_547*	558	62.4	-	1	-	Y	-	Y	Y	-	Y
P75237	30S ribosomal protein S20	*rpsT*	87	10.0	Non	0	-	-	-	-	Y	-	-
P75239	50S ribosomal protein L7/L12	*rplL*	122	13.1	-	1	Y	Y	Y	Y	Y	Y	-
P75240	50S ribosomal protein L10	*rplJ*	161	17.6	-	1	Y	-	-	Y	Y	-	-
P75245	Acetate kinase	*ackA*	390	43.7	-	0	Y	-	Y	-	Y	-	Y
P75247	Chaperone protein ClpB	*clpB*	715	81.3	-	1	Y	Y	Y	Y	Y	Y	Y
P75248	Uncharacterized protein MG354 homolog	*MPN_530*	136	15.6	-	1	Y	Y	-	Y	Y	Y	-

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75249	Uncharacterized protein MG353 homolog	*MPN_529*	109	12.4	Non	0	-	Y	-	-	-	-	-
P75250	Inorganic pyrophosphatase	*ppa*	184	21.4	-	1	-	-	Y	Y	-	Y	-
P75254	UPF0134 protein MPN_524	*MPN_524*	168	20.1	-	0	-	-	-	-	Y	-	Y
P75255	Uncharacterized lipoprotein MG348 homolog	*MPN_523*	305	33.5	Class.	1	-	Y	Y	Y	Y	Y	-
P75259	UPF0134 protein MPN_139	*MPN_139*	163	19.1	Non	0	-	-	-	-	Y	-	-
P75269	Uncharacterized protein MG343 homolog	*MPN_518*	348	40.7	-	0	-	Y	-	-	-	Y	-
P75271	DNA-directed RNA polymerase subunit beta	*rpoC*	1290	144.8	-	0	Y	Y	Y	Y	Y	Y	Y
P75295	Uncharacterized protein MPN_491	*MPN_491*	474	52.9	Non	1	-	Y	-	-	Y	-	Y
P75296	Uncharacterized lipoprotein MG338 homolog	*MPN_489*	1300	143.0	Non	2	-	-	Y	-	-	-	Y
P75305	FMN-dependent NADH-azoreductase	*azoR*	197	21.5	Non	>2	Y	Y	Y	Y	Y	Y	-
P75310	Uncharacterized protein MG328 homolog	*MPN_474*	1140	118.0	Non	0	Y	Y	Y	Y	Y	Y	Y
P75313	Putative Xaa-Pro aminopeptidase	*pepP*	354	39.6	-	0	Y	Y	-	Y	Y	Y	-
P75327	Uncharacterized lipoprotein MG321 homolog	*MPN_456*	1005	110.4	Class.	>2	Y	Y	Y	Y	Y	Y	Y
P75329	Uncharacterized protein MG319 homolog	*MPN_454*	193	21.4	Non	2	Y	Y	Y	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75334	Uncharacterized lipoprotein MG309 homolog	*MPN_444*	1325	146.2	Class.	1	Y	Y	Y	Y	Y	-	Y
P75342	Uncharacterized lipoprotein MG307 homolog	*MPN_436*	1244	139.0	Class.	2	-	Y	Y	Y	Y	-	Y
P75344	Chaperone protein DnaK	*dnaK*	595	65.1	-	0	Y	Y	Y	Y	Y	Y	Y
P75353	Uncharacterized protein MG202 homolog	*MPN_121*	121	13.3	Non	0	Y	-	-	-	-	-	-
P75354	DnaJ-like protein MG200 homolog (TopJ accessory protein)	*MPN_119*	910	100.1	Non	0	-	Y	Y	-	Y	-	Y
P75358	Glyceraldehyde-3-phosphate dehydrogenase	*gapA*	337	36.8	-	0	Y	Y	Y	Y	Y	Y	Y
P75359	Phosphate acetyltransferase	*pta*	320	35.2	-	1	Y	Y	Y	Y	Y	-	Y
P75360	Uncharacterized protein MPN_427	*MPN_427*	290	33.2	Non	0	Y	Y	Y	Y	-	-	-
P75376	Uncharacterized lipoprotein MPN_408	*MPN_408*	760	83.3	Class.	>2	-	Y	-	-	-	-	Y
P75383	Uncharacterized protein MG281 homolog	*MPN_400*	582	65.9	Non	1	-	Y	Y	Y	-	Y	Y
P75385	Uncharacterized protein MG279 homolog	*MPN_398*	218	24.9	Class.	2	-	Y	-	-	Y	-	-
P75389	Probable NADH oxidase	*nox*	479	52.8	-	>2	Y	Y	Y	Y	Y	Y	Y
P75390	Pyruvate dehydrogenase E1 component subunit alpha	*pdhA*	358	40.6	-	1	Y	Y	Y	Y	Y	Y	Y
P75391	Pyruvate dehydrogenase E1 component subunit beta	*pdhB*	327	35.9	-	2	Y	Y	Y	Y	Y	Y	Y

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75392	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	*pdhC*	402	42.4	-	2	Y	-	Y	Y	Y	Y	Y
P75393	Dihydrolipoyl dehydrogenase	*pdhD*	457	49.4	-	>2	Y	Y	Y	Y	Y	Y	Y
P75394	Probable lipoate-protein ligase A	*lplA*	339	39.2	-	0	Y	Y	Y	Y	Y	Y	-
P75395	Uncharacterized protein MG269 homolog	*MPN_387*	358	42.6	Non	0	Y	Y	-	Y	Y	Y	Y
P75396	Uncharacterized protein MG268 homolog	*MPN_386*	229	27.0	-	1	Y	Y	-	Y	Y	-	-
P75399	Putative phosphatase MPN_383	*MPN_383*	282	31.9	-	0	Y	Y	-	Y	Y	-	-
P75405	Uncharacterized protein MPN_376	*MPN_376*	1140	130.3	Class.	2	-	Y	-	Y	Y	-	Y
P75409	ADP-ribosylating toxin CARDS	*MPN_372*	591	68.0	Non	0	Y	Y	Y	Y	-	Y	Y
P75454	Trigger factor	*tig*	444	51.3	-	0	-	Y	Y	Y	Y	Y	Y
P75460	Protein NrdI	*nrdI*	153	17.1	-	0	-	Y	-	Y	Y	-	-
P75461	Ribonucleoside-diphosphate reductase subunit beta	*nrdF*	339	39.4	-	1	-	Y	-	Y	Y	Y	-
P75466	Ribosomal RNA small subunit methyltransferase H	*rsmH*	308	35.2	-	0	Y	Y	Y	Y	-	Y	-
P75467	Protein MraZ	*mraZ*	141	16.3	-	0	Y	-	Y	-	Y	-	-
P75470	Uncharacterized protein MG218.1 homolog (P41 accessory protein)	*MPN_311*	357	40.6	Non	0	Y	Y	Y	Y	-	Y	-

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75471	Cytadherence high molecular weight protein 2	*hmw2*	1818	215.5	-	0	-	Y	Y	-	Y	Y	Y
P75477	Segregation and condensation protein B	*scpB*	208	23.7	-	1	-	Y	Y	Y	Y	-	-
P75482	Uncharacterized protein MG210.1 homolog	*MPN_295*	220	25.7	Non	0	Y	Y	-	Y	Y	Y	Y
P75489	Uncharacterized lipoprotein MPN_288	*MPN_288*	787	86.8	Class.	2	-	Y	Y	Y	Y	-	Y
P75493	Uncharacterized lipoprotein MPN_284	*MPN_284*	794	87.1	Class.	2	Y	Y	Y	Y	Y	Y	Y
P75497	Ribonuclease J	*MPN_280*	569	64.0	-	0	Y	Y	-	Y	Y	Y	Y
P75504	Uncharacterized 16.1 kDa HIT-like protein	*MPN_273*	144	16.1	-	0	Y	Y	-	Y	Y	-	-
P75509	Uncharacterized protein MG127 homolog	*MPN_266*	145	16.8	Non	0	Y	Y	-	-	Y	-	-
P75512	Thioredoxin	*trxA*	102	11.2	-	0	Y	Y	Y	Y	Y	-	-
P75518	Uncharacterized protein MG117 homolog	*MPN_256*	223	25.7	Non	0	-	-	-	Y	Y	-	-
P75526	Guanylate kinase	*gmk*	189	27.0	-	1	-	Y	Y	Y	Y	Y	Y
P75529	Ribonuclease R	*rnr*	726	83.2	-	>2	-	Y	Y	Y	Y	-	-
P75538	Uncharacterized lipoprotein MG095 homolog	*MPN_233*	454	49.8	Class.	1	-	Y	Y	Y	Y	-	Y
P75539	Replicative DNA helicase	*dnaB*	473	54.5	-	0	Y	Y	Y	Y	Y	Y	Y
P75542	Single-stranded DNA-binding protein	*ssb*	166	18.4	Non	0	-	-	-	-	Y	-	-
P75544	Elongation factor G	*fusA*	688	76.5	-	1	Y	Y	Y	Y	Y	-	Y

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75545	30S ribosomal protein S7	*rpsG*	155	17.9	-	0	Y	Y	Y	-	Y	Y	-
P75550	50S ribosomal protein L11	*rplK*	137	14.8	-	0	Y	Y	-	Y	Y	Y	-
P75551	Oligopeptide transport ATP-binding protein OppF	*oppF*	851	98.6	-	0	-	Y	-	Y	-	-	Y
P75556	Uncharacterized protein MG075 homolog (P116 adhesin)	*MPN_213*	1030	115.9	Class.	2	Y	Y	Y	Y	Y	Y	Y
P75558	UvrABC system protein B	*uvrB*	657	75.1	-	0	-	Y	-	-	Y	-	-
P75559	Protein translocase subunit SecA	*secA*	808	91.8	-	1	Y	Y	Y	Y	-	Y	Y
P75560	30S ribosomal protein S2	*rpsB*	294	33.4	-	0	-	Y	Y	Y	Y	Y	Y
P75563	Phenylalanine--tRNA ligase beta subunit	*pheT*	805	91.7	-	0	Y	Y	-	Y	-	-	-
P75569	PTS system glucose-specific EIICBA component	*ptsG*	940	101.6	Non	>2	Y	Y	Y	Y	Y	Y	Y
P75576	30S ribosomal protein S19	*rpsS*	87	10.0	Non	0	-	-	-	-	Y	-	-
P75577	50S ribosomal protein L2	*rplB*	287	31.9	-	0	-	Y	-	-	Y	-	-
P75578	50S ribosomal protein L23	*rplW*	237	25.8	Non	0	-	Y	-	-	Y	-	-
P75579	50S ribosomal protein L4	*rplD*	212	23.6	Non	0	-	Y	-	-	Y	-	-
P75580	50S ribosomal protein L3	*rplC*	287	31.2	Non	1	Y	Y	-	Y	Y	Y	Y
P75581	30S ribosomal protein S10	*rpsJ*	108	12.2	-	0	Y	Y	-	Y	Y	Y	-
P75583	Uncharacterized lipoprotein MG149 homolog	*MPN_162*	320	36.1	Class.	1	-	-	-	-	-	Y	Y
P75588	Uncharacterized protein MG144 homolog	*MPN_157*	402	44.3	Non	>2	-	-	-	-	Y	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
P75590	Translation initiation factor IF-2	*infB*	617	67.9	-	1	-	Y	-	Y	Y	-	-
P75591	Transcription termination/antitermination protein NusA	*nusA*	540	60.2	Non	0	Y	Y	Y	Y	Y	Y	Y
P78007	L-lactate dehydrogenase	*ldh*	312	33.9	-	1	Y	Y	Y	Y	Y	Y	Y
P78009	Elongation factor Ts	*tsf*	298	33.6	-	0	Y	Y	Y	Y	Y	-	Y
P78010	Triosephosphate isomerase	*tpiA*	244	26.9	-	1	-	-	Y	-	Y	-	-
P78012	60 kDa chaperonin	*groL*	543	58.1	-	2	Y	Y	Y	Y	Y	-	Y
P78013	DNA-directed RNA polymerase subunit beta	*rpoB*	1391	155.5	-	1	Y	Y	Y	Y	Y	-	Y
P78017	Protein GrpE	*grpE*	217	24.7	-	0	Y	Y	Y	Y	Y	Y	Y
P78018	Phosphoglycerate kinase	*pgk*	409	44.2	-	2	Y	Y	Y	Y	Y	Y	Y
P78019	Transcription elongation factor greA	*greA*	160	18.1	-	0	Y	Y	Y	Y	Y	-	-
P78022	RNA polymerase sigma factor SigA	*sigA*	499	57.8	-	0	-	Y	-	Y	Y	Y	Y
P78025	Lon protease	*lon*	795	90.1	-	0	Y	Y	-	Y	Y	Y	Y
P78027	Ribonucleoside-diphosphate reductase subunit alpha	*nrdE*	721	82.3	-	1	Y	Y	Y	Y	Y	Y	Y
P78031	Pyruvate kinase	*pyk*	508	57.2	-	1	Y	Y	Y	Y	Y	Y	Y
P78032	DNA topoisomerase 1	*topA*	711	81.9	Non	0	-	Y	-	-	Y	-	Y
P78035	50S ribosomal protein L1	*rplA*	226	24.3	-	0	Y	Y	Y	Y	Y	-	-
Q50288	Uncharacterized lipoprotein MPN_200	*MPN_200*	798	87.6	Class.	>2	-	-	-	Y	-	-	-

UniProt #	Name	Gene	Length	Mass (kDa)	Secretion	Tm	A549	Act	Fet	Fn	Hep	Plg	Cleav
Q50293	Energy-coupling factor transporter ATP-binding protein EcfA2	*ecfA2*	303	34.3	-	0	Y	-	-	-	-	-	-
Q50295	DNA-directed RNA polymerase subunit alpha	*rpoA*	327	36.6	-	2	Y	Y	Y	Y	Y	Y	-
Q50296	30S ribosomal protein S11	*rpsK*	121	12.7	Class.	0	Y	Y	-	-	Y	-	-
Q50298	Translation initiation factor IF-1	*infA*	78	9.0	-	0	-	-	-	-	Y	-	-
Q50301	30S ribosomal protein S5	*rpsE*	219	24.1	-	1	Y	Y	-	-	-	-	-
Q50303	50S ribosomal protein L6	*rplF*	184	20.6	-	0	-	Y	-	Y	Y	Y	Y
Q50313	DNA polymerase III subunit beta	*dnaN*	380	43.8	-	0	Y	Y	Y	-	Y	Y	-
Q50314	ParA family protein	*MPN_688*	270	30.1	-	1	Y	Y	Y	Y	-	Y	Y
Q50329	ATP synthase subunit alpha	*atpA*	518	57.3	-	1	Y	Y	Y	Y	Y	Y	Y
Q50331	ATP synthase subunit beta	*atpD*	475	52.2	-	0	Y	Y	Y	Y	Y	Y	Y
Q50341	Mgp-operon protein 3	*MPN_142*	1218	130.4	Class.	>2	Y	Y	Y	Y	Y	Y	Y
Q50360	Cytadherence high molecular weight protein 3	*hmw3*	672	73.7	Non	0	Y	Y	Y	Y	Y	Y	Y
Q50365	Cytadherence high molecular weight protein 1	*hmw1*	1018	112.1	Non	0	Y	Y	Y	Y	Y	Y	Y
Q9EXD3	Uncharacterized protein MPN_377	*MPN_377*	74	9.1	Non	0	-	-	-	-	Y	-	-

Results in the Secretion column are from a combination of SignalP 4.1 and SecretomeP predictions of classically (Class.), non-classically (Non), or not (-) secreted proteins. Results in the Tm column are from TMpred predictions where the number listed indicates the number of predicted transmembrane domains. Affinity chromatography columns were shortened to: A549 for human A549 epithelial cells, Act for actin, Fet for fetuin,

Fn for fibronectin, Hep for heparin, and Plg for plasminogen. Cleav stands for whether or not the same protein was identified at a lower mass on SDS-PAGE from either of the affinity chromatography experiments. Y under any of these columns in the table indicates whether or not the protein was identified by the respective methodologies. Proteins listed had a minimum MASCOT protein score of 43.

8.4 Appendix 4: Excised gel sections from 1D-SDS PAGE experiments

Appendix 4: Table 1: Avidin affinity chromatography of gel sections and the mass range for each section.

Gel section	Mass range (kDa)	Gel section	Mass range (kDa)
1	Greater than 250	**7**	25 – 37
2	150 – 250	**8**	20 – 25
3	100 – 150	**9**	15 – 20
4	75 – 100	**10**	10 – 15
5	50 – 75	**11**	Less than 10
6	37 – 50		

Appendix 4: Table 2: Actin 'Bait and Prey' affinity chromatography of gel sections and the mass range for each section.

Gel section	Mass range (kDa)	Gel section	Mass range (kDa)
1	Greater than 300	**7**	38 – 50
2	150 – 300	**8**	35 – 38
3	100 – 150	**9**	27 – 35
4	75 – 100	**10**	20 – 27
5	62 – 75	**11**	15 – 20
6	50 – 62	**12**	Less than 15

Appendix 4: Table 3: Fetuin 'Bait and Prey' affinity chromatography of gel sections and the mass range for each section.

Gel section	Mass range (kDa)	Gel section	Mass range (kDa)
1	Greater than 300	**6**	35 – 45
2	150 – 300	**7**	20 – 35
3	100 – 150	**8**	15 – 20
4	62 – 100	**9**	Less than 15
5	45 – 62		

Appendix 4: Table 4: Fibronectin 'Bait and Prey' affinity chromatography of gel sections and the mass range for each section.

Gel section	Mass range (kDa)	Gel section	Mass range (kDa)
1	25 – 40	**4**	Greater than 350
2	15 – 25	**5**	250 – 350
3	Less than 15	**6**	200 – 250
		7	150 – 200
		8	125 – 150
		9	100 – 125
		10	90 – 100
		11	68 – 90
		12	62 – 68
		13	50 – 62
		14	45 – 50
		15	40 – 45
		16	32 – 40
		17	20 – 32
		18	15 – 20
		19	Less than 15

Sections are from two lanes separated by the different colours. Section 1 – 3 are from the first elution and section 4 – 19 are from the pooled samples of elution two to four.

Appendix 4: Table 5: Plasminogen 'Bait and Prey' affinity chromatography of gel sections and the mass range for each section.

Gel section	Mass range (kDa)	Gel section	Mass range (kDa)
1	Greater than 250	**7**	43 – 58
2	150 – 250	**8**	35 – 43
3	112 – 150	**9**	22 – 35
4	84 – 112	**10**	15 – 22
5	68 – 84	**11**	Less than 15
6	58 – 68		

Appendix 4: Table 6: Heparin 'Bait and Prey' affinity chromatography of gel sections and the mass range for each section.

Gel section	Mass range (kDa)	Gel section	Mass range (kDa)	Gel section	Mass range (kDa)
1	125 – 175	**27**	20 – 26	**52**	65 – 90
2	90 – 125	**28**	17 – 20	**53**	45 – 65
3	80 – 90	**29**	12 – 17	**54**	37 – 45
4	62 – 80	**30**	Less than 12	**55**	16 – 37
5	50 – 62	**31**	95 – 125	**56**	Less than 16
6	42 – 50	**32**	80 – 95	**57**	37 – 45
7	37 – 42	**33**	70 – 80	**58**	26 – 37
8	34 -37	**34**	62 – 75	**59**	20 – 26
9	31 – 34	**35**	50 – 62	**60**	Less than 20
10	28 – 31	**36**	43 – 50		
11	25 – 28	**37**	33 – 43		
12	23 – 25	**38**	28 – 33		
13	20 – 23	**39**	25 – 28		
14	17 – 20	**40**	20 – 25		
15	14 – 17	**41**	15 – 20		
16	12 – 14	**42**	13 – 15		
17	Less than 12	**43**	Less than 13		
18	150 – 300	**44**	65 – 90		
19	80 – 150	**45**	42 – 65		
20	74 – 80	**46**	37 – 42		
21	62 – 74	**47**	31 – 37		
22	53 – 62	**48**	20 – 31		
23	48 – 53	**49**	16 – 20		
24	38 – 48	**50**	13 – 16		
25	31 – 38	**51**	Less than 13		
26	26 – 31				

Sections are from six lanes separated by the different colours. Lanes are separated based on increasing concentrations of salt.

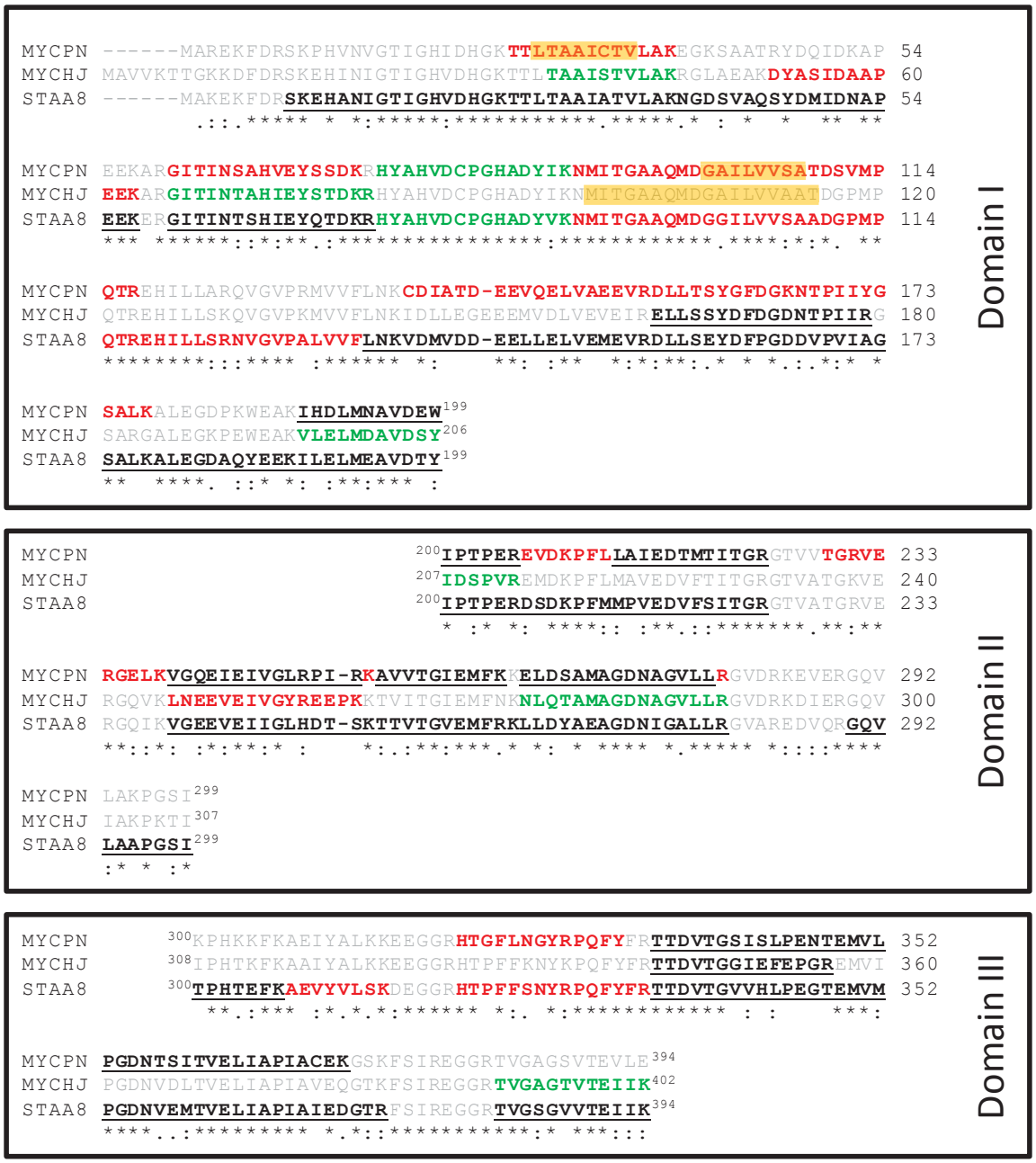
Appendix 4: Table 7: A549 surface protein complexes 'Bait and Prey' affinity chromatography of gel sections and the mass range for each section.

Gel section	Mass range (kDa)	Gel section	Mass range (kDa)
1	Greater than 300	**12**	Greater than 300
2	150 – 300	**13**	150 – 300
3	100 – 150	**14**	100 – 150
4	75 – 100	**15**	75 – 100
5	50 – 75	**16**	50 – 75
6	37 – 50	**17**	37 – 50
7	25 – 37	**18**	25 – 37
8	20 – 25	**19**	20 – 25
9	15 – 20	**20**	15 – 20
10	10 – 15	**21**	10 – 15
11	Less than 10	**22**	Less than 10

Sections are from the two different elution fractions separated by the different colours. Section 1 – 11 are from the 'Urea' fraction and section 12 – 22 are from the 'Acid' fraction.

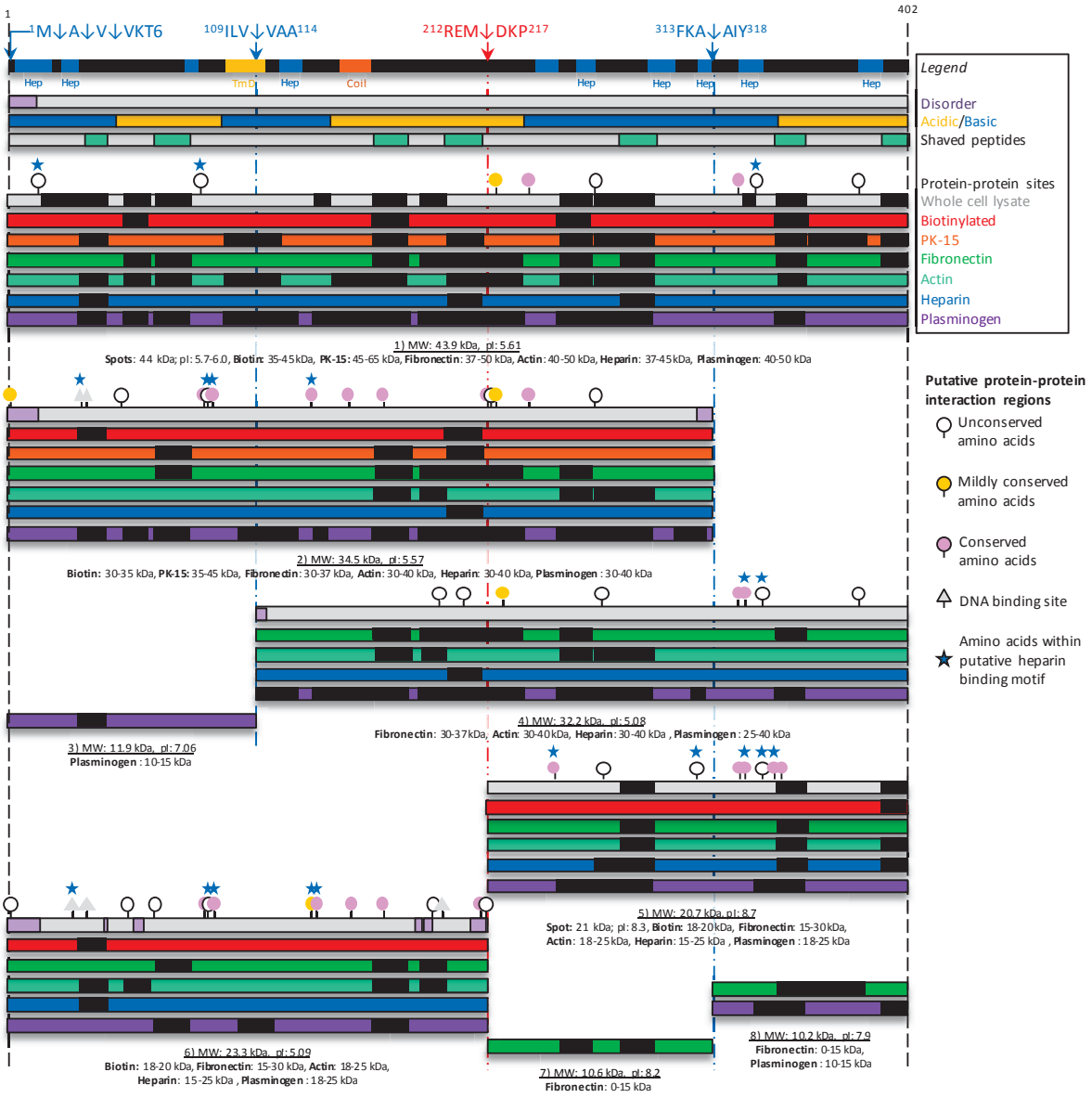
8.5 Appendix 5: Chapter 5 Supplementary Materials

8.5.1 Supplementary Figures



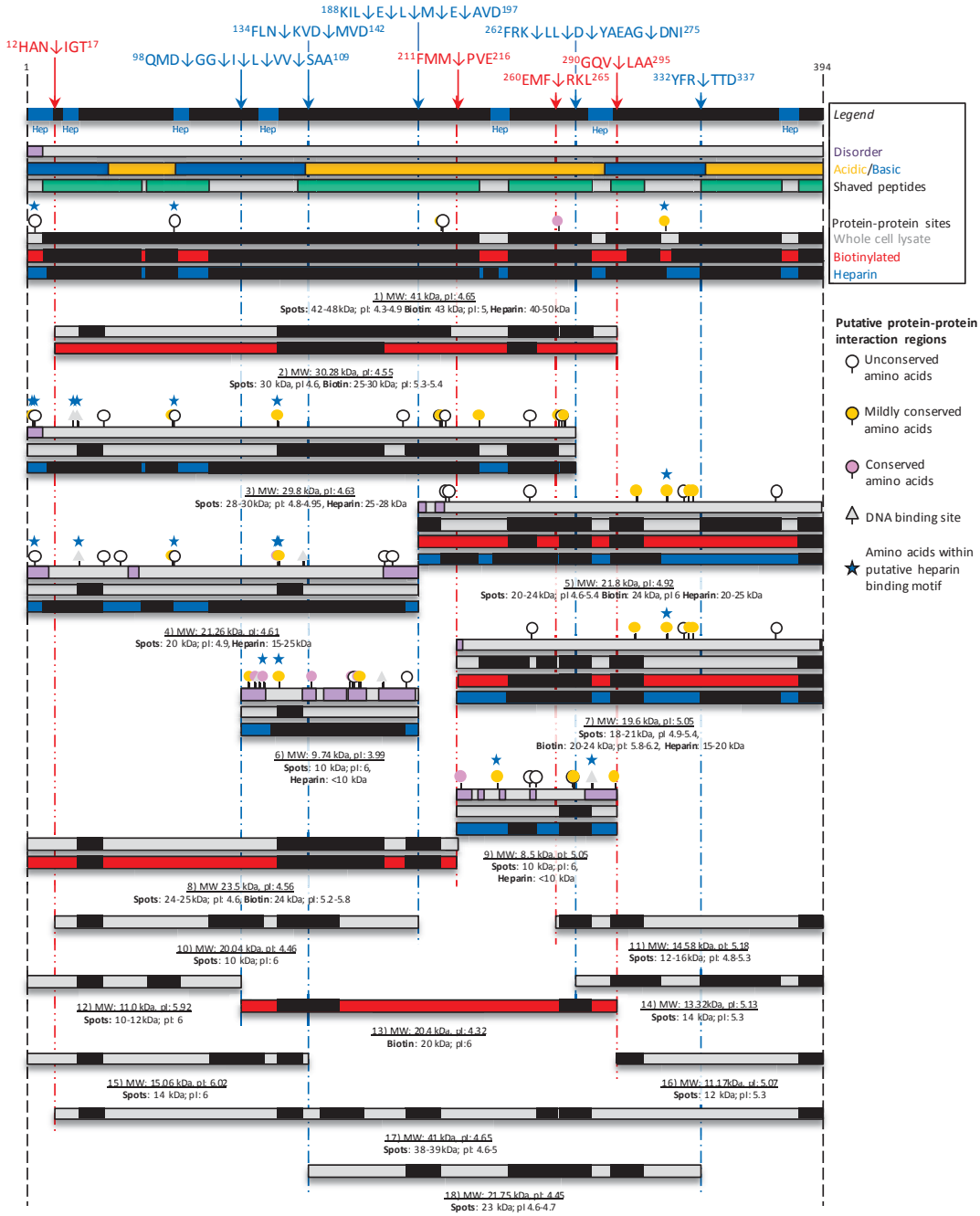
8.5.1: Figure 1: Peptides that map to Ef-Tu identified from surface biotinylation and shaving experiments. The sequence alignment for all three pathogens are shown separated into the three domains of Ef-Tu. Peptides identified from surface trypsin shaving of whole cells are coloured green where red text

indicate peptides from biotinylated surface proteins. Peptides that were identified from both experiments are highlighted as black underlined text. Yellow boxes indicate transmembrane domains for: *M. pneumoniae* adapted from[389] and *M. hyopneumoniae* predicted by TMpred[258].



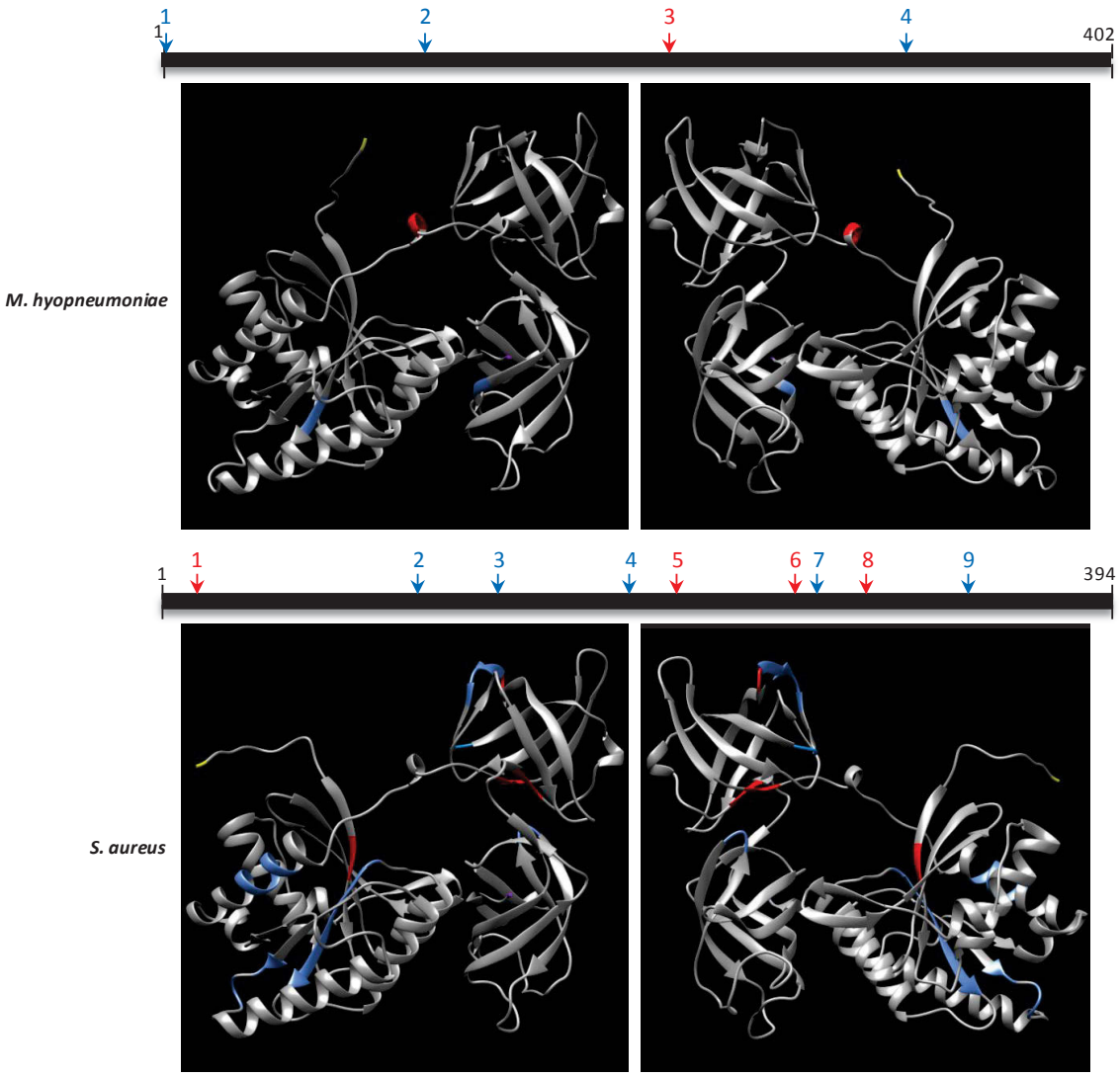
8.5.1: Figure 2: Cleavage map of $Mhp_{Ef\text{-}Tu}$. Full length Ef-Tu is represented as the black bar. Cleavage sites were identified by identifying dimethyl labelled peptides (blue arrows and broken lines) and by characterising semi-tryptic peptides generated after trypsin digestion (red arrows and broken lines). Exact cleavage sites are shown in the amino acid sequences above the black bar.

Bioinformatic tools such as ScanProsite[315], TMpred[258], COILS[312] and Meta-Disorder[384] were used to predict putative heparin binding motifs (Hep, blue boxes), transmembrane domains (TmD, yellow box), coiled-coils (Coil, orange boxes), and disordered regions (purple boxes in grey bars), respectively. Peptides released from trypsin shaving of cells are shown as the green boxes in the grey bar (labelled 'Shaved peptides'). Tryptic peptides identified by mass spectrometry (black boxes within coloured bars) within Ef-Tu fragments were obtained from 1D- and 2D-SDS PAGE of bacterial whole cell lysates (grey bars); avidin affinity chromatography of: biotinylated surface proteins (red bars), PK-15 surface proteins (orange bars), fibronectin (green bars), actin (teal bars), and plasminogen (purple bars); and peptides identified from heparin agarose affinity chromatography (blue bars). Circles and triangles just above fragments denote amino acid positions that are predicted to be surface exposed and represent putative protein:protein interaction regions (visual cues can be seen on the right of the cleavage map, sites listed in Appendix 5: 8.5.2: Table 4). Those marked with an additional star denote amino acid residues that fall within predicted putative heparin binding domains. White circles mark evolutionary unconserved binding regions, whilst yellow circles are mildly conserved and pink are highly conserved. Amino acid positions marked by grey triangles depict predicted nucleic acid interaction regions.

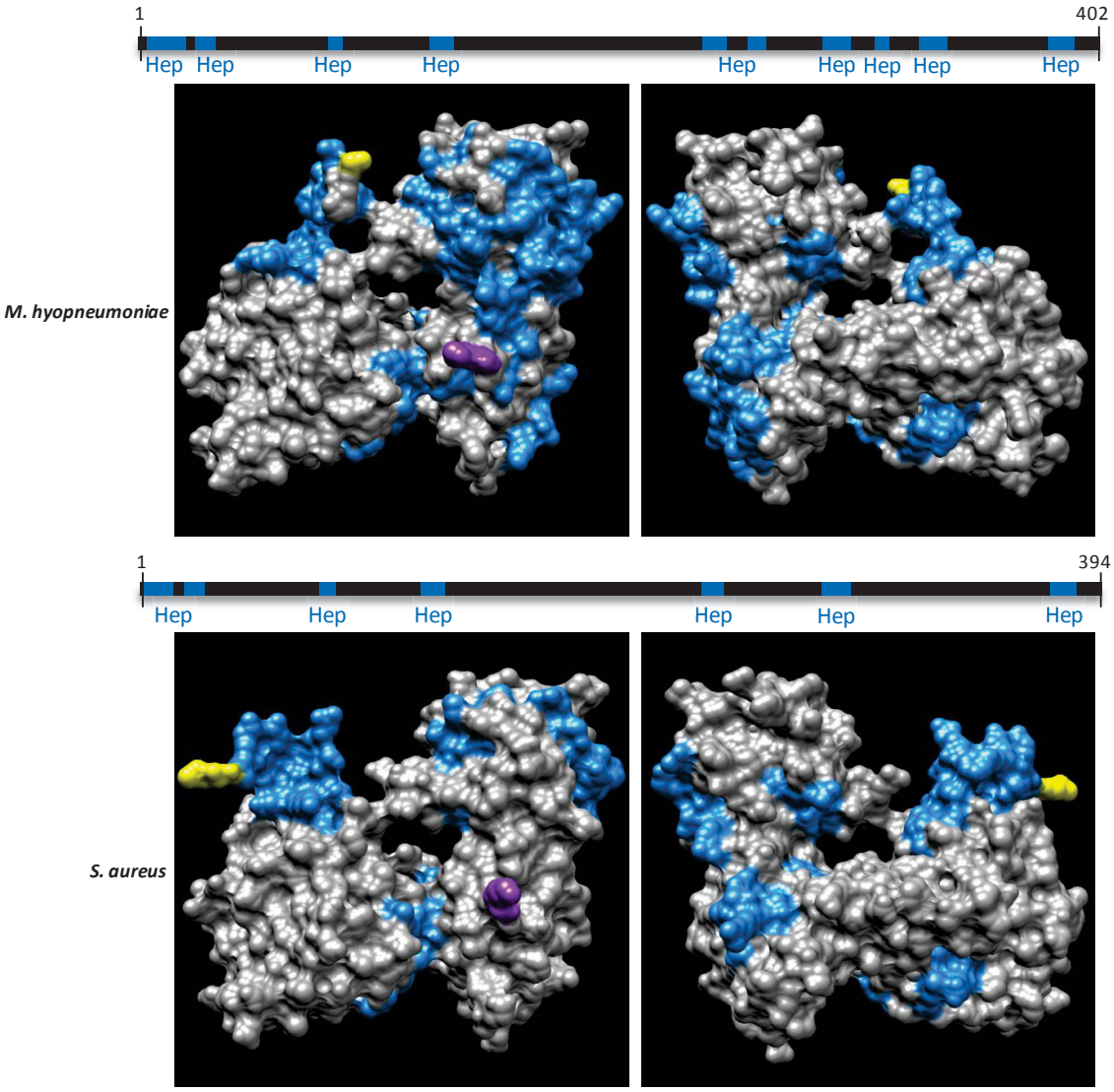


8.5.1: Figure 3: Cleavage map of Ef-Tu$_{Sa}$. Full length Ef-Tu is represented as the black bar. Cleavage sites were identified by identifying dimethyl labelled peptides (blue arrows and broken lines) and by characterising semi-tryptic peptides generated after trypsin digestion (red arrows and broken lines). Exact cleavage sites are shown in the amino acid sequences above the black bar. Bioinformatic tools such as ScanProsite[315] and Meta-Disorder[384] were used to

predict putative heparin binding motifs (Hep, blue boxes) and disordered regions (purple boxes in grey bars), respectively. Peptides released from trypsin shaving of cells are shown as the green boxes in the grey bar (labelled 'Shaved peptides'). Tryptic peptides identified by mass spectrometry (black boxes within coloured bars) within Ef-Tu fragments were obtained from 1D- and 2D-SDS PAGE of bacterial whole cell lysates (grey bars); avidin affinity chromatography of biotinylated surface proteins (red bars), and from heparin agarose affinity chromatography (blue bars). Circles and triangles just above fragments denote amino acid positions that are predicted to be surface exposed and represent putative protein:protein interaction regions (visual cues can be seen on the right of the cleavage map, sites listed in Appendix 5: 8.5.2: Table 5). Those marked with an additional star denote amino acid residues that fall within predicted putative heparin binding domains. White circles mark evolutionary unconserved binding regions, whilst yellow circles are mildly conserved and pink are highly conserved. Amino acid positions marked by grey triangles depict predicted nucleic acid interaction regions.



8.5.1: Figure 4: Predicted 3D structures of $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$. Images are reversed sides for each structure. Ef-Tu proteins are represented as black bars with cleavage sites shown in the ribbon structure. The first amino acid (yellow) and last amino acid (purple) have been coloured in the 3D structures. Identified cleavage sites are displayed as arrows (blue, dimethyl labelling and red, semi-tryptic) with numbers to indicate each distinct cleavage site. Cleavage sites can also be seen in the ribbon structures as blue and red sections for dimethyl labelling and semi-tryptic sites, respectively. Structures were predicted by MODELLER and were based on *E. coli* Ef-Tu homologues: PDB: 1DG1_H (*M. hyopneumoniae*) and PDB: 1DG1_H (*S. aureus*).



8.5.1: Figure 5: Predicted 3D structures of $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$. Images are reversed sides for each structure. Ef-Tu molecules are represented as black bars with heparin binding domains shown as blue boxes. The first amino acid (yellow) and last amino acid (purple) have been coloured in the 3D structures. All 10 and 7 predicted heparin binding sites are shown in blue regions in the surface structures for $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$, respectively. Structures were predicted by MODELLER and were based on *E. coli* Ef-Tu homologues: PDB: 1DG1_H (*M. hyopneumoniae*) and PDB: 1DG1_H (*S. aureus*).

8.5.2 Supplementary Tables

8.5.2: Table 1: Number of binding sites in full length and fragments of $Mpn_{Ef\text{-}Tu}$.

Fragment number	Full length	3	4	5	6	7	8	10	11	12	13
Range (amino acid)	1 – 394	1 – 110	131 – 394	1 – 130	218 – 394	1 – 217	244 – 394	111 – 242	308 – 394	111 – 217	218 – 307
Exposed P:P	8	15	5	13	10	17	6	12	9	19	11
Buried P:P	6	5	2	7	5	2	5	3	6	2	0
DNA-binding	0	8	0	5	0	6	0	3	2	1	1

Analysis of the $Mpn_{Ef\text{-}Tu}$ for the putative protein:protein (P:P) and protein-nucleic acid interaction sites using ISIS[383].

8.5.2: Table 2: Dimethyl labelled and semi-tryptic peptides identified in $Mhp_{Ef\text{-}Tu}$ and $Sa_{Ef\text{-}Tu}$.

No.	ID	Peptide Sequence	Score	E-value
		Dimethyl labelled peptides from $Mhp_{Ef\text{-}Tu}$		
	N1	M.^{2}**A**VVKTTGKKDFR^{14}.S	84	5.70E^{-07}
1	N2	A.^{3}**V**VKTTGKKDFR^{14}.S	36	0.021
	N3	V.^{4}**V**KTTGKKDFR^{14}.S	34	0.019
2	N4	V.^{112}**V**AATDGPMPQTR^{123}.E	74	1.70E^{-05}
4	N5	A.^{316}**A**IYALKKEEGGR^{327}.H	50	3.00E^{-05}
		Semi-tryptic N-terminal Peptides from $Mhp_{Ef\text{-}Tu}$		
3	S1	M.^{215}**D**KPFLMAVEDVFTITGR^{231}.G	68	2.50E^{-05}
		Dimethyl labelled peptides from $Sa_{Ef\text{-}Tu}$		
2	N1	D.^{101}**G**GILVVSAADGPMPQTR^{117}.E	98	2.90E^{-05}
	N2	G.^{103}**I**LVVSAADGPMPQTR^{117}.E	81	1.30E^{-03}
	N3	I.^{104}**L**VVSAADGPMPQTR^{117}.E	104	5.90E^{-06}
	N4	L.^{105}**V**VSAADGPMPQTR^{117}.E	91	1.10E^{-04}
	N5	V.^{107}**S**AADGPMPQTR^{117}.E	69	9.90E^{-03}
3	N6	N.^{137}**K**VDMVDDEELLELVEMEVR^{155}.D	80	3.10E^{-03}

	N7	D.[140]**M**VDDEELLELVEMEVR[155].D	78	3.40E^{-03}
4	N8	L.[191]**E**LMEAVDTYIPTPER[205].D	78	3.60E^{-03}
	N9	E.[192]**L**MEAVDTYIPTPER[205].D	78	3.20E^{-03}
	N10	L.[193]**M**EAVDTYIPTPER[205].D	72	2.80E^{-03}
	N11	M.[194]**E**AVDTYIPTPER[205].D	103	8.40E^{-06}
	N12	E.[195]**A**VDTYIPTPER[205].D	77	9.60E^{-04}
7	N13	K.[265]**L**LDYAEAGDNIGALLR[280].G	98	4.10E^{-05}
	N14	L.[267]**D**YAEAGDNIGALLR[280].G	88	1.80E^{-04}
	N15	D.[268]**Y**AEAGDNIGALLR[280].G	89	2.40E^{-04}
	N16	G.[273]**D**NIGALLR[280].G	67	1.80E^{-02}
9	N17	R.[335]**T**TDVTGVVHLPEGTEMVMPGDNVEMTVELIAPIAIED GTR[374].F	86	7.70E^{-07}
		Semi-tryptic N-terminal Peptides from $Sa_{Ef\text{-}Tu}$		
8	S1	V.[293]**L**AAPGSITPHTEFK[306].A	106	3.20E^{-05}
5	S2	M.[214]**P**VEDVFSITGR[224].G	80	0.010
1	S3	N.[15]**I**GTIGHVDHGK[28].T	80	1.10E^{-04}
6	S4	F.[263]**R**KLLDYAEAGDNIGALLR[280].G	91	1.20E^{-03}

Identified peptides have a Mascot score > 33 and an E-value < 0.05. The exact site of cleavage is to the left of the amino acid that is bold and underlined for N-terminal cleavage fragments and to the right of C-terminal cleavage fragments. Amino acid numbers are written at the start and end of each peptide identified by LC-MS/MS.

8.5.2: Table 3: Putative heparin binding motifs identified in $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$.

Amino acid range	Sequence
	$Mhp_{Ef\text{-}Tu}$ (Uniprot #: Q4A9G1)
4 – 19:	VKTTGKKDFDRSKEHI
25 – 32:	GHVDHGKT^{+}
79 – 86:	DKRHYAHV*
122 – 131:	TREHILLSKQ^{+}
237 – 246:	GKVERGQVKL^{+}
255 – 262:	YREEPKKT
287 – 298:	LRGVDRKDIERG
302 – 315:	AKPKTIIPHTKFKA
327 – 337:	GRHTPFFKNYKP
381 – 391:	TKFSIREGGRT^{+}
	$Sa_{Ef\text{-}Tu}$ (Uniprot #: Q2G0N0)
2 – 13:	AKEKFDRSKEHA
19 – 26:	GHVDHGKT^{+}
73 – 80:	DKRHYAHV*
116 – 125:	TREHILLSRN^{+}
230 – 239:	GRVERGQIKV^{+}
279 – 290:	LRGVAREDVQRG
373 – 383:	TRFSIREGGRT^{+}

$Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$ was searched for patters x-[HKR]-x(0,2)-[HKR]-x(0,2)-[HKR]-x and x-[HKR]-x(1,4)-[HKR]-x(1,4)-[HKR]-x using ScanProsite (142). * indicates the motif (dKRHyaH) which is found in all three pathogens. $^{+}$ indicates the motifs that are highly homologous (up to four non-basic residues different) in all three pathogens.

8.5.2: Table 4: Number of binding sites in full length and fragments of $Mhp_{Ef\text{-}Tu}$.

Fragment number	Full length	1	4	5	6
Range (amino acid)	Full	1 – 315	112 – 402	215 – 402	1 – 215
Exposed P:P	10	13	9	10	18
Buried P:P	2	3	1	3	4
DNA-binding	0	3	0	0	4

Analysis of the $Mhp_{Ef\text{-}Tu}$ for the putative protein:protein (P:P) and protein-nucleic acid interaction sites using ISIS[383].

8.5.2: Table 5: Number of binding sites in full length and fragments of $Sa_{Ef\text{-}Tu}$.

Fragment number	Full length	3	4	5	6	7	9
Range (amino acid)	Full	1 – 266	1 – 192	193 – 394	104 – 192	214 – 394	214 – 292
Exposed P:P	10	13	10	7	12	5	10
Buried P:P	6	12	6	5	2	5	1
DNA-binding	0	3	4	0	2	0	1

Analysis of the $Sa_{Ef\text{-}Tu}$ for the putative protein:protein (P:P) and protein-nucleic acid interaction sites using ISIS[383].

8.5.3 Supplementary Methodology

8.5.3.1 Strains and cultures and reagents

M. hyopneumoniae (J strain) was cultured in modified Friis medium at 37°C with shaking as described previously[484,485].

S. aureus (SH 1000 strain) was cultured in TSB (Oxoid) at 37°C with shaking and harvested during early stationary phase. Protease inhibitors (Roche Diagnostics®) in PBS were added to the cells during harvest and washes with PBS. For *S. aureus* lysis, cell pellets were freeze-dried overnight before added to pre-cooled metal milling canisters with 12 small metal beads. The canister was cooled in liquid nitrogen and milled at a maximum frequency of 30 Hz for 1 minute for 15 rounds; cooling in liquid nitrogen between rounds.

Proteins were than solubilised in 7 M urea, 2 M thiourea, 50 mM LiCl, 50 mM Tris-HCl (pH 8.8), 1% (w/v) C7bZ0 with protease inhibitors followed by sonication at maximum intensity for 30 seconds for 20 rounds resting on ice in between.

Porcine kidney epithelial (PK-15) cells were cultured in DMEM medium (Invitrogen) supplemented with 10% heat inactivated fetal bovine serum at 37°C with 5% CO_2 in tissue culture flasks.

8.5.3.2 Enrichment of *M. hyopneumoniae* and *S. aureus* surface proteins

8.5.3.2.1 Biotinylation

M. hyopneumoniae and *S. aureus* cells were washed with PBS after centrifugation before the resuspending in EZ-link sulfo-NHS-biotin (Thermo Fisher Scientific). *M. hyopneumoniae* and *S. aureus* cells were biotinylated for 30 seconds and 1 minute, respectively. Quenching, lysis (for *M. hyopneumoniae*), avidin purification and western blotting were the same as for *M. pneumoniae*. Lysis for *S. aureus* cells is described above.

8.5.3.2.2 Triton X-114 phase extraction of biotinylated *M. hyopneumoniae* proteins

Triton X-114 phase extraction of proteins was carried as described in[186,225,486]and biotinylated surface proteins were purified by avidin column chromatography.

8.5.3.2.3 Trypsin shaving

Trypsin shaving of *M. pneumoniae* cells was carried out as described previously[187] with modifications. *M. hyopneumoniae* and *S. aureus* cells were resuspended in trypsin for 5 minutes at 37°C.

8.5.3.3 Whole cell lysis preparation for one- and two-dimensional gel electrophoresis

M. hyopneumoniae whole cell lysates were prepared as previously described[187]. Lysis for *S. aureus* cells is described above. Proteins were reduced and alkylated with 5 mM tributylphosphine and 20 mM acrylamide monomers for 90 min at room temperature. Insoluble material was removed by centrifugation and five volumes of acetone added to precipitate protein. After centrifugation, the protein pellet was solubilized in 7 M urea, 2 M thiourea, 1% (w/v) C7BzO for one- and two-dimensional gel electrophoresis.

8.5.3.4 Heparin affinity chromatography

Affinity purification of *M. hyopneumoniae* and *S. aureus* heparin binding proteins performed as described in[230] with modifications. *M. hyopneumoniae* cells were and lysed in 10 mM sodium phosphate, pH 7 with three 30 second rounds of sonication. *S. aureus* cells were lysed as described in section 2.1 except that protein was solubilised in 10 mM sodium phosphate, pH 7 with protease inhibitors followed by sonication at maximum intensity for 30 seconds for 4 rounds resting on ice in between. After centrifugation, ~300 µg of soluble protein from both *M. hyopneumoniae* and *S. aureus* lysates were treated exactly the same as *M. pneumoniae*.

8.5.3.5 Avidin purification of host binding *M. hyopneumoniae* proteins

Purified fibronectin (Merck Millipore), plasminogen (Sigma) and actin (Sigma) used in this section are described in '5.6.1 Strains and cultures and reagents'. Avidin purification of these host binding *M. hyopneumoniae* proteins was performed as described in[226]. Avidin purification of *M. hyopneumoniae* proteins that bind PK-15 surface proteins was performed as described in[224].

8.5.3.6 Extra Peptide search parameters

Files were searched against the MSPnr100 database[251] with the following parameters. Fixed modifications: none. Variable modifications: propionamide, oxidized methionine, deamidation. Enzyme: semi-trypsin. Number of allowed missed cleavages: 3. Peptide mass tolerance: 100 ppm. MS/MS mass tolerance: 0.2 Da. Charge state: 2+, 3+ and 4+. For samples collected from the 'Avidin purification of PK-15 interacting proteins', variable modifications also included NHS-LC-Biotin (K) and NHS-LC-Biotin (N-term). 'Avidin purification of PK-15 interacting proteins' was also searched against *sus scrofa* entries in MSPnr100 to identify biotinylated surface PK-15 proteins. *S. aureus* proteins were also searched against a *S. aureus* NCTC 8325 database derived from the published genome.

8.5.3.7 LC-MS/MS of dimethyl labelled proteins

8.5.3.7.1 LC-MS/MS (Sciex 5600) of dimethyl labelled proteins

Peptides from dimethyl labelled proteins described in section 2.16.1 were separated by nanoLC using an Ultimate nanoRSLC UPLC and autosampler system (Dionex, Amsterdam, Netherlands). Samples (2.5 µl) were concentrated and desalted onto a micro C18 precolumn (300 µm x 5 mm, Dionex) with H_2O:CH_3CN (98:2, 0.1 % TFA) at 15 µl/min. After a 4 min wash the pre-column was switched (Valco 10 port UPLC valve, Valco, Houston, TX) into line with a fritless nano column (75µ x ~15cm) containing C18AQ media (1.9µ, 120 Å Dr Maisch, Ammerbuch-Entringen Germany). 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 200 nl/min over 240 min. High voltage 2000 V was applied to low volume Titanium union (Valco) with the tip positioned ~ 0.5 cm from the curtain plate (T=150°C) of a 5600^+ mass spectrometer (Sciex, Toronto, Canada). Positive ions were generated by electrospray and the 5600^+ operated in information dependent acquisition mode (IDA).

A survey scan m/z 350-1750 was acquired (PWHH resolution ~30,000, 0.25 sec acquisition time) with autocalibration enabled (at ~6 hr intervals). Up to the 10 most abundant ions (>300 counts) with charge states > +2 and <+5 were sequentially isolated (width m/z ~3) and fragmented by CID with an optimal CE chosen based on m/z (product ion spectra were acquired at a resolution ~20,000 PWHH in 0.15 sec). M/z ratios selected for MS/MS were dynamically excluded for 30 or 45 seconds.

Peak lists were generated using Mascot Daemon/Mascot Distiller (Matrix Science, London, England) or ProteinPilot (Sciex, v4.5) using default parameters, and submitted to the database search program Mascot (version 2.5.1, Matrix Science). Search parameters were: Precursor tolerance 10 ppm and product ion tolerances ± 0.05 Da; oxidation (M), deamidation (NQ), propionamide (C), Dimethyl (K), Dimethyl (N-term) specified as variable modifications; enzyme specificity was semi-ArgC; 1 missed cleavage was possible and the non-redundant protein database from NCBI (Jan 2015) searched.

8.5.3.7.2 LC-MS/MS (Thermo Scientific Q Exactive™) of dimethyl labelled proteins

Peptides from dimethyl labelled proteins described in section 2.18.1 were separated by nanoLC using an Ultimate nanoRSLC UPLC and autosampler system (Dionex, Amsterdam, Netherlands). Samples (2.5 µl) were concentrated and desalted onto a micro C18 precolumn (300 µm x 5 mm, Dionex) with H_2O:CH_3CN (98:2, 0.1 % TFA) at 15 µl/min. After a 4 min wash the pre-column was switched (Valco 10 port UPLC valve, Valco, Houston, TX) into line with a fritless nano column (75µ x ~35cm) containing C18AQ media (1.9µ, 120 Å Dr Maisch, Ammerbuch-Entringen Germany). 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 200 nl/min over 30 or 240 min. High voltage 2000 V was applied to low volume Titanium union (Valco) with the column oven heated to 45°C (Sonation, Biberach, Germany) and the tip positioned ~ 0.5 cm from the heated capillary (T=300°C) of a QExactive Plus mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Positive ions were generated by electrospray and the QExactive operated in data dependent acquisition mode (DDA).

A survey scan m/z 350-1750 was acquired (resolution = 70,000 at m/z 200, with an AGC target value of 10^6 ions) and lockmass was enabled (m/z 445.12003). Up to the 10 most abundant ions (>80,000 counts, underfill ratio 10%) with charge states > +2 and <+7 were sequentially isolated (width m/z 2.5) and fragmented by HCD (NCE = 30) with a AGC target of 10^5 ions (resolution = 17,500 at m/z 200). M/z ratios selected for MS/MS were dynamically excluded for 30 or 45 seconds.

Peak lists were generated using Mascot Daemon/Mascot Distiller (Matrix Science, London, England) or Proteome Discoverer (Thermo, v1.4) using default parameters, and submitted to the database search program Mascot (version 2.5.1, Matrix Science). Search parameters were: Precursor tolerance 4 ppm and product ion tolerances ± 0.05 Da; oxidation (M), deamidation (NQ), propionamide (C), Dimethyl (K), Dimethyl (N-term) specified as variable modifications; enzyme specificity was semi-ArgC; 1 missed cleavage was possible and the non-redundant protein database from NCBI (Jan 2015) searched.

8.5.3.8 Supplementary Bioinformatics: Analysis of conservation of amino acids in $Mpn_{Ef\text{-}Tu}$, $Mhp_{Ef\text{-}Tu}$, and $Sa_{Ef\text{-}Tu}$

Degree of conservation in $Mpn_{Ef\text{-}Tu}$ was calculated by The ConSurf server[382]. Colours indicate degree of conservation of the amino acid across species. Red diamond sticks indicate predicted protein binding sites by ISIS[383,487]. Grey, yellow and purple circle sticks indicate predicted nucleotide, DNA and RNA binding regions, respectively by SomeNA[487,488]. Below the sticks are three rows: i) Blue and red bars indicate predicted beta-strand and helix secondary structures, respectively by REPROFSec[386,487], ii) Blue and yellow bars indicate predicted regions that are exposed and buried to solvent accessibilities, respectively by PROFAcc[386,487]. iii) Green bars indicate predicted disordered regions by Meta-Disorder (MD)[385,487]. Circled amino acids represent predicted binding sites. Amino acids indicated by a triangle indicate predicted nucleotide binding sites.

Pleae find the supplementary bioinformatics data for this section on the attached CD.

References

1 Waites, K. B. & Talkington, D. F. Mycoplasma pneumoniae and its role as a human pathogen. *Clin Microbiol Rev* **17**, 697-728, table of contents, doi:10.1128/CMR.17.4.697-728.2004 (2004).

2 Othman, N., Isaacs, D. & Kesson, A. Mycoplasma pneumoniae infections in Australian children. *J Paediatr Child Health* **41**, 671-676, doi:10.1111/j.1440-1754.2005.00757.x (2005).

3 Hong, S. J. The Role of Mycoplasma pneumoniae Infection in Asthma. *Allergy Asthma Immunol Res* **4**, 59-61, doi:10.4168/aair.2012.4.2.59 (2012).

4 Lind, K., Benzon, M. W., Jensen, J. S. & Clyde, W. A., Jr. A seroepidemiological study of Mycoplasma pneumoniae infections in Denmark over the 50-year period 1946-1995. *Eur J Epidemiol* **13**, 581-586 (1997).

5 Chalker, V. J. *et al.* Mycoplasma pneumoniae infection in primary care investigated by real-time PCR in England and Wales. *Eur J Clin Microbiol Infect Dis* **30**, 915-921, doi:10.1007/s10096-011-1176-3 (2011).

6 Jacobs, E. Mycoplasma pneumoniae: now in the focus of clinicians and epidemiologists. *Euro Surveill* **17** (2012).

7 Nir-Paz, R., Abutbul, A., Moses, A. E., Block, C. & Hidalgo-Grass, C. Ongoing epidemic of Mycoplasma pneumoniae infection in Jerusalem, Israel, 2010 to 2012. *Euro Surveill* **17** (2012).

8 Brown, R. J. *et al.* Mycoplasma pneumoniae Epidemiology in England and Wales: A National Perspective. *Front Microbiol* **7**, 157, doi:10.3389/fmicb.2016.00157 (2016).

9 Sanchez-Vargas, F. M. & Gomez-Duarte, O. G. Mycoplasma pneumoniae-an emerging extra-pulmonary pathogen. *Clin Microbiol Infect* **14**, 105-117, doi:10.1111/j.1469-0691.2007.01834.x (2008).

10 Atkinson, T. P., Balish, M. F. & Waites, K. B. Epidemiology, clinical manifestations, pathogenesis and laboratory detection of Mycoplasma pneumoniae infections. *FEMS Microbiol Rev* **32**, 956-973, doi:10.1111/j.1574-6976.2008.00129.x (2008).

11 Youn, Y. S. & Lee, K. Y. Mycoplasma pneumoniae pneumonia in children. *Korean J Pediatr* **55**, 42-47, doi:10.3345/kjp.2012.55.2.42 (2012).

12 Mogabgab, W. J. Mycoplasma pneumoniae and adenovirus respiratory illnesses in military and university personnel, 1959-1966. *Am Rev Respir Dis* **97**, 345-358, doi:10.1164/arrd.1968.97.3.345 (1968).

13 Klement, E. *et al.* Identification of risk factors for infection in an outbreak of Mycoplasma pneumoniae respiratory tract disease. *Clin Infect Dis* **43**, 1239-1245, doi:10.1086/508458 (2006).

14 Kashiwagi, S. *et al.* An outbreak of Mycoplasma pneumoniae infections in a hospital in Japan. *Kurume Med J* **32**, 293-296 (1985).

15 Kleemola, M. & Jokinen, C. Outbreak of Mycoplasma pneumoniae infection among hospital personnel studied by a nucleic acid hybridization test. *J Hosp Infect* **21**, 213-221 (1992).

16 Klausner, J. D. *et al.* Enhanced control of an outbreak of Mycoplasma pneumoniae pneumonia with azithromycin prophylaxis. *J Infect Dis* **177**, 161-166 (1998).

17 Dorigo-Zetsma, J. W., de Wit, M., Szabo, J. S. & Schneeberger, P. M. Epidemic of respiratory tract infections by Mycoplasma pneumoniae in an institute for mentally

disabled, investigated with polymerase chain reaction of a throat swab specimen. *Ned Tijdschr Geneeskd* **143**, 1261-1265 (1999).

18 Leibowitz, Z., Schvartzman, P., Epstein, L., Lis, I. & Naot, Y. An outbreak of Mycoplasma pneumoniae pneumonia in two kibbutzim: a clinical and epidemiologic study. *Isr J Med Sci* **24**, 88-92 (1988).

19 Muldoon, R. L., Raucci, J., Kowalski, J. & Rajashekaraiah, K. An outbreak of mycoplasma pneumoniae respiratory illness in a semi-closed religious commune. *Ann Emerg Med* **11**, 613-615 (1982).

20 Waring, A. L. *et al.* Development of a genomics-based PCR assay for detection of Mycoplasma pneumoniae in a large outbreak in New York State. *J Clin Microbiol* **39**, 1385-1390, doi:10.1128/JCM.39.4.1385-1390.2001 (2001).

21 Muse, K. E., Powell, D. A. & Collier, A. M. Mycoplasma pneumoniae in hamster tracheal organ culture studied by scanning electron microscopy. *Infect Immun* **13**, 229-237 (1976).

22 Izumikawa, K. Clinical Features of Severe or Fatal Mycoplasma pneumoniae Pneumonia. *Front Microbiol* **7**, 800, doi:10.3389/fmicb.2016.00800 (2016).

23 Staugas, R. & Martin, A. J. Secondary bacterial infections in children with proved Mycoplasma pneumoniae. *Thorax* **40**, 546-548 (1985).

24 Niederman, M. S. Guidelines for the management of community-acquired pneumonia. Current recommendations and antibiotic selection issues. *Med Clin North Am* **85**, 1493-1509 (2001).

25 Suzuki, S. *et al.* Clinical evaluation of macrolide-resistant Mycoplasma pneumoniae. *Antimicrob Agents Chemother* **50**, 709-712, doi:10.1128/AAC.50.2.709-712.2006 (2006).

26 Okazaki, N. *et al.* Characteristics of macrolide-resistant Mycoplasma pneumoniae strains isolated from patients and induced with erythromycin in vitro. *Microbiol Immunol* **45**, 617-620 (2001).

27 Baseman, J. B., Cole, R. M., Krause, D. C. & Leith, D. K. Molecular basis for cytadsorption of Mycoplasma pneumoniae. *J Bacteriol* **151**, 1514-1522 (1982).

28 Su, C. J., Chavoya, A., Dallo, S. F. & Baseman, J. B. Sequence divergency of the cytadhesin gene of Mycoplasma pneumoniae. *Infect Immun* **58**, 2669-2674 (1990).

29 Su, C. J., Dallo, S. F., Chavoya, A. & Baseman, J. B. Possible origin of sequence divergence in the P1 cytadhesin gene of Mycoplasma pneumoniae. *Infect Immun* **61**, 816-822 (1993).

30 Kenri, T. *et al.* Identification of a new variable sequence in the P1 cytadhesin gene of Mycoplasma pneumoniae: evidence for the generation of antigenic variation by DNA recombination between repetitive sequences. *Infect Immun* **67**, 4557-4562 (1999).

31 Dorigo-Zetsma, J. W., Wilbrink, B., Dankert, J. & Zaat, S. A. Mycoplasma pneumoniae P1 type 1- and type 2-specific sequences within the P1 cytadhesin gene of individual strains. *Infect Immun* **69**, 5612-5618 (2001).

32 Spuesens, E. B. *et al.* Sequence variations in RepMP2/3 and RepMP4 elements reveal intragenomic homologous DNA recombination events in Mycoplasma pneumoniae. *Microbiology* **155**, 2182-2196, doi:10.1099/mic.0.028506-0 (2009).

33 Himmelreich, R. *et al.* Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. *Nucleic Acids Res* **24**, 4420-4449 (1996).

34 Dandekar, T. *et al.* Re-annotating the Mycoplasma pneumoniae genome sequence: adding value, function and reading frames. *Nucleic Acids Res* **28**, 3278-3288 (2000).

35 Chen, W. H. *et al.* Integration of multi-omics data of a genome-reduced bacterium: Prevalence of post-transcriptional regulation and its correlation with protein abundances. *Nucleic Acids Res* **44**, 1192-1202, doi:10.1093/nar/gkw004 (2016).

36 Krishnakumar, R. *et al.* Targeted chromosomal knockouts in Mycoplasma pneumoniae. *Appl Environ Microbiol* **76**, 5297-5299, doi:10.1128/AEM.00024-10 (2010).

37 Spuesens, E. B. *et al.* Comparison of Mycoplasma pneumoniae Genome Sequences from Strains Isolated from Symptomatic and Asymptomatic Patients. *Front Microbiol* **7**, 1701, doi:10.3389/fmicb.2016.01701 (2016).

38 Lluch-Senar, M. *et al.* Comparative "-omics" in Mycoplasma pneumoniae Clinical Isolates Reveals Key Virulence Factors. *PLoS One* **10**, e0137354, doi:10.1371/journal.pone.0137354 (2015).

39 Xiao, L. *et al.* Comparative genome analysis of Mycoplasma pneumoniae. *BMC Genomics* **16**, 610, doi:10.1186/s12864-015-1801-0 (2015).

40 Guell, M. *et al.* Transcriptome complexity in a genome-reduced bacterium. *Science* **326**, 1268-1271, doi:10.1126/science.1176951 (2009).

41 Vivancos, A. P., Guell, M., Dohm, J. C., Serrano, L. & Himmelbauer, H. Strand-specific deep sequencing of the transcriptome. *Genome Res* **20**, 989-999, doi:10.1101/gr.094318.109 (2010).

42 Regula, J. T. *et al.* Defining the mycoplasma 'cytoskeleton': the protein composition of the Triton X-100 insoluble fraction of the bacterium Mycoplasma pneumoniae determined by 2-D gel electrophoresis and mass spectrometry. *Microbiology* **147**, 1045-1057, doi:10.1099/00221287-147-4-1045 (2001).

43 Ueberle, B., Frank, R. & Herrmann, R. The proteome of the bacterium Mycoplasma pneumoniae: comparing predicted open reading frames to identified gene products. *Proteomics* **2**, 754-764, doi:10.1002/1615-9861(200206)2:6<754::AID-PROT754>3.0.CO;2-2 (2002).

44 Jaffe, J. D., Berg, H. C. & Church, G. M. Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* **4**, 59-77, doi:10.1002/pmic.200300511 (2004).

45 Yogev, D., Browning, G. F. & K.S., W. *Molecular Biology and Pathogenicity of Mycoplasmas (Chapter 19: Genetic Mechanisms of Surface Variation)*. 417 - 443 (Springer US, 2002).

46 Maier, T. *et al.* Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol Syst Biol* **7**, 511, doi:10.1038/msb.2011.38 (2011).

47 Borras, E. *et al.* Integrative quantitation enables a comprehensive proteome comparison of two Mycoplasma pneumoniae genetic perturbations. *Mol Biosyst* **9**, 1249-1256, doi:10.1039/c3mb25581f (2013).

48 Yus, E. *et al.* Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326**, 1263-1268, doi:10.1126/science.1177263 (2009).

49 Kuhner, S. *et al.* Proteome organization in a genome-reduced bacterium. *Science* **326**, 1235-1240, doi:10.1126/science.1176343 (2009).

50 van Noort, V. *et al.* Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol Syst Biol* **8**, 571, doi:10.1038/msb.2012.4 (2012).

51 Su, H. C., Hutchison, C. A., 3rd & Giddings, M. C. Mapping phosphoproteins in Mycoplasma genitalium and Mycoplasma pneumoniae. *BMC Microbiol* **7**, 63, doi:10.1186/1471-2180-7-63 (2007).

52 Schmidl, S. R. *et al.* The phosphoproteome of the minimal bacterium Mycoplasma pneumoniae: analysis of the complete known Ser/Thr kinome suggests the existence of novel kinases. *Mol Cell Proteomics* **9**, 1228-1242, doi:10.1074/mcp.M900267-MCP200 (2010).

53 Lluch-Senar, M. *et al.* Comprehensive methylome characterization of Mycoplasma genitalium and Mycoplasma pneumoniae at single-base resolution. *PLoS Genet* **9**, e1003191, doi:10.1371/journal.pgen.1003191 (2013).

54 Lluch-Senar, M. *et al.* Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. *Mol Syst Biol* **11**, 780, doi:10.15252/msb.20145558 (2015).

55 Junier, I., Unal, E. B., Yus, E., Llorens-Rico, V. & Serrano, L. Insights into the Mechanisms of Basal Coordination of Transcription Using a Genome-Reduced Bacterium. *Cell Syst* **2**, 391-401, doi:10.1016/j.cels.2016.04.015 (2016).

56 Wodke, J. A. *et al.* MyMpn: a database for the systems biology model organism Mycoplasma pneumoniae. *Nucleic Acids Res* **43**, D618-623, doi:10.1093/nar/gku1105 (2015).

57 Katz, B. & Waites, K. Emerging intracellular bacterial infections. *Clin Lab Med* **24**, 627-649, vi, doi:10.1016/j.cll.2004.05.005 (2004).

58 Biberfeld, G. & Biberfeld, P. Ultrastructural features of Mycoplasma pneumoniae. *J Bacteriol* **102**, 855-861 (1970).

59 Hegermann, J., Herrmann, R. & Mayer, F. Cytoskeletal elements in the bacterium Mycoplasma pneumoniae. *Naturwissenschaften* **89**, 453-458, doi:10.1007/s00114-002-0359-2 (2002).

60 Dallo, S. F., Horton, J. R., Su, C. J. & Baseman, J. B. Restriction fragment length polymorphism in the cytadhesin P1 gene of human clinical isolates of Mycoplasma pneumoniae. *Infect Immun* **58**, 2017-2020 (1990).

61 Su, C. J., Dallo, S. F. & Baseman, J. B. Molecular distinctions among clinical isolates of Mycoplasma pneumoniae. *J Clin Microbiol* **28**, 1538-1540 (1990).

62 Narita, M. Classification of Extrapulmonary Manifestations Due to Mycoplasma pneumoniae Infection on the Basis of Possible Pathogenesis. *Front Microbiol* **7**, 23, doi:10.3389/fmicb.2016.00023 (2016).

63 Biscardi, S. *et al.* Mycoplasma pneumoniae and asthma in children. *Clin Infect Dis* **38**, 1341-1346, doi:10.1086/392498 (2004).

64 Nisar, N., Guleria, R., Kumar, S., Chand Chawla, T. & Ranjan Biswas, N. Mycoplasma pneumoniae and its role in asthma. *Postgrad Med J* **83**, 100-104, doi:10.1136/pgmj.2006.049023 (2007).

65 Wood, P. R. *et al.* Mycoplasma pneumoniae in children with acute and refractory asthma. *Ann Allergy Asthma Immunol* **110**, 328-334 e321, doi:10.1016/j.anai.2013.01.022 (2013).

66 Kraft, M. *et al.* Detection of Mycoplasma pneumoniae in the airways of adults with chronic asthma. *Am J Respir Crit Care Med* **158**, 998-1001, doi:10.1164/ajrccm.158.3.9711092 (1998).

67 Smith-Norowitz, T. A. *et al.* Asthmatic children have increased specific anti-Mycoplasma pneumoniae IgM but not IgG or IgE-values independent of history of respiratory tract infection. *Pediatr Infect Dis J* **32**, 599-603, doi:10.1097/INF.0b013e3182862ea8 (2013).

68 Petitjean, J., Vabret, A., Gouarin, S. & Freymuth, F. Evaluation of four commercial immunoglobulin G (IgG)- and IgM-specific enzyme immunoassays for diagnosis of Mycoplasma pneumoniae infections. *J Clin Microbiol* **40**, 165-171 (2002).

69 Daxboeck, F., Krause, R. & Wenisch, C. Laboratory diagnosis of Mycoplasma pneumoniae infection. *Clin Microbiol Infect* **9**, 263-273 (2003).

70 Martinez, M. A., Ruiz, M., Zunino, E., Luchsinger, V. & Avendano, L. F. Detection of Mycoplasma pneumoniae in adult community-acquired pneumonia by PCR and serology. *J Med Microbiol* **57**, 1491-1495, doi:10.1099/jmm.0.2008/003814-0 (2008).

71 Nilsson, A. C., Bjorkman, P. & Persson, K. Polymerase chain reaction is superior to serology for the diagnosis of acute Mycoplasma pneumoniae infection and reveals a high rate of persistent infection. *BMC Microbiol* **8**, 93, doi:10.1186/1471-2180-8-93 (2008).

72 Chaudhry, R. *et al.* Molecular detection of Mycoplasma pneumoniae by quantitative real-time PCR in patients with community acquired pneumonia. *Indian J Med Res* **138**, 244-251 (2013).

73 Zhang, L., Zong, Z. Y., Liu, Y. B., Ye, H. & Lv, X. J. PCR versus serology for diagnosing Mycoplasma pneumoniae infection: a systematic review & meta-analysis. *Indian J Med Res* **134**, 270-280 (2011).

74 Ferguson, G. D. *et al.* Clinical outcomes and macrolide resistance in Mycoplasma pneumoniae infection in Scotland, UK. *J Med Microbiol* **62**, 1876-1882, doi:10.1099/jmm.0.066191-0 (2013).

75 Peuchant, O. *et al.* Increased macrolide resistance of Mycoplasma pneumoniae in France directly detected in clinical specimens by real-time PCR and melting curve analysis. *J Antimicrob Chemother* **64**, 52-58, doi:10.1093/jac/dkp160 (2009).

76 Dumke, R., von Baum, H., Luck, P. C. & Jacobs, E. Occurrence of macrolide-resistant Mycoplasma pneumoniae strains in Germany. *Clin Microbiol Infect* **16**, 613-616, doi:10.1111/j.1469-0691.2009.02968.x (2010).

77 Chironna, M. *et al.* Emergence of macrolide-resistant strains during an outbreak of Mycoplasma pneumoniae infections in children. *J Antimicrob Chemother* **66**, 734-737, doi:10.1093/jac/dkr003 (2011).

78 Meyer Sauteur, P. M. *et al.* Survey of macrolide-resistant Mycoplasma pneumoniae in children with community-acquired pneumonia in Switzerland. *Swiss Med Wkly* **144**, w14041, doi:10.4414/smw.2014.14041 (2014).

79 Averbuch, D., Hidalgo-Grass, C., Moses, A. E., Engelhard, D. & Nir-Paz, R. Macrolide resistance in Mycoplasma pneumoniae, Israel, 2010. *Emerg Infect Dis* **17**, 1079-1082, doi:10.3201/eid/1706.101558 (2011).

80 Okada, T. *et al.* Rapid effectiveness of minocycline or doxycycline against macrolide-resistant Mycoplasma pneumoniae infection in a 2011 outbreak among Japanese children. *Clin Infect Dis* **55**, 1642-1649, doi:10.1093/cid/cis784 (2012).

81 Wolff, B. J., Thacker, W. L., Schwartz, S. B. & Winchell, J. M. Detection of macrolide resistance in Mycoplasma pneumoniae by real-time PCR and high-resolution melt analysis. *Antimicrob Agents Chemother* **52**, 3542-3549, doi:10.1128/AAC.00582-08 (2008).

82 Li, X. *et al.* Emerging macrolide resistance in Mycoplasma pneumoniae in children: detection and characterization of resistant isolates. *Pediatr Infect Dis J* **28**, 693-696, doi:10.1097/INF.0b013e31819e3f7a (2009).

83 Yamada, M., Buller, R., Bledsoe, S. & Storch, G. A. Rising rates of macrolide-resistant Mycoplasma pneumoniae in the central United States. *Pediatr Infect Dis J* **31**, 409-400, doi:10.1097/INF.0b013e318247f3e0 (2012).

84 Zhao, F. *et al.* Surveillance of macrolide-resistant Mycoplasma pneumoniae in Beijing, China, from 2008 to 2012. *Antimicrob Agents Chemother* **57**, 1521-1523, doi:10.1128/AAC.02060-12 (2013).

85 Kawai, Y. *et al.* Nationwide surveillance of macrolide-resistant Mycoplasma pneumoniae infection in pediatric patients. *Antimicrob Agents Chemother* **57**, 4046-4049, doi:10.1128/AAC.00663-13 (2013).

86 Eshaghi, A. *et al.* Macrolide-resistant Mycoplasma pneumoniae in humans, Ontario, Canada, 2010-2011. *Emerg Infect Dis* **19**, doi:10.3201/eid1909.121466 (2013).

87 Diaz, M. H., Benitez, A. J. & Winchell, J. M. Investigations of Mycoplasma pneumoniae infections in the United States: trends in molecular typing and macrolide resistance from 2006 to 2013. *J Clin Microbiol* **53**, 124-130, doi:10.1128/JCM.02597-14 (2015).

88 Zheng, X. *et al.* Macrolide-Resistant Mycoplasma pneumoniae, United States(1). *Emerg Infect Dis* **21**, 1470-1472, doi:10.3201/eid2108.150273 (2015).

89 Jacobs, E. *et al.* Immunological reaction of guinea-pigs following intranasal Mycoplasma pneumoniae infection and immunization with the 168 kDa adherence protein. *J Gen Microbiol* **134**, 473-479, doi:10.1099/00221287-134-2-473 (1988).

90 Szczepanek, S. M. *et al.* Vaccination of BALB/c mice with an avirulent Mycoplasma pneumoniae P30 mutant results in disease exacerbation upon challenge with a virulent strain. *Infect Immun* **80**, 1007-1014, doi:10.1128/IAI.06078-11 (2012).

91 Linchevski, I., Klement, E. & Nir-Paz, R. Mycoplasma pneumoniae vaccine protective efficacy and adverse reactions--Systematic review and meta-analysis. *Vaccine* **27**, 2437-2446, doi:10.1016/j.vaccine.2009.01.135 (2009).

92 Cimolai, N., Mah, D. G., Taylor, G. P. & Morrison, B. J. Bases for the early immune response after rechallenge or component vaccination in an animal model of acute Mycoplasma pneumoniae pneumonitis. *Vaccine* **13**, 305-309 (1995).

93 Schurwanz, N., Jacobs, E. & Dumke, R. Strategy to create chimeric proteins derived from functional adhesin regions of Mycoplasma pneumoniae for vaccine development. *Infect Immun* **77**, 5007-5015, doi:10.1128/IAI.00268-09 (2009).

94 Hausner, M., Schamberger, A., Naumann, W., Jacobs, E. & Dumke, R. Development of protective anti-Mycoplasma pneumoniae antibodies after immunization of guinea pigs with the combination of a P1-P30 chimeric recombinant protein and chitosan. *Microb Pathog* **64**, 23-32, doi:10.1016/j.micpath.2013.07.004 (2013).

95 Zhu, C. *et al.* Protective efficacy of a Mycoplasma pneumoniae P1C DNA vaccine fused with the B subunit of Escherichia coli heat-labile enterotoxin. *Can J Microbiol* **58**, 802-810, doi:10.1139/w2012-051 (2012).

96 Chen, C. *et al.* Designing, Expression and Immunological Characterization of a Chimeric Protein of Mycoplasma pneumoniae. *Protein Pept Lett* **23**, 592-596 (2016).

97 Parrott, G. L., Kinjo, T. & Fujita, J. A Compendium for Mycoplasma pneumoniae. *Front Microbiol* **7**, 513, doi:10.3389/fmicb.2016.00513 (2016).

98 Almagor, M., Kahane, I. & Yatziv, S. Role of superoxide anion in host cell injury induced by mycoplasma pneumoniae infection. A study in normal and trisomy 21 cells. *J Clin Invest* **73**, 842-847, doi:10.1172/JCI111279 (1984).

99 Sun, G. *et al.* Mycoplasma pneumoniae infection induces reactive oxygen species and DNA damage in A549 human lung carcinoma cells. *Infect Immun* **76**, 4405-4413, doi:10.1128/IAI.00575-08 (2008).

100 Grosshennig, S. *et al.* Hydrogen sulfide is a novel potential virulence factor of Mycoplasma pneumoniae: characterization of the unusual cysteine desulfurase/desulfhydrase HapE. *Mol Microbiol* **100**, 42-54, doi:10.1111/mmi.13300 (2016).

101 Kannan, T. R., Provenzano, D., Wright, J. R. & Baseman, J. B. Identification and characterization of human surfactant protein A binding protein of Mycoplasma pneumoniae. *Infect Immun* **73**, 2828-2834, doi:10.1128/IAI.73.5.2828-2834.2005 (2005).

102 Kannan, T. R. & Baseman, J. B. ADP-ribosylating and vacuolating cytotoxin of Mycoplasma pneumoniae represents unique virulence determinant among bacterial pathogens. *Proc Natl Acad Sci U S A* **103**, 6724-6729, doi:10.1073/pnas.0510644103 (2006).

103 Razin, S. Adherence of pathogenic mycoplasmas to host cells. *Biosci Rep* **19**, 367-372 (1999).

104 Kannan, T. R. *et al.* Mycoplasma pneumoniae Community Acquired Respiratory Distress Syndrome toxin expression reveals growth phase and infection-dependent regulation. *Mol Microbiol* **76**, 1127-1141, doi:10.1111/j.1365-2958.2010.07092.x (2010).

105 Clyde, W. A., Jr. Mycoplasma pneumoniae respiratory disease symposium: summation and significance. *Yale J Biol Med* **56**, 523-527 (1983).

106 Mansel, J. K., Rosenow, E. C., Smith, T. F. & Martin, J. W. Mycoplasma pneumoniae pneumonia. *CHEST Journal* **95**, 639-646 (1989).

107 Kannan, T. R. *et al.* Synthesis and distribution of CARDS toxin during Mycoplasma pneumoniae infection in a murine model. *J Infect Dis* **204**, 1596-1604, doi:10.1093/infdis/jir557 (2011).

108 Waites, K. B., Balish, M. F. & Atkinson, T. P. New insights into the pathogenesis and detection of Mycoplasma pneumoniae infections. *Future Microbiol* **3**, 635-648, doi:10.2217/17460913.3.6.635 (2008).

109 Johnson, C., Kannan, T. R. & Baseman, J. B. Cellular vacuoles induced by Mycoplasma pneumoniae CARDS toxin originate from Rab9-associated compartments. *PLoS One* **6**, e22877, doi:10.1371/journal.pone.0022877 (2011).

110 Su, C. J., Chavoya, A. & Baseman, J. B. Regions of Mycoplasma pneumoniae cytadhesin P1 structural gene exist as multiple copies. *Infect Immun* **56**, 3157-3161 (1988).

111 Wenzel, R. & Herrmann, R. Repetitive DNA sequences in Mycoplasma pneumoniae. *Nucleic Acids Res* **16**, 8337-8350 (1988).

112 Ruland, K., Wenzel, R. & Herrmann, R. Analysis of three different repeated DNA elements present in the P1 operon of Mycoplasma pneumoniae: size, number and distribution on the genome. *Nucleic Acids Res* **18**, 6311-6317 (1990).

113 Kenri, T. *et al.* Genotyping analysis of Mycoplasma pneumoniae clinical strains in Japan between 1995 and 2005: type shift phenomenon of M. pneumoniae clinical strains. *J Med Microbiol* **57**, 469-475, doi:10.1099/jmm.0.47634-0 (2008).

114 Simmons, W. L. *et al.* Type 1 and type 2 strains of Mycoplasma pneumoniae form different biofilms. *Microbiology* **159**, 737-747, doi:10.1099/mic.0.064782-0 (2013).

115 Yamazaki, T. & Kenri, T. Epidemiology of Mycoplasma pneumoniae Infections in Japan and Therapeutic Strategies for Macrolide-Resistant M. pneumoniae. *Front Microbiol* **7**, 693, doi:10.3389/fmicb.2016.00693 (2016).

116 Techasaensiri, C. *et al.* Variation in colonization, ADP-ribosylating and vacuolating cytotoxin, and pulmonary disease severity among mycoplasma pneumoniae strains. *Am J Respir Crit Care Med* **182**, 797-804, doi:10.1164/rccm.201001-0080OC (2010).

117 Krause, D. C. & Balish, M. F. Structure, function, and assembly of the terminal organelle of Mycoplasma pneumoniae. *FEMS Microbiol Lett* **198**, 1-7 (2001).

118 Krause, D. C. & Balish, M. F. Cellular engineering in a minimal microbe: structure and assembly of the terminal organelle of Mycoplasma pneumoniae. *Mol Microbiol* **51**, 917-924 (2004).

119 Balish, M. F. Mycoplasma pneumoniae, an underutilized model for bacterial cell biology. *J Bacteriol* **196**, 3675-3682, doi:10.1128/JB.01865-14 (2014).

120 Radestock, U. & Bredt, W. Motility of Mycoplasma pneumoniae. *J Bacteriol* **129**, 1495-1501 (1977).

121 Miyata, M. & Hamaguchi, T. Integrated Information and Prospects for Gliding Mechanism of the Pathogenic Bacterium Mycoplasma pneumoniae. *Front Microbiol* **7**, 960, doi:10.3389/fmicb.2016.00960 (2016).

122 Chilvers, M. A. & O'Callaghan, C. Analysis of ciliary beat pattern and beat frequency using digital high speed imaging: comparison with the photomultiplier and photodiode methods. *Thorax* **55**, 314-317 (2000).

123 Krunkosky, T. M., Jordan, J. L., Chambers, E. & Krause, D. C. Mycoplasma pneumoniae host-pathogen studies in an air-liquid culture of differentiated human airway epithelial cells. *Microb Pathog* **42**, 98-103, doi:10.1016/j.micpath.2006.11.003 (2007).

124 Prince, O. A., Krunkosky, T. M. & Krause, D. C. In vitro spatial and temporal analysis of Mycoplasma pneumoniae colonization of human airway epithelium. *Infect Immun* **82**, 579-586, doi:10.1128/IAI.01036-13 (2014).

125 Seybert, A., Herrmann, R. & Frangakis, A. S. Structural analysis of Mycoplasma pneumoniae by cryo-electron tomography. *J Struct Biol* **156**, 342-354, doi:10.1016/j.jsb.2006.04.010 (2006).

126 Henderson, G. P. & Jensen, G. J. Three-dimensional structure of Mycoplasma pneumoniae's attachment organelle and a model for its role in gliding motility. *Mol Microbiol* **60**, 376-385, doi:10.1111/j.1365-2958.2006.05113.x (2006).

127 Miyata, M. Centipede and inchworm models to explain Mycoplasma gliding. *Trends Microbiol* **16**, 6-12, doi:10.1016/j.tim.2007.11.002 (2008).

128 Kawamoto, A. *et al.* Periodicity in Attachment Organelle Revealed by Electron Cryotomography Suggests Conformational Changes in Gliding Mechanism of Mycoplasma pneumoniae. *MBio* **7**, e00243-00216, doi:10.1128/mBio.00243-16 (2016).

129 Meng, K. E. & Pfister, R. M. Intracellular structures of Mycoplasma pneumoniae revealed after membrane removal. *J Bacteriol* **144**, 390-399 (1980).

130 Gobel, U., Speth, V. & Bredt, W. Filamentous structures in adherent Mycoplasma pneumoniae cells treated with nonionic detergents. *J Cell Biol* **91**, 537-543 (1981).

131 Collier, A. M., Hu, P. C. & Clyde, W. A., Jr. Location of attachment moiety on Mycoplasma pneumoniae. *Yale J Biol Med* **56**, 671-677 (1983).

132 Nakane, D., Kenri, T., Matsuo, L. & Miyata, M. Systematic Structural Analyses of Attachment Organelle in Mycoplasma pneumoniae. *PLoS Pathog* **11**, e1005299, doi:10.1371/journal.ppat.1005299 (2015).

133 Neimark, H. C. Extraction of an actin-like protein from the prokaryote Mycoplasma pneumoniae. *Proc Natl Acad Sci U S A* **74**, 4041-4045 (1977).

134 Miyata, M. & Nakane, D. (eds G. Browning & C. Citti) 237 - 252 (Norfolk: Horizon Press, 2013).

135 Balish, M. F., Hahn, T. W., Popham, P. L. & Krause, D. C. Stability of Mycoplasma pneumoniae cytadherence-accessory protein HMW1 correlates with its association with the triton shell. *J Bacteriol* **183**, 3680-3688, doi:10.1128/JB.183.12.3680-3688.2001 (2001).

136 Krause, D. C., Leith, D. K. & Baseman, J. B. Reacquisition of specific proteins confers virulence in Mycoplasma pneumoniae. *Infect Immun* **39**, 830-836 (1983).

137 Hasselbring, B. M. & Krause, D. C. Cytoskeletal protein P41 is required to anchor the terminal organelle of the wall-less prokaryote Mycoplasma pneumoniae. *Mol Microbiol* **63**, 44-53, doi:10.1111/j.1365-2958.2006.05507.x (2007).

138 Hasselbring, B. M. & Krause, D. C. Proteins P24 and P41 function in the regulation of terminal-organelle development and gliding motility in Mycoplasma pneumoniae. *J Bacteriol* **189**, 7442-7449, doi:10.1128/JB.00867-07 (2007).

139 Jordan, J. L., Berry, K. M., Balish, M. F. & Krause, D. C. Stability and subcellular localization of cytadherence-associated protein P65 in Mycoplasma pneumoniae. *J Bacteriol* **183**, 7387-7391, doi:10.1128/JB.183.24.7387-7891.2001 (2001).

140 Jordan, J. L. *et al.* Protein P200 is dispensable for Mycoplasma pneumoniae hemadsorption but not gliding motility or colonization of differentiated bronchial epithelium. *Infect Immun* **75**, 518-522, doi:10.1128/IAI.01344-06 (2007).

141 Cloward, J. M. & Krause, D. C. Mycoplasma pneumoniae J-domain protein required for terminal organelle function. *Mol Microbiol* **71**, 1296-1307, doi:10.1111/j.1365-2958.2009.06602.x (2009).

142 Hasselbring, B. M., Page, C. A., Sheppard, E. S. & Krause, D. C. Transposon mutagenesis identifies genes associated with Mycoplasma pneumoniae gliding motility. *J Bacteriol* **188**, 6335-6345, doi:10.1128/JB.00698-06 (2006).

143 Kawakita, Y. *et al.* Structural Study of MPN387, an Essential Protein for Gliding Motility of a Human-Pathogenic Bacterium, Mycoplasma pneumoniae. *J Bacteriol* **198**, 2352-2359, doi:10.1128/JB.00160-16 (2016).

144 Loomes, L. M., Uemura, K. & Feizi, T. Interaction of Mycoplasma pneumoniae with erythrocyte glycolipids of I and i antigen types. *Infect Immun* **47**, 15-20 (1985).

145 Loomes, L. M. *et al.* Erythrocyte receptors for Mycoplasma pneumoniae are sialylated oligosaccharides of Ii antigen type. *Nature* **307**, 560-563 (1984).

146 Roberts, D. D., Olson, L. D., Barile, M. F., Ginsburg, V. & Krivan, H. C. Sialic acid-dependent adhesion of Mycoplasma pneumoniae to purified glycoproteins. *J Biol Chem* **264**, 9289-9293 (1989).

147 Loveless, R. W. & Feizi, T. Sialo-oligosaccharide receptors for Mycoplasma pneumoniae and related oligosaccharides of poly-N-acetyllactosamine series are polarized at the cilia and apical-microvillar domains of the ciliated cells in human bronchial epithelium. *Infect Immun* **57**, 1285-1289 (1989).

148 Singer, G. A. & Hickey, D. A. Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Mol Biol Evol* **17**, 1581-1588 (2000).

149 Noel, G. J., Love, D. C. & Mosser, D. M. High-molecular-weight proteins of nontypeable Haemophilus influenzae mediate bacterial adhesion to cellular proteoglycans. *Infect Immun* **62**, 4028-4033 (1994).

150 Zhang, Q., Young, T. F. & Ross, R. F. Microtiter plate adherence assay and receptor analogs for Mycoplasma hyopneumoniae. *Infect Immun* **62**, 1616-1622 (1994).

151 Hu, P. C., Collier, A. M. & Baseman, J. B. Surface parasitism by Mycoplasma pneumoniae of respiratory epithelium. *J Exp Med* **145**, 1328-1343 (1977).

152 Krause, D. C., Leith, D. K., Wilson, R. M. & Baseman, J. B. Identification of Mycoplasma pneumoniae proteins associated with hemadsorption and virulence. *Infect Immun* **35**, 809-817 (1982).

153 Leith, D. K., Trevino, L. B., Tully, J. G., Senterfit, L. B. & Baseman, J. B. Host discrimination of Mycoplasma pneumoniae proteinaceous immunogens. *J Exp Med* **157**, 502-514 (1983).

154 Baseman, J. B. *et al.* Identification of a 32-kilodalton protein of Mycoplasma pneumoniae associated with hemadsorption. *Isr J Med Sci* **23**, 474-479 (1987).

155 Svenstrup, H. F., Nielsen, P. K., Drasbek, M., Birkelund, S. & Christiansen, G. Adhesion and inhibition assay of Mycoplasma genitalium and M. pneumoniae by immunofluorescence microscopy. *J Med Microbiol* **51**, 361-373, doi:10.1099/0022-1317-51-5-361 (2002).

156 Hu, P. C. *et al.* Mycoplasma pneumoniae infection: role of a surface protein in the attachment organelle. *Science* **216**, 313-315 (1982).

157 Kahane, I., Tucker, S., Leith, D. K., Morrison-Plummer, J. & Baseman, J. B. Detection of the major adhesin P1 in triton shells of virulent Mycoplasma pneumoniae. *Infect Immun* **50**, 944-946 (1985).

158 Morrison-Plummer, J., Leith, D. K. & Baseman, J. B. Biological effects of anti-lipid and anti-protein monoclonal antibodies on Mycoplasma pneumoniae. *Infect Immun* **53**, 398-403 (1986).

159 Razin, S. & Jacobs, E. Mycoplasma adhesion. *J Gen Microbiol* **138**, 407-422, doi:10.1099/00221287-138-3-407 (1992).

160 Kornspan, J. D., Tarshis, M. & Rottem, S. Adhesion and biofilm formation of Mycoplasma pneumoniae on an abiotic surface. *Arch Microbiol* **193**, 833-836, doi:10.1007/s00203-011-0749-y (2011).

161 Feldner, J., Gobel, U. & Bredt, W. Mycoplasma pneumoniae adhesin localized to tip structure by monoclonal antibody. *Nature* **298**, 765-767 (1982).

162 Seto, S., Kenri, T., Tomiyama, T. & Miyata, M. Involvement of P1 adhesin in gliding motility of Mycoplasma pneumoniae as revealed by the inhibitory effects of antibody under optimized gliding conditions. *J Bacteriol* **187**, 1875-1877, doi:10.1128/JB.187.5.1875-1877.2005 (2005).

163 Inamine, J. M. *et al.* Nucleotide sequence of the P1 attachment-protein gene of Mycoplasma pneumoniae. *Gene* **64**, 217-229 (1988).

164 Romero-Arroyo, C. E. *et al.* Mycoplasma pneumoniae protein P30 is required for cytadherence and associated with proper cell development. *J Bacteriol* **181**, 1079-1087 (1999).

165 Krause, D. C. Mycoplasma pneumoniae cytadherence: unravelling the tie that binds. *Mol Microbiol* **20**, 247-253 (1996).

166 Sperker, B., Hu, P. & Herrmann, R. Identification of gene products of the P1 operon of Mycoplasma pneumoniae. *Mol Microbiol* **5**, 299-306 (1991).

167 Layh-Schmitt, G. & Herrmann, R. Localization and biochemical characterization of the ORF6 gene product of the Mycoplasma pneumoniae P1 operon. *Infect Immun* **60**, 2906-2913 (1992).

168 Catrein, I., Herrmann, R., Bosserhoff, A. & Ruppert, T. Experimental proof for a signal peptidase I like activity in Mycoplasma pneumoniae, but absence of a gene encoding a conserved bacterial type I SPase. *FEBS J* **272**, 2892-2900, doi:10.1111/j.1742-4658.2005.04710.x (2005).

169 Layh-Schmitt, G. & Herrmann, R. Spatial arrangement of gene products of the P1 operon in the membrane of Mycoplasma pneumoniae. *Infect Immun* **62**, 974-979 (1994).

170 Waldo, R. H., 3rd & Krause, D. C. Synthesis, stability, and function of cytadhesin P1 and accessory protein B/C complex of Mycoplasma pneumoniae. *J Bacteriol* **188**, 569-575, doi:10.1128/JB.188.2.569-575.2006 (2006).

171 Willby, M. J. *et al.* HMW1 is required for stability and localization of HMW2 to the attachment organelle of Mycoplasma pneumoniae. *J Bacteriol* **186**, 8221-8228, doi:10.1128/JB.186.24.8221-8228.2004 (2004).

172 Seto, S. & Miyata, M. Attachment organelle formation represented by localization of cytadherence proteins and formation of the electron-dense core in wild-type and mutant strains of Mycoplasma pneumoniae. *J Bacteriol* **185**, 1082-1091 (2003).

173 Willby, M. J. & Krause, D. C. Characterization of a Mycoplasma pneumoniae hmw3 mutant: implications for attachment organelle assembly. *J Bacteriol* **184**, 3061-3068 (2002).

174 Cloward, J. M. & Krause, D. C. Loss of co-chaperone TopJ impacts adhesin P1 presentation and terminal organelle maturation in Mycoplasma pneumoniae. *Mol Microbiol* **81**, 528-539, doi:10.1111/j.1365-2958.2011.07712.x (2011).

175 Monod, J. From enzymatic adaptation to allosteric transitions. *Science* **154**, 475-483 (1966).

176 Becher, D. *et al.* A proteomic view of an important human pathogen--towards the quantification of the entire Staphylococcus aureus proteome. *PLoS One* **4**, e8176, doi:10.1371/journal.pone.0008176 (2009).

177 McNamara, M., Tzeng, S. C., Maier, C., Zhang, L. & Bermudez, L. E. Surface proteome of "Mycobacterium avium subsp. hominissuis" during the early stages of macrophage infection. *Infect Immun* **80**, 1868-1880, doi:10.1128/IAI.06151-11 (2012).

178 Olaya-Abril, A., Gomez-Gascon, L., Jimenez-Munguia, I., Obando, I. & Rodriguez-Ortega, M. J. Another turn of the screw in shaving Gram-positive bacteria: Optimization of proteomics surface protein identification in Streptococcus pneumoniae. *J Proteomics* **75**, 3733-3746, doi:10.1016/j.jprot.2012.04.037 (2012).

179 Solis, N., Larsen, M. R. & Cordwell, S. J. Improved accuracy of cell surface shaving proteomics in Staphylococcus aureus using a false-positive control. *Proteomics* **10**, 2037-2049, doi:10.1002/pmic.200900564 (2010).

180 Cincarova, L., Polansky, O., Babak, V., Kulich, P. & Kralik, P. Changes in the Expression of Biofilm-Associated Surface Proteins in Staphylococcus aureus Food-Environmental Isolates Subjected to Sublethal Concentrations of Disinfectants. *Biomed Res Int* **2016**, 4034517, doi:10.1155/2016/4034517 (2016).

181 Walters, M. S. & Mobley, H. L. Identification of uropathogenic Escherichia coli surface proteins by shotgun proteomics. *J Microbiol Methods* **78**, 131-135, doi:10.1016/j.mimet.2009.04.013 (2009).

182 Sears, K. T. *et al.* Surface proteome analysis and characterization of surface cell antigen (Sca) or autotransporter family of Rickettsia typhi. *PLoS Pathog* **8**, e1002856, doi:10.1371/journal.ppat.1002856 (2012).

183 Voss, B. J. & Cover, T. L. Biotinylation and Purification of Surface-exposed Helicobacter pylori Proteins. *Bio Protoc* **5**, e1455 (2015).

184 Zhu, W. & Lee, S. W. Surface interactions between two of the main periodontal pathogens: Porphyromonas gingivalis and Tannerella forsythia. *J Periodontal Implant Sci* **46**, 2-9, doi:10.5051/jpis.2016.46.1.2 (2016).

185 Wu, H. N. & Miyata, M. Whole surface image of Mycoplasma mobile, suggested by protein identification and immunofluorescence microscopy. *J Bacteriol* **194**, 5848-5855, doi:10.1128/JB.00976-12 (2012).

186 Bogema, D. R. *et al.* Sequence TTKF downward arrow 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, doi:10.1074/jbc.M111.226084 (2011).

187 Deutscher, A. T. *et al.* Repeat regions R1 and R2 in the P97 paralogue Mhp271 of Mycoplasma hyopneumoniae bind heparin, fibronectin and porcine cilia. *Mol Microbiol* **78**, 444-458, doi:10.1111/j.1365-2958.2010.07345.x (2010).

188 Henderson, B. & Martin, A. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect Immun* **79**, 3476-3491, doi:10.1128/IAI.00179-11 (2011).

189 Jeffery, C. J. Moonlighting proteins. *Trends Biochem Sci* **24**, 8-11 (1999).

190 Jeffery, C. J. Moonlighting proteins--an update. *Mol Biosyst* **5**, 345-350, doi:10.1039/b900658n (2009).

191 Kaparakis-Liaskos, M. & Ferrero, R. L. Immune modulation by bacterial outer membrane vesicles. *Nat Rev Immunol* **15**, 375-387, doi:10.1038/nri3837 (2015).

192 Schwechheimer, C. & Kuehn, M. J. Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nat Rev Microbiol* **13**, 605-619, doi:10.1038/nrmicro3525 (2015).
193 Turnbull, L. *et al.* Explosive cell lysis as a mechanism for the biogenesis of bacterial membrane vesicles and biofilms. *Nat Commun* **7**, 11220, doi:10.1038/ncomms11220 (2016).
194 Wang, G. *et al.* The Roles of Moonlighting Proteins in Bacteria. *Curr Issues Mol Biol* **16**, 15-22 (2014).
195 Wang, W. & Jeffery, C. J. An analysis of surface proteomics results reveals novel candidates for intracellular/surface moonlighting proteins in bacteria. *Mol Biosyst* **12**, 1420-1431, doi:10.1039/c5mb00550g (2016).
196 Mani, M. *et al.* MoonProt: a database for proteins that are known to moonlight. *Nucleic Acids Res* **43**, D277-282, doi:10.1093/nar/gku954 (2015).
197 Sha, J. *et al.* Surface-expressed enolase contributes to the pathogenesis of clinical isolate SSU of Aeromonas hydrophila. *J Bacteriol* **191**, 3095-3107, doi:10.1128/JB.00005-09 (2009).
198 Dinis, M. *et al.* Oral therapeutic vaccination with Streptococcus sobrinus recombinant enolase confers protection against dental caries in rats. *J Infect Dis* **199**, 116-123, doi:10.1086/594372 (2009).
199 Dutta, S., DasSarma, P., DasSarma, S. & Jarori, G. K. Immunogenicity and protective potential of a Plasmodium spp. enolase peptide displayed on archaeal gas vesicle nanoparticles. *Malar J* **14**, 406, doi:10.1186/s12936-015-0914-x (2015).
200 Perez-Casal, J. & Potter, A. A. Glyceradehyde-3-phosphate dehydrogenase as a suitable vaccine candidate for protection against bacterial and parasitic diseases. *Vaccine* **34**, 1012-1017, doi:10.1016/j.vaccine.2015.11.072 (2016).
201 Daniely, D. *et al.* Pneumococcal 6-phosphogluconate-dehydrogenase, a putative adhesin, induces protective immune response in mice. *Clin Exp Immunol* **144**, 254-263, doi:10.1111/j.1365-2249.2006.03047.x (2006).
202 Tan, C. *et al.* Vaccination with Streptococcus suis serotype 2 recombinant 6PGD protein provides protection against S. suis infection in swine. *FEMS Microbiol Lett* **296**, 78-83, doi:10.1111/j.1574-6968.2009.01617.x (2009).
203 Collado, R., Prenafeta, A., Gonzalez-Gonzalez, L., Perez-Pons, J. A. & Sitja, M. Probing vaccine antigens against bovine mastitis caused by Streptococcus uberis. *Vaccine* **34**, 3848-3854, doi:10.1016/j.vaccine.2016.05.044 (2016).
204 Argiro, L. *et al.* Induction of a protection against S. mansoni with a MAP containing epitopes of Sm37-GAPDH and Sm10-DLC. Effect of coadsorption with GM-CSF on alum. *Vaccine* **18**, 2033-2038 (2000).
205 Dallo, S. F., Kannan, T. R., Blaylock, M. W. & Baseman, J. B. Elongation factor Tu and E1 beta subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in Mycoplasma pneumoniae. *Mol Microbiol* **46**, 1041-1051 (2002).
206 Daubenspeck, J. M., Liu, R. & Dybvig, K. Rhamnose Links Moonlighting Proteins to Membrane Phospholipid in Mycoplasmas. *PLoS One* **11**, e0162505, doi:10.1371/journal.pone.0162505 (2016).
207 Jordan, D. S., Daubenspeck, J. M. & Dybvig, K. Rhamnose biosynthesis in mycoplasmas requires precursor glycans larger than monosaccharide. *Mol Microbiol* **89**, 918-928, doi:10.1111/mmi.12320 (2013).

208 Dumke, R., Hausner, M. & Jacobs, E. Role of Mycoplasma pneumoniae glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in mediating interactions with the human extracellular matrix. *Microbiology* **157**, 2328-2338, doi:10.1099/mic.0.048298-0 (2011).

209 Gründel, A., Pfeiffer, M., Jacobs, E. & Dumke, R. Network of Surface-Displayed Glycolytic Enzymes in Mycoplasma pneumoniae and Their Interactions with Human Plasminogen. *Infect Immun* **84**, 666-676, doi:10.1128/IAI.01071-15 (2016).

210 Gründel, A., Jacobs, E. & Dumke, R. Interactions of surface-displayed glycolytic enzymes of Mycoplasma pneumoniae with components of the human extracellular matrix. *Int J Med Microbiol*, doi:10.1016/j.ijmm.2016.09.001 (2016).

211 Lahteenmaki, K., Edelman, S. & Korhonen, T. K. Bacterial metastasis: the host plasminogen system in bacterial invasion. *Trends Microbiol* **13**, 79-85, doi:10.1016/j.tim.2004.12.003 (2005).

212 Raymond, B. B. & Djordjevic, S. Exploitation of plasmin(ogen) by bacterial pathogens of veterinary significance. *Vet Microbiol* **178**, 1-13, doi:10.1016/j.vetmic.2015.04.008 (2015).

213 Hagemann, L., Gründel, A., Jacobs, E. & Dumke, R. The surface-displayed chaperones GroEL and DnaK of Mycoplasma pneumoniae interact with human plasminogen and components of the extracellular matrix. *Pathog Dis*, doi:10.1093/femspd/ftx017 (2017).

214 Thomas, C., Jacobs, E. & Dumke, R. Characterization of pyruvate dehydrogenase subunit B and enolase as plasminogen-binding proteins in Mycoplasma pneumoniae. *Microbiology* **159**, 352-365, doi:10.1099/mic.0.061184-0 (2013).

215 Gründel, A., Friedrich, K., Pfeiffer, M., Jacobs, E. & Dumke, R. Subunits of the Pyruvate Dehydrogenase Cluster of Mycoplasma pneumoniae Are Surface-Displayed Proteins that Bind and Activate Human Plasminogen. *PLoS One* **10**, e0126600, doi:10.1371/journal.pone.0126600 (2015).

216 Djordjevic, S. P., Cordwell, S. J., Djordjevic, M. A., Wilton, J. & Minion, F. C. Proteolytic processing of the Mycoplasma hyopneumoniae cilium adhesin. *Infect Immun* **72**, 2791-2802 (2004).

217 Burnett, T. A. *et al.* 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, doi:10.1111/j.1365-2958.2006.05139.x (2006).

218 Wilton, J. *et al.* Mhp493 (P216) is a proteolytically processed, cilium and heparin binding protein of Mycoplasma hyopneumoniae. *Mol Microbiol* **71**, 566-582, doi:10.1111/j.1365-2958.2008.06546.x (2009).

219 Seymour, L. M. *et al.* A processed multidomain mycoplasma hyopneumoniae adhesin binds fibronectin, plasminogen, and swine respiratory cilia. *J Biol Chem* **285**, 33971-33978, doi:10.1074/jbc.M110.104463 (2010).

220 Seymour, L. M. *et al.* Mhp107 is a member of the multifunctional adhesin family of Mycoplasma hyopneumoniae. *J Biol Chem* **286**, 10097-10104, doi:10.1074/jbc.M110.208140 (2011).

221 Bogema, D. R. *et al.* Characterization of cleavage events in the multifunctional cilium adhesin Mhp684 (P146) reveals a mechanism by which Mycoplasma hyopneumoniae regulates surface topography. *MBio* **3**, doi:10.1128/mBio.00282-11 (2012).

222 Deutscher, A. T. *et al.* 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, doi:10.1021/pr201115v (2012).

223 Seymour, L. M. *et al.* Mhp182 (P102) binds fibronectin and contributes to the recruitment of plasmin(ogen) to the Mycoplasma hyopneumoniae cell surface. *Cell Microbiol* **14**, 81-94, doi:10.1111/j.1462-5822.2011.01702.x (2012).

224 Raymond, B. B. *et al.* 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, doi:10.1021/pr400903s (2013).

225 Tacchi, J. L. *et al.* 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, doi:10.1021/pr500087c (2014).

226 Raymond, B. B. *et al.* Proteolytic processing of the cilium adhesin MHJ_0194 (P123J) in Mycoplasma hyopneumoniae generates a functionally diverse array of cleavage fragments that bind multiple host molecules. *Cell Microbiol* **17**, 425-444, doi:10.1111/cmi.12377 (2015).

227 Tacchi, J. L. *et al.* Post-translational processing targets functionally diverse proteins in Mycoplasma hyopneumoniae. *Open Biol* **6**, 150210, doi:10.1098/rsob.150210 (2016).

228 Shimizu, T., Kida, Y. & Kuwano, K. Triacylated lipoproteins derived from Mycoplasma pneumoniae activate nuclear factor-kappaB through toll-like receptors 1 and 2. *Immunology* **121**, 473-483, doi:10.1111/j.1365-2567.2007.02594.x (2007).

229 Shimizu, T., Kida, Y. & Kuwano, K. Mycoplasma pneumoniae-derived lipopeptides induce acute inflammatory responses in the lungs of mice. *Infect Immun* **76**, 270-277, doi:10.1128/IAI.00955-07 (2008).

230 Widjaja, M., Berry, I. J., Pont, E. J., Padula, M. P. & Djordjevic, S. P. P40 and P90 from Mpn142 are Targets of Multiple Processing Events on the Surface of Mycoplasma pneumoniae. *Proteomes* **3**, 512-537, doi:10.3390/proteomes3040512 (2015).

231 Szczepanek, S. M. *et al.* Identification of lipoprotein MslA as a neoteric virulence factor of Mycoplasma gallisepticum. *Infect Immun* **78**, 3475-3483, doi:10.1128/IAI.00154-10 (2010).

232 Calcutt, M. J., Kim, M. F., Karpas, A. B., Muhlradt, P. F. & Wise, K. S. Differential posttranslational processing confers intraspecies variation of a major surface lipoprotein and a macrophage-activating lipopeptide of Mycoplasma fermentans. *Infect Immun* **67**, 760-771 (1999).

233 Davis, K. L. & Wise, K. S. 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 (2002).

234 Muhlradt, P. F., Kiess, M., Meyer, H., Sussmuth, R. & Jung, G. Isolation, structure elucidation, and synthesis of a macrophage stimulatory lipopeptide from Mycoplasma fermentans acting at picomolar concentration. *J Exp Med* **185**, 1951-1958 (1997).

235 Shimizu, T., Kida, Y. & Kuwano, K. A triacylated lipoprotein from Mycoplasma genitalium activates NF-kappaB through Toll-like receptor 1 (TLR1) and TLR2. *Infect Immun* **76**, 3672-3678, doi:10.1128/IAI.00257-08 (2008).

236 Dubrana, M. P. *et al.* Proteolytic post-translational processing of adhesins in the plant pathogen Spiroplasma citri. *J Mol Biol*, doi:10.1016/j.jmb.2017.05.004 (2017).

237 Suh, M. J. *et al.* Using chemical derivatization and mass spectrometric analysis to characterize the post-translationally modified Staphylococcus aureus surface protein G. *Biochim Biophys Acta* **1804**, 1394-1404, doi:10.1016/j.bbapap.2010.02.006 (2010).

238 Wall, E. A. *et al.* Specific N-terminal cleavage of ribosomal protein L27 in Staphylococcus aureus and related bacteria. *Mol Microbiol* **95**, 258-269, doi:10.1111/mmi.12862 (2015).

239 Scott, N. E. *et al.* 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, doi:10.1002/pmic.200900440 (2010).

240 Veith, P. D. *et al.* Protein substrates of a novel secretion system are numerous in the Bacteroidetes phylum and have in common a cleavable C-terminal secretion signal, extensive post-translational modification, and cell-surface attachment. *J Proteome Res* **12**, 4449-4461, doi:10.1021/pr400487b (2013).

241 Noriea, N. F., Clark, T. R., Mead, D. & Hackstadt, T. Proteolytic Cleavage of the Immunodominant Outer Membrane Protein rOmpA in Rickettsia rickettsii. *J Bacteriol*, doi:10.1128/JB.00826-16 (2016).

242 Georgiou, T. *et al.* Specific peptide-activated proteolytic cleavage of Escherichia coli elongation factor Tu. *Proc Natl Acad Sci U S A* **95**, 2891-2895 (1998).

243 Robinson, M. W. *et al.* MHJ_0125 is an M42 glutamyl aminopeptidase that moonlights as a multifunctional adhesin on the surface of Mycoplasma hyopneumoniae. *Open Biol* **3**, 130017, doi:10.1098/rsob.130017 (2013).

244 Jarocki, V. M. *et al.* MHJ_0461 is a multifunctional leucine aminopeptidase on the surface of Mycoplasma hyopneumoniae. *Open Biol* **5**, 140175, doi:10.1098/rsob.140175 (2015).

245 Su, C. J., Tryon, V. V. & Baseman, J. B. Cloning and sequence analysis of cytadhesin P1 gene from Mycoplasma pneumoniae. *Infect Immun* **55**, 3023-3029 (1987).

246 Jacobs, E., Fuchte, K. & Bredt, W. Amino acid sequence and antigenicity of the amino-terminus of the 168 kDa adherence protein of Mycoplasma pneumoniae. *J Gen Microbiol* **133**, 2233-2236, doi:10.1099/00221287-133-8-2233 (1987).

247 Layh-Schmitt, G., Podtelejnikov, A. & Mann, M. Proteins complexed to the P1 adhesin of Mycoplasma pneumoniae. *Microbiology* **146 (Pt 3)**, 741-747 (2000).

248 Regula, J. T. *et al.* Towards a two-dimensional proteome map of Mycoplasma pneumoniae. *Electrophoresis* **21**, 3765-3780, doi:10.1002/1522-2683(200011)21:17<3765::AID-ELPS3765>3.0.CO;2-6 (2000).

249 Baseman, J. B. The cytadhesins of Mycoplasma pneumoniae and M. genitalium. *Subcell Biochem* **20**, 243-259 (1993).

250 Hayflick, L. Tissue cultures and mycoplasmas. *Tex Rep Biol Med* **23**, Suppl 1:285+ (1965).

251 Webb, A. *Systems Biology Mascot Server: Databases (MSPnr100)*, <http://www.wehi.edu.au/people/andrew-webb/1295/andrew-webb-resources> (2015).

252 Oliveros, J. C. *Venny. An interactive tool for comparing lists with Venn's diagrams.*, <http://bioinfogp.cnb.csic.es/tools/venny/index.html> (2015).

253 Scientific, T. F. *EZ-Link™ Sulfo-NHS-Biotin*, <https://www.thermofisher.com/order/catalog/product/21217> (2017).

254 Sigma-Aldrich. *Trypsin from porcine pancreas*, <http://www.sigmaaldrich.com/catalog/product/sigma/t4799> (2017).

255 Tjalsma, H., Lambooy, L., Hermans, P. W. & Swinkels, D. W. Shedding & shaving: disclosure of proteomic expressions on a bacterial face. *Proteomics* **8**, 1415-1428, doi:10.1002/pmic.200700550 (2008).

256 Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785-786, doi:10.1038/nmeth.1701 (2011).

257 Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G. & Brunak, S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* **17**, 349-356, doi:10.1093/protein/gzh037 (2004).

258 Hofmann K, S. W. TMbase - A database of membrane spanning proteins segments. *Biol. Chem.* **374**, 166 (1993).

259 Apweiler, R. *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**, D115-119, doi:10.1093/nar/gkh131 (2004).

260 Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608-1615, doi:10.1093/bioinformatics/btq249 (2010).

261 Flemming, H. C. & Wingender, J. The biofilm matrix. *Nat Rev Microbiol* **8**, 623-633, doi:10.1038/nrmicro2415 (2010).

262 Goodman, S. D. *et al.* Biofilms can be dispersed by focusing the immune system on a common family of bacterial nucleoid-associated proteins. *Mucosal Immunol* **4**, 625-637, doi:10.1038/mi.2011.27 (2011).

263 Henderson, B. & Martin, A. Bacterial moonlighting proteins and bacterial virulence. *Curr Top Microbiol Immunol* **358**, 155-213, doi:10.1007/82_2011_188 (2013).

264 Okshevsky, M. & Meyer, R. L. The role of extracellular DNA in the establishment, maintenance and perpetuation of bacterial biofilms. *Crit Rev Microbiol* **41**, 341-352, doi:10.3109/1040841X.2013.841639 (2015).

265 Dallo, S. F., Lazzell, A. L., Chavoya, A., Reddy, S. P. & Baseman, J. B. Biofunctional domains of the Mycoplasma pneumoniae P30 adhesin. *Infect Immun* **64**, 2595-2601 (1996).

266 Layh-Schmitt, G., Himmelreich, R. & Leibfried, U. The adhesin related 30-kDa protein of Mycoplasma pneumoniae exhibits size and antigen variability. *FEMS Microbiol Lett* **152**, 101-108 (1997).

267 Corporation, P. *Trypsin Gold, Mass Spectrometry Grade*, <https://www.promega.com.au/-/media/files/resources/protocols/technical-bulletins/101/trypsin-gold-mass-spectrometry-grade-protocol.pdf?la=en> (2015).

268 Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**, 531-552 (1999).

269 Hegermann, J. *et al.* The acidic, glutamine-rich Mpn474 protein of Mycoplasma pneumoniae is surface exposed and covers the complete cell. *Microbiology* **154**, 1185-1192, doi:10.1099/mic.0.2007/013342-0 (2008).

270 Proft, T. & Herrmann, R. Identification and characterization of hitherto unknown Mycoplasma pneumoniae proteins. *Mol Microbiol* **13**, 337-348 (1994).

271 Kahane, I., Banai, M., Razin, S. & Feldner, J. Attachment of mycoplasmas to host cell membranes. *Rev Infect Dis* **4 Suppl**, S185-192 (1982).

272 Jenkins, C., Geary, S. J., Gladd, M. & Djordjevic, S. P. The Mycoplasma gallisepticum OsmC-like protein MG1142 resides on the cell surface and binds heparin. *Microbiology* **153**, 1455-1463, doi:10.1099/mic.0.2006/004937-0 (2007).

273 Sokoli, A. *et al.* Mycoplasma suis infection results endothelial cell damage and activation: new insight into the cell tropism and pathogenicity of hemotrophic mycoplasma. *Vet Res* **44**, 6, doi:10.1186/1297-9716-44-6 (2013).

274 Zhang, Y. *et al.* Identification of Mycoplasma suis MSG1 interaction proteins on porcine erythrocytes. *Arch Microbiol* **197**, 277-283, doi:10.1007/s00203-014-1050-7 (2015).

275 Healthcare, G. *HiTrap™ Heparin HP*, <https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1314716762536/litdoc18113477_20161012164805.pdf> (2014).

276 Cardin, A. D. & Weintraub, H. J. Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **9**, 21-32 (1989).

277 Sigma-Aldrich. *Actin From Bovine Muscle*, <http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/a3653pis.pdf> (2017).

278 Baseman, J. B., Banai, M. & Kahane, I. Sialic acid residues mediate Mycoplasma pneumoniae attachment to human and sheep erythrocytes. *Infect Immun* **38**, 389-391 (1982).

279 Das, D. *et al.* Crystal structure of a novel single-stranded DNA binding protein from Mycoplasma pneumoniae. *Proteins* **67**, 776-782, doi:10.1002/prot.21340 (2007).

280 May, M., Papazisi, L., Gorton, T. S. & Geary, S. J. Identification of fibronectin-binding proteins in Mycoplasma gallisepticum strain R. *Infect Immun* **74**, 1777-1785, doi:10.1128/IAI.74.3.1777-1785.2006 (2006).

281 Kamhi, E., Joo, E. J., Dordick, J. S. & Linhardt, R. J. Glycosaminoglycans in infectious disease. *Biol Rev Camb Philos Soc* **88**, 928-943, doi:10.1111/brv.12034 (2013).

282 Chen, Y., Gotte, M., Liu, J. & Park, P. W. Microbial subversion of heparan sulfate proteoglycans. *Mol Cells* **26**, 415-426 (2008).

283 Shanks, R. M. *et al.* Heparin stimulates Staphylococcus aureus biofilm formation. *Infect Immun* **73**, 4596-4606, doi:10.1128/IAI.73.8.4596-4606.2005 (2005).

284 Chen, X. e., Ling, P., Duan, R. & Zhang, T. Effects of heparosan and heparin on the adhesion and biofilm formation of several bacteria in vitro. *Carbohydrate polymers* **88**, 1288-1292 (2012).

285 Madureira, P. *et al.* Streptococcus agalactiae GAPDH is a virulence-associated immunomodulatory protein. *J Immunol* **178**, 1379-1387 (2007).

286 Bergonzelli, G. E. *et al.* GroEL of Lactobacillus johnsonii La1 (NCC 533) is cell surface associated: potential role in interactions with the host and the gastric pathogen Helicobacter pylori. *Infect Immun* **74**, 425-434, doi:10.1128/IAI.74.1.425-434.2006 (2006).

287 Gozalbo, D. *et al.* The cell wall-associated glyceraldehyde-3-phosphate dehydrogenase of Candida albicans is also a fibronectin and laminin binding protein. *Infect Immun* **66**, 2052-2059 (1998).

288 Hurmalainen, V. *et al.* Extracellular proteins of Lactobacillus crispatus enhance activation of human plasminogen. *Microbiology* **153**, 1112-1122, doi:10.1099/mic.0.2006/000901-0 (2007).

289 Pancholi, V. & Fischetti, V. A. A major surface protein on group A streptococci is a glyceraldehyde-3-phosphate-dehydrogenase with multiple binding activity. *J Exp Med* **176**, 415-426 (1992).

290 Seifert, K. N., McArthur, W. P., Bleiweis, A. S. & Brady, L. J. Characterization of group B streptococcal glyceraldehyde-3-phosphate dehydrogenase: surface localization, enzymatic activity, and protein-protein interactions. *Can J Microbiol* **49**, 350-356, doi:10.1139/w03-042 (2003).

291 Candela, M. *et al.* DnaK from Bifidobacterium animalis subsp. lactis is a surface-exposed human plasminogen receptor upregulated in response to bile salts. *Microbiology* **156**, 1609-1618, doi:10.1099/mic.0.038307-0 (2010).

292 Knaust, A. *et al.* Cytosolic proteins contribute to surface plasminogen recruitment of Neisseria meningitidis. *J Bacteriol* **189**, 3246-3255, doi:10.1128/JB.01966-06 (2007).

293 Cookson, E. J. & Beynon, R. J. Degradation artefacts during sample preparation for sodium dodecyl sulphate polyacrylamide gel electrophoresis. *Biosci Rep* **7**, 209-215 (1987).

294 Waites, K. B. & Atkinson, T. P. The role of Mycoplasma in upper respiratory infections. *Curr Infect Dis Rep* **11**, 198-206 (2009).

295 Taylor-Robinson, D. & Bebear, C. Antibiotic susceptibilities of mycoplasmas and treatment of mycoplasmal infections. *J Antimicrob Chemother* **40**, 622-630 (1997).

296 Razin, S., Yogev, D. & Naot, Y. Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev* **62**, 1094-1156 (1998).

297 Meseguer, M. A. *et al.* Mycoplasma pneumoniae: a reduced-genome intracellular bacterial pathogen. *Infect Genet Evol* **3**, 47-55 (2003).

298 Hasselbring, B. M., Jordan, J. L., Krause, R. W. & Krause, D. C. Terminal organelle development in the cell wall-less bacterium Mycoplasma pneumoniae. *Proc Natl Acad Sci U S A* **103**, 16478-16483, doi:10.1073/pnas.0608051103 (2006).

299 Chang, H. Y., Prince, O. A., Sheppard, E. S. & Krause, D. C. Processing is required for a fully functional protein P30 in Mycoplasma pneumoniae gliding and cytadherence. *J Bacteriol* **193**, 5841-5846, doi:10.1128/JB.00104-11 (2011).

300 Balish, M. F. & Krause, D. C. Mycoplasmas: a distinct cytoskeleton for wall-less bacteria. *J Mol Microbiol Biotechnol* **11**, 244-255, doi:10.1159/000094058 (2006).

301 Nakane, D., Adan-Kubo, J., Kenri, T. & Miyata, M. Isolation and characterization of P1 adhesin, a leg protein of the gliding bacterium Mycoplasma pneumoniae. *J Bacteriol* **193**, 715-722, doi:10.1128/JB.00796-10 (2011).

302 Layh-Schmitt, G. & Harkenthal, M. The 40- and 90-kDa membrane proteins (ORF6 gene product) of Mycoplasma pneumoniae are responsible for the tip structure formation and P1 (adhesin) association with the Triton shell. *FEMS Microbiol Lett* **174**, 143-149 (1999).

303 Waldo, R. H., 3rd, Jordan, J. L. & Krause, D. C. Identification and complementation of a mutation associated with loss of Mycoplasma pneumoniae virulence-specific proteins B and C. *J Bacteriol* **187**, 747-751, doi:10.1128/JB.187.2.747-751.2005 (2005).

304 Dallo, S. F., Chavoya, A. & Baseman, J. B. Characterization of the gene for a 30-kilodalton adhesion-related protein of Mycoplasma pneumoniae. *Infect Immun* **58**, 4163-4165 (1990).

305 Hu, P. C., Collier, A. M. & Clyde, W. A., Jr. Serological comparison of virulent and avirulent Mycoplasma pneumoniae by monoclonal antibodies. *Isr J Med Sci* **20**, 870-873 (1984).

306 Inamine, J. M., Loechel, S. & Hu, P. C. Analysis of the nucleotide sequence of the P1 operon of Mycoplasma pneumoniae. *Gene* **73**, 175-183 (1988).

307 Aravind, L. & Koonin, E. V. A novel family of predicted phosphoesterases includes Drosophila prune protein and bacterial RecJ exonuclease. *Trends Biochem Sci* **23**, 17-19 (1998).

308 Hansen, E. J., Wilson, R. M. & Baseman, J. B. Two-dimensional gel electrophoretic comparison of proteins from virulent and avirulent strains of Mycoplasma pneumoniae. *Infect Immun* **24**, 468-475 (1979).

309 Lipman, R. P., Clyde, W. A., Jr. & Denny, F. W. Characteristics of virulent, attenuated, and avirulent Mycoplasma pneumoniae strains. *J Bacteriol* **100**, 1037-1043 (1969).

310 Krivan, H. C., Olson, L. D., Barile, M. F., Ginsburg, V. & Roberts, D. D. Adhesion of Mycoplasma pneumoniae to sulfated glycolipids and inhibition by dextran sulfate. *J Biol Chem* **264**, 9283-9288 (1989).

311 McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* **41**, W597-600, doi:10.1093/nar/gkt376 (2013).

312 Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164, doi:10.1126/science.252.5009.1162 (1991).

313 Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K. & Obradovic, Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**, 208, doi:10.1186/1471-2105-7-208 (2006).

314 Radivojac, P., Obradovic, Z., Brown, C. J. & Dunker, A. K. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*, 216-227 (2003).

315 de Castro, E. *et al.* ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* **34**, W362-365, doi:10.1093/nar/gkl124 (2006).

316 Klimstra, W. B., Heidner, H. W. & Johnston, R. E. The furin protease cleavage recognition sequence of Sindbis virus PE2 can mediate virion attachment to cell surface heparan sulfate. *J Virol* **73**, 6299-6306 (1999).

317 Seto, S., Layh-Schmitt, G., Kenri, T. & Miyata, M. Visualization of the attachment organelle and cytadherence proteins of Mycoplasma pneumoniae by immunofluorescence microscopy. *J Bacteriol* **183**, 1621-1630, doi:10.1128/JB.183.5.1621-1630.2001 (2001).

318 Furano, A. V. Content of elongation factor Tu in Escherichia coli. *Proc Natl Acad Sci U S A* **72**, 4780-4784 (1975).

319 Sprinzl, M. Elongation factor Tu: a regulatory GTPase with an integrated effector. *Trends Biochem Sci* **19**, 245-250 (1994).

320 Polekhina, G. *et al.* Helix unwinding in the effector region of elongation factor EF-Tu-GDP. *Structure* **4**, 1141-1151 (1996).

321 Kjeldgaard, M., Nissen, P., Thirup, S. & Nyborg, J. The crystal structure of elongation factor EF-Tu from Thermus aquaticus in the GTP conformation. *Structure* **1**, 35-50 (1993).

322 Kjeldgaard, M. & Nyborg, J. Refined structure of elongation factor EF-Tu from Escherichia coli. *J Mol Biol* **223**, 721-742 (1992).

323 Clark, B. F., Kjeldgaard, M., la Cour, T. F., Thirup, S. & Nyborg, J. Structural determination of the functional sites of E. coli elongation factor Tu. *Biochim Biophys Acta* **1050**, 203-208 (1990).

324 Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc Natl Acad Sci U S A* **93**, 7749-7754 (1996).

325 Andersen, G. R., Valente, L., Pedersen, L., Kinzy, T. G. & Nyborg, J. Crystal structures of nucleotide exchange intermediates in the eEF1A-eEF1Balpha complex. *Nat Struct Biol* **8**, 531-534, doi:10.1038/88598 (2001).

326 Gaucher, E. A., Das, U. K., Miyamoto, M. M. & Benner, S. A. The crystal structure of eEF1A refines the functional predictions of an evolutionary analysis of rate changes among elongation factors. *Mol Biol Evol* **19**, 569-573 (2002).

327 Ejiri, S. Moonlighting functions of polypeptide elongation factor 1: from actin bundling to zinc finger protein R1-associated nuclear localization. *Biosci Biotechnol Biochem* **66**, 1-21, doi:10.1271/bbb.66.1 (2002).

328 Abbas, W., Kumar, A. & Herbein, G. The eEF1A Proteins: At the Crossroads of Oncogenesis, Apoptosis, and Viral Infections. *Front Oncol* **5**, 75, doi:10.3389/fonc.2015.00075 (2015).

329 Lamberti, A. *et al.* The translation elongation factor 1A in tumorigenesis, signal transduction and apoptosis: review article. *Amino Acids* **26**, 443-448, doi:10.1007/s00726-004-0088-2 (2004).

330 Mateyak, M. K. & Kinzy, T. G. eEF1A: thinking outside the ribosome. *J Biol Chem* **285**, 21209-21213, doi:10.1074/jbc.R110.113795 (2010).

331 Sasikumar, A. N., Perez, W. B. & Kinzy, T. G. The many roles of the eukaryotic elongation factor 1 complex. *Wiley Interdiscip Rev RNA* **3**, 543-555, doi:10.1002/wrna.1118 (2012).

332 Thornton, S., Anand, N., Purcell, D. & Lee, J. Not just for housekeeping: protein initiation and elongation factors in cell growth and tumorigenesis. *J Mol Med (Berl)* **81**, 536-548, doi:10.1007/s00109-003-0461-8 (2003).

333 Matsubayashi, M. *et al.* Elongation factor-1alpha is a novel protein associated with host cell invasion and a potential protective antigen of Cryptosporidium parvum. *J Biol Chem* **288**, 34111-34120, doi:10.1074/jbc.M113.515544 (2013).

334 Inomata, A. *et al.* Heparin interacts with elongation factor 1alpha of Cryptosporidium parvum and inhibits invasion. *Sci Rep* **5**, 11599, doi:10.1038/srep11599 (2015).

335 Nandan, D., Yi, T., Lopez, M., Lai, C. & Reiner, N. E. Leishmania EF-1alpha activates the Src homology 2 domain containing tyrosine phosphatase SHP-1 leading to macrophage deactivation. *J Biol Chem* **277**, 50190-50197, doi:10.1074/jbc.M209210200 (2002).

336 Alves, L. R., Oliveira, C. & Goldenberg, S. Eukaryotic translation elongation factor-1 alpha is associated with a specific subset of mRNAs in Trypanosoma cruzi. *BMC Microbiol* **15**, 104, doi:10.1186/s12866-015-0436-2 (2015).

337 Crowe, J. D. *et al.* Candida albicans binds human plasminogen: identification of eight plasminogen-binding proteins. *Mol Microbiol* **47**, 1637-1651 (2003).

338 Granato, D. *et al.* Cell surface-associated elongation factor Tu mediates the attachment of Lactobacillus johnsonii NCC533 (La1) to human intestinal cells and mucins. *Infect Immun* **72**, 2160-2169 (2004).

339 Kunert, A. *et al.* Immune evasion of the human pathogen Pseudomonas aeruginosa: elongation factor Tuf is a factor H and plasminogen binding protein. *J Immunol* **179**, 2979-2988 (2007).

340 Pasztor, L. *et al.* Staphylococcal major autolysin (Atl) is involved in excretion of cytoplasmic proteins. *J Biol Chem* **285**, 36794-36803, doi:10.1074/jbc.M110.167312 (2010).

341 Mohan, S. *et al.* Tuf of Streptococcus pneumoniae is a surface displayed human complement regulator binding protein. *Mol Immunol* **62**, 249-264, doi:10.1016/j.molimm.2014.06.029 (2014).

342 Wolff, D. G. *et al.* Interaction of Leptospira elongation factor Tu with plasminogen and complement factor H: a metabolic leptospiral protein with moonlighting activities. *PLoS One* **8**, e81818, doi:10.1371/journal.pone.0081818 (2013).

343 Vanden Bergh, P., Heller, M., Braga-Lagache, S. & Frey, J. The Aeromonas salmonicida subsp. salmonicida exoproteome: global analysis, moonlighting proteins and putative antigens for vaccination against furunculosis. *Proteome Sci* **11**, 44, doi:10.1186/1477-5956-11-44 (2013).

344 Schaumburg, J. *et al.* The cell wall subproteome of Listeria monocytogenes. *Proteomics* **4**, 2991-3006, doi:10.1002/pmic.200400928 (2004).

345 Barel, M. *et al.* 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, doi:10.1186/1471-2180-8-145 (2008).

346 Barel, M. & Charbit, A. Detection of the interaction between host and bacterial proteins: eukaryotic nucleolin interacts with Francisella elongation factor Tu. *Methods Mol Biol* **1197**, 123-139, doi:10.1007/978-1-4939-1261-2_7 (2014).

347 Xolalpa, W. *et al.* Identification of novel bacterial plasminogen-binding proteins in the human pathogen Mycobacterium tuberculosis. *Proteomics* **7**, 3332-3341, doi:10.1002/pmic.200600876 (2007).

348 Li, Q. *et al.* Identification of Novel Laminin- and Fibronectin-binding Proteins by Far-Western Blot: Capturing the Adhesins of Streptococcus suis Type 2. *Frontiers in cellular and infection microbiology* **5**, 82, doi:10.3389/fcimb.2015.00082 (2015).

349 Balbo, M., Barel, M., Lottin-Divoux, S., Jean, D. & Frade, R. Infection of human B lymphoma cells by Mycoplasma fermentans induces interaction of its elongation factor with the intracytoplasmic domain of Epstein-Barr virus receptor (gp140, EBV/C3dR, CR2, CD21). *FEMS Microbiol Lett* **249**, 359-366, doi:10.1016/j.femsle.2005.06.052 (2005).

350 Viale, M. N. *et al.* Description of a novel adhesin of Mycobacterium avium subsp. paratuberculosis. *Biomed Res Int* **2014**, 729618, doi:10.1155/2014/729618 (2014).

351 Munoz-Provencio, D., Perez-Martinez, G. & Monedero, V. Identification of Surface Proteins from Lactobacillus casei BL23 Able to Bind Fibronectin and Collagen. *Probiotics and antimicrobial proteins* **3**, 15-20, doi:10.1007/s12602-011-9065-8 (2011).

352 Jiang, F. *et al.* Elongation Factor Tu and Heat Shock Protein 70 Are Membrane-Associated Proteins from Mycoplasma ovipneumoniae Capable of Inducing Strong Immune Response in Mice. *PLoS One* **11**, e0161170, doi:10.1371/journal.pone.0161170 (2016).

353 Nishiyama, K. *et al.* Identification and characterization of sulfated carbohydrate-binding protein from Lactobacillus reuteri. *PLoS One* **8**, e83703, doi:10.1371/journal.pone.0083703 (2013).

354 Glowalla, E., Tosetti, B., Kronke, M. & Krut, O. Proteomics-based identification of anchorless cell wall proteins as vaccine candidates against Staphylococcus aureus. *Infect Immun* **77**, 2719-2729, doi:10.1128/IAI.00617-08 (2009).

355 Kloppot, P. *et al.* Microarray-based identification of human antibodies against Staphylococcus aureus antigens. *Proteomics. Clinical applications* **9**, 1003-1011, doi:10.1002/prca.201400123 (2015).

356 Churchward, C. P. *et al.* Immunoproteomic characterisation of Mycoplasma mycoides subspecies capri by mass spectrometry analysis of two-dimensional electrophoresis spots and western blot. *The Journal of pharmacy and pharmacology* **67**, 364-371, doi:10.1111/jphp.12344 (2015).

357 Sanchez-Campillo, M. *et al.* Identification of immunoreactive proteins of Chlamydia trachomatis by Western blot analysis of a two-dimensional electrophoresis map with patient sera. *Electrophoresis* **20**, 2269-2279, doi:10.1002/(SICI)1522-2683(19990801)20:11<2269::AID-ELPS2269>3.0.CO;2-D (1999).

358 Nieves, W. *et al.* Immunospecific responses to bacterial elongation factor Tu during Burkholderia infection and immunization. *PLoS One* **5**, e14361, doi:10.1371/journal.pone.0014361 (2010).

359 Pinto, P. M. *et al.* Proteomic survey of the pathogenic Mycoplasma hyopneumoniae strain 7448 and identification of novel post-translationally modified and antigenic proteins. *Vet Microbiol* **121**, 83-93, doi:10.1016/j.vetmic.2006.11.018 (2007).

360 Hempel, K. *et al.* 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, doi:10.1021/pr901143a (2010).

361 Dreisbach, A. *et al.* Profiling the surfacome of Staphylococcus aureus. *Proteomics* **10**, 3082-3096, doi:10.1002/pmic.201000062 (2010).

362 Ventura, C. L. *et al.* Identification of a novel Staphylococcus aureus two-component leukotoxin using cell surface proteomics. *PLoS One* **5**, e11634, doi:10.1371/journal.pone.0011634 (2010).

363 Hempel, K., Herbst, F. A., Moche, M., Hecker, M. & Becher, D. 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, doi:10.1021/pr1009838 (2011).

364 Monteiro, R. *et al.* Surfaceome and exoproteome of a clinical sequence type 398 methicillin resistant Staphylococcus aureus strain. *Biochemistry and Biophysics Reports* **3**, 7-13, doi:http://dx.doi.org/10.1016/j.bbrep.2015.07.004 (2015).

365 Liew, Y. K., Awang Hamat, R., van Belkum, A., Chong, P. P. & Neela, V. Comparative Exoproteomics and Host Inflammatory Response in Staphylococcus aureus Skin and Soft Tissue Infections, Bacteremia, and Subclinical Colonization. *Clin Vaccine Immunol* **22**, 593-603, doi:10.1128/CVI.00493-14 (2015).

366 Dallo, S. F. *et al.* Association of Acinetobacter baumannii EF-Tu with cell surface, outer membrane vesicles, and fibronectin. *TheScientificWorldJournal* **2012**, 128705, doi:10.1100/2012/128705 (2012).

367 Ebner, P. *et al.* Excretion of cytoplasmic proteins (ECP) in Staphylococcus aureus. *Mol Microbiol* **97**, 775-789, doi:10.1111/mmi.13065 (2015).

368 Jones, D. A. & Takemoto, D. Plant innate immunity - direct and indirect recognition of general and specific pathogen-associated molecules. *Current opinion in immunology* **16**, 48-62 (2004).

369 Zipfel, C. *et al.* Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell* **125**, 749-760, doi:10.1016/j.cell.2006.03.037 (2006).

370 Gomez-Gomez, L. & Boller, T. FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Molecular cell* **5**, 1003-1011 (2000).

371 Sharp, J. K., McNeil, M. & Albersheim, P. The primary structures of one elicitor-active and seven elicitor-inactive hexa(beta-D-glucopyranosyl)-D-glucitols isolated from the mycelial walls of Phytophthora megasperma f. sp. glycinea. *J Biol Chem* **259**, 11321-11336 (1984).

372 Miya, A. *et al.* CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. *Proc Natl Acad Sci U S A* **104**, 19613-19618, doi:10.1073/pnas.0705147104 (2007).

373 Ponchet, M. *et al.* Are elicitins cryptograms in plant-Oomycete communications? *Cell Mol Life Sci* **56**, 1020-1047 (1999).

374 Wan, J. *et al.* A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in Arabidopsis. *The Plant cell* **20**, 471-481, doi:10.1105/tpc.107.056754 (2008).

375 Nurnberger, T., Brunner, F., Kemmerling, B. & Piater, L. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunological reviews* **198**, 249-266 (2004).

376 Boller, T. & Felix, G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual review of plant biology* **60**, 379-406, doi:10.1146/annurev.arplant.57.032905.105346 (2009).

377 Kunze, G. *et al.* The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. *The Plant cell* **16**, 3496-3507, doi:10.1105/tpc.104.026765 (2004).

378 Furukawa, T., Inagaki, H., Takai, R., Hirai, H. & Che, F. S. Two distinct EF-Tu epitopes induce immune responses in rice and Arabidopsis. *Molecular plant-microbe interactions : MPMI* **27**, 113-124, doi:10.1094/MPMI-10-13-0304-R (2014).

379 Lu, F. *et al.* Enhancement of innate immune system in monocot rice by transferring the dicotyledonous elongation factor Tu receptor EFR. *Journal of integrative plant biology* **57**, 641-652, doi:10.1111/jipb.12306 (2015).

380 Chang, C. C. *et al.* Fragmentation of CagA Reduces Hummingbird Phenotype Induction by Helicobactor pylori. *PLoS One* **11**, e0150061, doi:10.1371/journal.pone.0150061 (2016).

381 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539, doi:10.1038/msb.2011.75 (2011).

382 Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529-533, doi:10.1093/nar/gkq399 (2010).

383 Ofran, Y. & Rost, B. ISIS: interaction sites identified from sequence. *Bioinformatics* **23**, e13-16, doi:10.1093/bioinformatics/btl303 (2007).

384 Kozlowski, L. P. & Bujnicki, J. M. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **13**, 111, doi:10.1186/1471-2105-13-111 (2012).

385 Schlessinger, A., Punta, M., Yachdav, G., Kajan, L. & Rost, B. Improved disorder prediction by combination of orthogonal approaches. *PLoS One* **4**, e4433, doi:10.1371/journal.pone.0004433 (2009).

386 Rost, B., Yachdav, G. & Liu, J. The PredictProtein server. *Nucleic Acids Res* **32**, W321-326, doi:10.1093/nar/gkh377 (2004).

387 P., H. *Diploma thesis*, Technische Universität München, (2012).

388 Rawlings, N. D., Barrett, A. J. & Finn, R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **44**, D343-350, doi:10.1093/nar/gkv1118 (2016).

389 Balasubramanian, S., Kannan, T. R. & Baseman, J. B. The surface-exposed carboxyl region of Mycoplasma pneumoniae elongation factor Tu interacts with fibronectin. *Infect Immun* **76**, 3116-3123, doi:10.1128/IAI.00173-08 (2008).

390 Balasubramanian, S., Kannan, T. R., Hart, P. J. & Baseman, J. B. Amino acid changes in elongation factor Tu of Mycoplasma pneumoniae and Mycoplasma genitalium influence fibronectin binding. *Infect Immun* **77**, 3533-3541, doi:10.1128/IAI.00081-09 (2009).

391 Liang, O. D., Ascencio, F., Fransson, L. A. & Wadstrom, T. Binding of heparan sulfate to Staphylococcus aureus. *Infect Immun* **60**, 899-906 (1992).

392 Fallgren, C., Utt, M. & Ljungh, A. Isolation and characterisation of a 17-kDa staphylococcal heparin-binding protein with broad specificity. *J Med Microbiol* **50**, 547-557, doi:10.1099/0022-1317-50-6-547 (2001).

393 Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **47**, 5 6 1-32, doi:10.1002/0471250953.bi0506s47 (2014).

394 Gotz, F., Yu, W., Dube, L., Prax, M. & Ebner, P. Excretion of cytosolic proteins (ECP) in bacteria. *Int J Med Microbiol* **305**, 230-237, doi:10.1016/j.ijmm.2014.12.021 (2015).

395 Pancholi, V. & Chhatwal, G. S. Housekeeping enzymes as virulence factors for pathogens. *Int J Med Microbiol* **293**, 391-401, doi:10.1078/1438-4221-00283 (2003).

396 Wang, G. *et al.* The Roles of Moonlighting Proteins in Bacteria. *Curr Issues Mol Biol* **16**, 15-22 (2013).

397 Davey, N. E. *et al.* Attributes of short linear motifs. *Mol Biosyst* **8**, 268-281, doi:10.1039/c1mb05231d (2012).

398 Davey, N. E., Cyert, M. S. & Moses, A. M. Short linear motifs - ex nihilo evolution of protein regulation. *Cell communication and signaling : CCS* **13**, 43, doi:10.1186/s12964-015-0120-z (2015).

399 Gibson, T. J., Dinkel, H., Van Roey, K. & Diella, F. Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell communication and signaling : CCS* **13**, 42, doi:10.1186/s12964-015-0121-y (2015).

400 Wright, P. E. & Dyson, H. J. Linking folding and binding. *Current opinion in structural biology* **19**, 31-38, doi:10.1016/j.sbi.2008.12.003 (2009).

401 Boone, T. J. & Tyrrell, G. J. Identification of the actin and plasminogen binding regions of group B streptococcal phosphoglycerate kinase. *J Biol Chem* **287**, 29035-29044, doi:10.1074/jbc.M112.361261 (2012).

402 Sun, Y. *et al.* Factors influencing the nuclear targeting ability of nuclear localization signals. *J Drug Target* **24**, 927-933, doi:10.1080/1061186X.2016.1184273 (2016).

403 Robbins, J., Dilworth, S. M., Laskey, R. A. & Dingwall, C. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**, 615-623 (1991).

404 Konkel, M. E. *et al.* Identification of a fibronectin-binding domain within the Campylobacter jejuni CadF protein. *Mol Microbiol* **57**, 1022-1035, doi:10.1111/j.1365-2958.2005.04744.x (2005).

405 Davis, F. P. & Sali, A. The overlap of small molecule and protein binding sites within families of protein structures. *PLoS computational biology* **6**, e1000668, doi:10.1371/journal.pcbi.1000668 (2010).

406 Duensing, T. D., Wing, J. S. & van Putten, J. P. Sulfated polysaccharide-directed recruitment of mammalian host proteins: a novel strategy in microbial pathogenesis. *Infect Immun* **67**, 4463-4468 (1999).

407 Otto, M. Staphylococcal biofilms. *Curr Top Microbiol Immunol* **322**, 207-228 (2008).

408 Lister, J. L. & Horswill, A. R. Staphylococcus aureus biofilms: recent developments in biofilm dispersal. *Frontiers in cellular and infection microbiology* **4**, 178, doi:10.3389/fcimb.2014.00178 (2014).

409 Foulston, L., Elsholz, A. K., DeFrancesco, A. S. & Losick, R. 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, doi:10.1128/mBio.01667-14 (2014).

410 Speziale, P., Pietrocola, G., Foster, T. J. & Geoghegan, J. A. Protein-based biofilm matrices in Staphylococci. *Frontiers in cellular and infection microbiology* **4**, 171, doi:10.3389/fcimb.2014.00171 (2014).

411 Izano, E. A., Amarante, M. A., Kher, W. B. & Kaplan, J. B. Differential roles of poly-N-acetylglucosamine surface polysaccharide and extracellular DNA in Staphylococcus aureus and Staphylococcus epidermidis biofilms. *Appl Environ Microbiol* **74**, 470-476, doi:10.1128/AEM.02073-07 (2008).

412 Mann, E. E. *et al.* Modulation of eDNA release and degradation affects Staphylococcus aureus biofilm maturation. *PLoS One* **4**, e5822, doi:10.1371/journal.pone.0005822 (2009).

413 Schwartz, K., Ganesan, M., Payne, D. E., Solomon, M. J. & Boles, B. R. Extracellular DNA facilitates the formation of functional amyloids in Staphylococcus aureus biofilms. *Mol Microbiol* **99**, 123-134, doi:10.1111/mmi.13219 (2016).

414 Dengler, V., Foulston, L., DeFrancesco, A. S. & Losick, R. An Electrostatic Net Model for the Role of Extracellular DNA in Biofilm Formation by Staphylococcus aureus. *J Bacteriol* **197**, 3779-3787, doi:10.1128/JB.00726-15 (2015).

415 Hou, S. *et al.* Chlamydia trachomatis outer membrane complex protein B (OmcB) is processed by the protease CPAF. *J Bacteriol* **195**, 951-957, doi:10.1128/JB.02087-12 (2013).

416 Guillot, A. *et al.* Mass Spectrometry Analysis of the Extracellular Peptidome of Lactococcus lactis: Lines of Evidence for the Coexistence of Extracellular Protein Hydrolysis and Intracellular Peptide Excretion. *J Proteome Res* **15**, 3214-3224, doi:10.1021/acs.jproteome.6b00424 (2016).

417 Duffy, M. F., Walker, I. D. & Browning, G. F. The immunoreactive 116 kDa surface protein of Mycoplasma pneumoniae is encoded in an operon. *Microbiology* **143 (Pt 10)**, 3391-3402, doi:10.1099/00221287-143-10-3391 (1997).

418 Sluijter, M., Spuesens, E. B., Hartwig, N. G., van Rossum, A. M. & Vink, C. The Mycoplasma pneumoniae MPN490 and Mycoplasma genitalium MG339 genes encode reca homologs that promote homologous DNA strand exchange. *Infect Immun* **77**, 4905-4911, doi:10.1128/IAI.00747-09 (2009).

419 Jacobs, E., Vonski, M., Oberle, K., Opitz, O. & Pietsch, K. Are outbreaks and sporadic respiratory infections by Mycoplasma pneumoniae due to two distinct subtypes? *Eur J Clin Microbiol Infect Dis* **15**, 38-44 (1996).

420 Dumke, R., Catrein, I., Pirkil, E., Herrmann, R. & Jacobs, E. Subtyping of Mycoplasma pneumoniae isolates based on extended genome sequencing and on expression profiles. *Int J Med Microbiol* **292**, 513-525 (2003).

421 Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318-1322 (1999).

422 Hirschberg, L., Holme, T. & Krook, A. Human antibody response to the major adhesin of Mycoplasma pneumoniae: increase in titers against synthetic peptides in patients with pneumonia. *APMIS* **99**, 515-520 (1991).

423 Tuuminen, T., Suni, J., Kleemola, M. & Jacobs, E. Improved sensitivity and specificity of enzyme immunoassays with P1-adhesin enriched antigen to detect acute Mycoplasma pneumoniae infection. *J Microbiol Methods* **44**, 27-37 (2001).

424 Hahn, T. W., Willby, M. J. & Krause, D. C. HMW1 is required for cytadhesin P1 trafficking to the attachment organelle in Mycoplasma pneumoniae. *J Bacteriol* **180**, 1270-1276 (1998).

425 Dallo, S. F., Su, C. J., Horton, J. R. & Baseman, J. B. Identification of P1 gene domain containing epitope(s) mediating Mycoplasma pneumoniae cytoadherence. *J Exp Med* **167**, 718-723 (1988).

426 Gerstenecker, B. & Jacobs, E. Topological mapping of the P1-adhesin of Mycoplasma pneumoniae with adherence-inhibiting monoclonal antibodies. *J Gen Microbiol* **136**, 471-476, doi:10.1099/00221287-136-3-471 (1990).

427 Jacobs, E., Bartl, A., Oberle, K. & Schiltz, E. Molecular mimicry by Mycoplasma pneumoniae to evade the induction of adherence inhibiting antibodies. *J Med Microbiol* **43**, 422-429, doi:10.1099/00222615-43-6-422 (1995).

428 Chaudhry, R., Nisar, N., Hora, B., Chirasani, S. R. & Malhotra, P. Expression and immunological characterization of the carboxy-terminal region of the P1 adhesin protein of Mycoplasma pneumoniae. *J Clin Microbiol* **43**, 321-325, doi:10.1128/JCM.43.1.321-325.2005 (2005).

429 Drasbek, M., Christiansen, G., Drasbek, K. R., Holm, A. & Birkelund, S. Interaction between the P1 protein of Mycoplasma pneumoniae and receptors on HEp-2 cells. *Microbiology* **153**, 3791-3799, doi:10.1099/mic.0.2007/010736-0 (2007).

430 Beghetto, E., De Paolis, F., Montagnani, F., Cellesi, C. & Gargano, N. Discovery of new Mycoplasma pneumoniae antigens by use of a whole-genome lambda display library. *Microbes Infect* **11**, 66-73, doi:10.1016/j.micinf.2008.10.004 (2009).

431 Schmidl, S. R. *et al.* The stability of cytadherence proteins in Mycoplasma pneumoniae requires activity of the protein kinase PrkC. *Infect Immun* **78**, 184-192, doi:10.1128/IAI.00958-09 (2010).

432 Kyhse-Andersen, J. 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 (1984).

433 Padula, M. P. *et al.* A Comprehensive Guide for Performing Sample Preparation and Top-Down Protein Analysis. *Proteomes* **5**, doi:10.3390/proteomes5020011 (2017).

434 Grover, R. K. *et al.* A structurally distinct human mycoplasma protein that generically blocks antigen-antibody union. *Science* **343**, 656-661, doi:10.1126/science.1246135 (2014).

435 Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochim Biophys Acta* **1804**, 1231-1264, doi:10.1016/j.bbapap.2010.01.017 (2010).

436 Boguslavsky, S. *et al.* Molecular characterization of the Mycoplasma gallisepticum pvpA gene which encodes a putative variable cytadhesin protein. *Infect Immun* **68**, 3956-3964 (2000).

437 Williamson, M. P. The structure and function of proline-rich regions in proteins. *Biochem J* **297 (Pt 2)**, 249-260 (1994).

438 Kay, B. K., Williamson, M. P. & Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* **14**, 231-241 (2000).

439 Cork, A. J. *et al.* Defining the structural basis of human plasminogen binding by streptococcal surface enolase. *J Biol Chem* **284**, 17129-17137, doi:10.1074/jbc.M109.004317 (2009).

440 Lee, A. Y., Hsu, C. H. & Wu, S. H. Functional domains of Brevibacillus thermoruber lon protease for oligomerization and DNA binding: role of N-terminal and sensor and substrate discrimination domains. *J Biol Chem* **279**, 34903-34912, doi:10.1074/jbc.M403562200 (2004).

441 Yavlovich, A., Tarshis, M. & Rottem, S. Internalization and intracellular survival of Mycoplasma pneumoniae by non-phagocytic cells. *FEMS Microbiol Lett* **233**, 241-246, doi:10.1016/j.femsle.2004.02.016 (2004).

442 Tsiodras, S., Kelesidis, I., Kelesidis, T., Stamboulis, E. & Giamarellou, H. Central nervous system manifestations of Mycoplasma pneumoniae infections. *J Infect* **51**, 343-354, doi:10.1016/j.jinf.2005.07.005 (2005).

443 van Niekerk, C. C., Jap, P. H., Ramaekers, F. C., van de Molengraft, F. & Poels, L. G. Immunohistochemical demonstration of keratin 7 in routinely fixed paraffin-embedded human tissues. *J Pathol* **165**, 145-152, doi:10.1002/path.1711650210 (1991).

444 Sandilands, A. *et al.* Generation and characterisation of keratin 7 (K7) knockout mice. *PLoS One* **8**, e64404, doi:10.1371/journal.pone.0064404 (2013).

445 Owens, D. W. & Lane, E. B. The quest for the function of simple epithelial keratins. *Bioessays* **25**, 748-758, doi:10.1002/bies.10316 (2003).

446 Dobashi, N. *et al.* Detection of anti-cytokeratin 8 antibody in the serum of patients with cryptogenic fibrosing alveolitis and pulmonary fibrosis associated with collagen vascular disorders. *Thorax* **53**, 969-974 (1998).

447 Haim, M. *et al.* Cytokeratin 8 interacts with clumping factor B: a new possible virulence factor target. *Microbiology* **156**, 3710-3721, doi:10.1099/mic.0.034413-0 (2010).

448 Godfroid, E., Geuskens, M., Dupressoir, T., Parent, I. & Szpirer, C. Cytokeratins are exposed on the outer surface of established human mammary carcinoma cells. *J Cell Sci* **99 (Pt 3)**, 595-607 (1991).

449 Hembrough, T. A., Vasudevan, J., Allietta, M. M., Glass, W. F., 2nd & Gonias, S. L. A cytokeratin 8-like protein with plasminogen-binding activity is present on the external

surfaces of hepatocytes, HepG2 cells and breast carcinoma cell lines. *J Cell Sci* **108 (Pt 3)**, 1071-1082 (1995).

450 Lu, X. & Lane, E. B. Retrovirus-mediated transgenic keratin expression in cultured fibroblasts: specific domain functions in keratin stabilization and filament formation. *Cell* **62**, 681-696 (1990).

451 Moll, R., Franke, W. W., Schiller, D. L., Geiger, B. & Krepler, R. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell* **31**, 11-24 (1982).

452 Franke, W. W., Schiller, D. L., Hatzfeld, M. & Winter, S. Protein complexes of intermediate-sized filaments: melting of cytokeratin complexes in urea reveals different polypeptide separation characteristics. *Proc Natl Acad Sci U S A* **80**, 7113-7117 (1983).

453 Mendez, M. G., Kojima, S. & Goldman, R. D. Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. *FASEB J* **24**, 1838-1851, doi:10.1096/fj.09-151639 (2010).

454 Mor-Vaknin, N., Punturieri, A., Sitwala, K. & Markovitz, D. M. Vimentin is secreted by activated macrophages. *Nat Cell Biol* **5**, 59-63, doi:10.1038/ncb898 (2003).

455 Tamura, G. S. & Nittayajarn, A. Group B streptococci and other gram-positive cocci bind to cytokeratin 8. *Infect Immun* **68**, 2129-2134 (2000).

456 Scherer, C. A., Cooper, E. & Miller, S. I. The Salmonella type III secretion translocon protein SspC is inserted into the epithelial cell plasma membrane upon infection. *Mol Microbiol* **37**, 1133-1145 (2000).

457 Carlson, S. A., Omary, M. B. & Jones, B. D. Identification of cytokeratins as accessory mediators of Salmonella entry into eukaryotic cells. *Life Sci* **70**, 1415-1426 (2002).

458 Mak, T. N. *et al.* Propionibacterium acnes host cell tropism contributes to vimentin-mediated invasion and induction of inflammation. *Cell Microbiol* **14**, 1720-1733, doi:10.1111/j.1462-5822.2012.01833.x (2012).

459 Batchelor, M. *et al.* Involvement of the intermediate filament protein cytokeratin-18 in actin pedestal formation during EPEC infection. *EMBO Rep* **5**, 104-110, doi:10.1038/sj.embor.7400038 (2004).

460 Viswanathan, V. K. *et al.* Cytokeratin 18 interacts with the enteropathogenic Escherichia coli secreted protein F (EspF) and is redistributed after infection. *Cell Microbiol* **6**, 987-997, doi:10.1111/j.1462-5822.2004.00416.x (2004).

461 Icenogle, L. M. *et al.* Molecular and biological characterization of Streptococcal SpyA-mediated ADP-ribosylation of intermediate filament protein vimentin. *J Biol Chem* **287**, 21481-21491, doi:10.1074/jbc.M112.370791 (2012).

462 Murli, S., Watson, R. O. & Galan, J. E. Role of tyrosine kinases and the tyrosine phosphatase SptP in the interaction of Salmonella with host cells. *Cell Microbiol* **3**, 795-810 (2001).

463 Guignot, J. & Servin, A. L. Maintenance of the Salmonella-containing vacuole in the juxtanuclear area: a role for intermediate filaments. *Microb Pathog* **45**, 415-422, doi:10.1016/j.micpath.2008.09.007 (2008).

464 Saberi, S. *et al.* A potential association between Helicobacter pylori CagA EPIYA and multimerization motifs with cytokeratin 18 cleavage rate during early apoptosis. *Helicobacter* **17**, 350-357, doi:10.1111/j.1523-5378.2012.00954.x (2012).

465 Kumar, Y. & Valdivia, R. H. Actin and intermediate filaments stabilize the Chlamydia trachomatis vacuole by forming dynamic structural scaffolds. *Cell Host Microbe* **4**, 159-169, doi:10.1016/j.chom.2008.05.018 (2008).

466 Bove, J. M., Renaudin, J., Saillard, C., Foissac, X. & Garnier, M. Spiroplasma citri, a plant pathogenic molligute: relationships with its two hosts, the plant and the leafhopper vector. *Annu Rev Phytopathol* **41**, 483-500, doi:10.1146/annurev.phyto.41.052102.104034 (2003).

467 Bebear, C., Pereyre, S. & Peuchant, O. Mycoplasma pneumoniae: susceptibility and resistance to antibiotics. *Future Microbiol* **6**, 423-431, doi:10.2217/fmb.11.18 (2011).

468 Li, A. *et al.* Chemical cleavage at aspartyl residues for protein identification. *Anal Chem* **73**, 5395-5402 (2001).

469 Birktoft, J. J. & Breddam, K. Glutamyl endopeptidases. *Methods Enzymol* **244**, 114-126 (1994).

470 Pizza, M. *et al.* Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**, 1816-1820 (2000).

471 Sette, A. & Rappuoli, R. Reverse vaccinology: developing vaccines in the era of genomics. *Immunity* **33**, 530-541, doi:10.1016/j.immuni.2010.09.017 (2010).

472 Rappuoli, R., Pizza, M., Del Giudice, G. & De Gregorio, E. Vaccines, new opportunities for a new society. *Proc Natl Acad Sci U S A* **111**, 12288-12293, doi:10.1073/pnas.1402981111 (2014).

473 Bayles, K. W. The biological role of death and lysis in biofilm development. *Nat Rev Microbiol* **5**, 721-726, doi:10.1038/nrmicro1743 (2007).

474 Jacob, F. Evolution and tinkering. *Science* **196**, 1161-1166 (1977).

475 Gancedo, C. & Flores, C. L. Moonlighting proteins in yeasts. *Microbiol Mol Biol Rev* **72**, 197-210, table of contents, doi:10.1128/MMBR.00036-07 (2008).

476 Flores, C. L. & Gancedo, C. Unraveling moonlighting functions with yeasts. *IUBMB Life* **63**, 457-462, doi:10.1002/iub.454 (2011).

477 James, L. C. & Tawfik, D. S. Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. *Trends Biochem Sci* **28**, 361-368, doi:10.1016/S0968-0004(03)00135-X (2003).

478 Tompa, P., Szasz, C. & Buday, L. Structural disorder throws new light on moonlighting. *Trends Biochem Sci* **30**, 484-489, doi:10.1016/j.tibs.2005.07.008 (2005).

479 Tokuriki, N. & Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668-673, doi:10.1038/nature08009 (2009).

480 Noda-Garcia, L. & Barona-Gomez, F. Enzyme evolution beyond gene duplication: A model for incorporating horizontal gene transfer. *Mob Genet Elements* **3**, e26439, doi:10.4161/mge.26439 (2013).

481 Espinosa-Cantu, A., Ascencio, D., Barona-Gomez, F. & DeLuna, A. Gene duplication and the evolution of moonlighting proteins. *Front Genet* **6**, 227, doi:10.3389/fgene.2015.00227 (2015).

482 Huang, S. H. *et al.* Vimentin, a Novel NF-kappaB Regulator, Is Required for Meningitic Escherichia coli K1-Induced Pathogen Invasion and PMN Transmigration across the Blood-Brain Barrier. *PLoS One* **11**, e0162641, doi:10.1371/journal.pone.0162641 (2016).

483 Woof, J. M. Immunoglobulins and their receptors, and subversion of their protective roles by bacterial pathogens. *Biochem Soc Trans* **44**, 1651-1658, doi:10.1042/BST20160246 (2016).

484 Bereiter, M., Young, T. F., Joo, H. S. & Ross, R. F. 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 (1990).

485 Scarman, A. L., Chin, J. C., Eamens, G. J., Delaney, S. F. & Djordjevic, S. P. Identification of novel species-specific antigens of Mycoplasma hyopneumoniae by preparative SDS-PAGE ELISA profiling. *Microbiology* **143 (Pt 2)**, 663-673, doi:10.1099/00221287-143-2-663 (1997).

486 Bordier, C. Phase separation of integral membrane proteins in Triton X-114 solution. *J Biol Chem* **256**, 1604-1607 (1981).

487 Yachdav, G. *et al.* PredictProtein--an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res* **42**, W337-343, doi:10.1093/nar/gku366 (2014).

488 Hönigschmid, P. *Improvement of DNA- and RNA protein binding prediction.*, Technische Universität München, (2012).