UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

CENTRE FOR BIOLOGICAL SCIENCES

THE DEVELOPMENT AND APPLICATION OF PROTEOMICS TO THE ANALYSIS OF *CHLAMYDIA TRACHOMATIS*

Paul James Stuart Skipp

Thesis for the degree of Doctor of Philosophy

May 2012

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

CENTRE FOR BIOLOGICAL SCIENCES

Doctor of Philosophy

THE DEVELOPMENT AND APPLICATION OF PROTEOMICS TO THE ANALYSIS OF *CHLAMYDIA TRACHOMATIS*

Paul James Stuart Skipp

The bacterial pathogen *Chlamydia trachomatis* causes Trachoma, the worlds leading cause of preventable blindness and is also responsible for the most common curable sexually transmitted disease in the UK and United States. *C. trachomatis* is an obligate intracellular organism characterised by a unique and complex growth cycle. Its study presents many challenges since it has historically been recalcitrant to genetic manipulation and growth in the absence of a host cell. Nevertheless, the sequencing of the *C. trachomatis* genome and its relatively small size by comparison to genomes from other bacterial pathogens, has paved the way for studies at the proteomic level.

This thesis describes the development and application of proteomic approaches to study *C. trachomatis* L2. To survey the expressed chlamydial proteome, a combination of the qualitative approaches, 2-DGE, MudPIT and GeLC-MS/MS; and the quantitative approaches AQUA, iTRAQ and LC-MS^E were used. Collectively, the approaches efficiently identified 648 expressed proteins, representing ~72% of the predicted proteome of *C. trachomatis* L2, from both the infectious (elementary body, EB) and replicating (reticulate body, RB) form of the pathogen. In the infectious EB, the entire set of predicted glycolytic enzymes were detected, indicating that metabolite flux rather than *de novo* synthesis of this pathway is triggered upon infection of host cells. Further, proteomic analysis of the RB form also uncovered biosynthetic enzymes for chlamydial cell wall synthesis, indicating that peptidoglycan is produced in some form during growth in host cells.

Comparison of the quantitative approaches iTRAQ and LC-MS^E demonstrated that LC-MS^E quantitative data was significantly more robust and extensive relative to iTRAQ data. In addition to information on relative amounts of these proteins between the two forms, LC-MS^E data also yielded the cellular concentration (molecules per cell) for 489 proteins.

This extensive set of absolute quantitation data permits estimates of the energy invested in the synthesis of various classes of proteins. The results indicate that *C. trachomatis* devotes most of its energy into maintenance of the translational machinery. However, it also expends significant amounts of energy into making cell envelope components and a set of hitherto hypothetical proteins. These proteins, which account for the bulk of the energy invested by the intracellular RB form of the pathogen as it converts to the extracellular EB form, highlight the importance of absolute quantitation data for understanding the biological processing status of the cell.

The datasets also revealed a large number of proteins that were differentially expressed between replicating RBs and infectious EBs, ranging from 8.4-fold down-regulation to 3.5-fold up-regulation. Consistent with transcriptomic studies (Belland *et al*., 2003), proteins involved in protein synthesis, ATP generation, central metabolism, secretion and nutrient uptake were predominant in the metabolically active RB at 15 h PI. Although many of the proteins in these functional categories were down-regulated in EBs, proteins required for glycolysis, central metabolism, protein synthesis, and type III secretion were present in significant amounts in EBs suggesting that the infectious EB is primed ‘ready-to-go’ upon contact with the host cell.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

.

DECLARATION OF AUTHORSHIP

I, **Paul James Stuart Skipp**

declare that the thesis entitled **"The development and application of proteomics to the analysis of *Chlamydia trachomatis*"** and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: Skipp, P., Robinson, J., O'Connor, C.D., and Clarke, I.N. (2005). Shotgun proteomic analysis of *Chlamydia trachomatis*. Proteomics 11, 1868-1892.

Signed:

Date: 17th May 2012

Acknowledgements

I would like to thank a number of people who have supported me throughout the course of my PhD. Firstly; I would like to extend my gratitude to Professor David O'Connor (Centre for Biological Sciences, Southampton) for his continual support and encouragement. In particular, I would like to thank him for the many enjoyable and insightful discussions and for sharing his enthusiasm for science. I am also indebted to Professor Ian Clarke (Microbiology, Southampton General Hospital), not only for his support and encouragement, but also for introducing and sharing his extensive knowledge of *Chlamydia* research with me.

I would like to acknowledge many other people who have been a source of help through the course of my PhD; My gratitude extends to Mrs Leslie Cutcliffe (Microbiology, Southampton General Hospital) for the excellent technical support provided in culturing and purifying chlamydiae. Dr Joanne Spencer for support in culturing and purifying chlamydiae. Dr Chris Hughes (Waters Life Sciences Laboratory, Manchester, UK) for his training and expertise in the operation of the Synapt G2 for the label-free analyses. Dr Simon Miles and Mr James Rodger (Electronics and Computer Science, Southampton) for assistance in development of the iTRAQ parser. Dr Richard Edwards (Centre for Biological Sciences, Southampton) for bioinformatic assistance in the identification and removal of human homologues from the chlamydial datasets. Dr Erika Parkinson (Centre for Proteomic Research, Southampton) for her encouragement and insightful discussions during the preparation of this thesis. Mrs Lorraine Prout (Centre for Biological Sciences) and Mrs Maureen Smith (Centre for Biological Sciences), whom have been a pillar of support and encouragment (Centre for Biological Sciences) and Dr Vincent O'Connor (Centre for Biological Sciences) for his open advice and support.

Finally, I am indebted to my partner Stephen Forster and my parents James and Marilyn Skipp for their immeasurable support and continual encouragement throughout the course of my PhD.

LIST OF ABBREVIATIONS

1-DGE	one dimensional gel electrophoresis
2-DGE	two dimensional gel electrophoresis
ATP	adenosine triphosphate
BCA	bicinchoninic acid
BGMK	Buffalo Green Monkey Kidney
BSA	bovine serum albumin
CAD	collisionally activated dissociation
CID	collisionally induced dissociation
CIM	chlamydial inclusion membrane
COMC	chlamydial outer membrane complex
DTT	dithiothreitol
EB	Elementary body
EIA	enzyme immunoassays
ELISA	enzyme-linked immunosorbant assays
ESI	electropsray ionisation
FT-ICR	fourier transform ion cyclotron resonance
GeLC-MS/MS	gel electrophoresis liquid chromatography tandem mass spectrometry
HPLC	high performance liquid chromatography
ICAT	isotope coded affinity tags
IDO	Indoleamine 2,3 - dioxygenase
IFN	Interferron
IPG	immobilised pH gradient
iTRAQ	isobaric tags for relative and absolute quantitation
LB	luria broth
LC	liquid chromatography
LCR	ligase chain reaction
LGV	Lymphogranuloma venereum
LOQ	lower limit of quantitation
LPS	Lipopolysaccharide
m/z	mass to charge ratio
MALDI	matrix assisted laser desporption ionisation
MIF	micro-immunofluorescence
MMTS	methyl methane-thiosulfonate
MOI	multiplicity of infection
MOMP	major outer membrane protein
MRM	multiple reaction monitoring
MS	mass spectrometry

MS/MS	tandem mass spectrometry
MudPIT	multidimensional protein identification technology
NGU	non-gonococcal urethritis
OaTof	orthogonal acceleration time of flight
OM	outer membrane
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PG	peptidoglycan
PGU	post-gonococcal urethritis
PI	post-infection
pI	isoelectric point
PID	pelvic inflammatory disease
PLV	Psittacosis lymphogranuloma venereum
Pmp	polymorphic membrane protein
ppm	parts per million
Q-Tof	quadrupole time of flight
QqQ	triple quadrupole
RB	Reticulate body
RP	reverse phase
SCX	strong cation exchange
SDS	sodium dodecyl sulphate
SILAC	stable isotope labelling in amino acid cell culture
STD	sexually transmitted disease
TCA	tricarboxylic acid
TCEP	tris-(2-carboxyethyl) phosphine
TEAB	triethylammonium bicarbonate
TFA	trifluoroacetic acid
Tof	time of flight
TRIC	Trachoma inclusion conjunctivitis
TTSS	type three secretion system
UPLC	ultra performance liquid chromatography
UV	ultraviolet
WHO	World Health Organisation

CHAPTER 1

GENERAL INTRODUCTION

1. GENERAL INTRODUCTION

1.0 Introduction

1.1 Infection and Immunity – the global prospective

From the common cold to tuberculosis, malaria and AIDS, infectious diseases are found all over the world and in many cases present a life-threatening risk. Infectious disease remains the world's No. 1 cause of death, with around 17 million people dying every year – almost one third of all deaths worldwide. A major contributing factor to these deaths is the lack of appropriate vaccines and treatments against these diseases. However, the idea that these deaths are only a consequence of poor socioeconomic conditions and as such are restricted to third world countries would be incorrect. For example, over the past decade the increasing resistance of pathogens to antibiotics has become a major cause for concern. The Centre for Disease Control (Klevens *et al*., 2002) estimates that each year, nearly 2 million people in the United States acquire an infection while in a hospital, resulting in 99,000 deaths. As life expectancy increases and we spend longer in our old age with a weakened immune system, there is a high probability we will fall ill to an infectious disease. As well as direct mortality and morbidity, research indicates that there are many indirect connections between infections and diseases such as asthma, rheumatic arthritis, allergies, and certain forms of diabetes. Bacteria, viruses and parasites are also linked with triggering cancer and diseases of the heart. It is therefore quite clear that finding effective treatments and prevention regimes against infectious diseases provides an enormous and important challenge for science.

The chronic ocular disease trachoma, caused by the bacterial pathogen *Chlamydia trachomatis* is one of the oldest and commonest infectious diseases known to mankind, dating back several thousand years and being first documented as early as the pharaonic era in Egypt (Ebers & Stern, 1875). Trachoma is the leading cause of preventable blindness and represents a major public health problem. The causative agent is the bacterial pathogen *Chlamydia trachomatis*. The World Health Organisation (WHO) estimates that 146 million people have trachoma due to ocular infection by *C. trachomatis* serovars A to C and that 4.9 million of these are totally blind (Whitcher *et al.*, 2001). In some parts of the developing world, over 90% of the population are infected and is attributed to poor socioeconomic environment where limited access to water and sanitation leads to poor personal and environmental hygiene. Where improvements in the socioeconomic status of a population in an endemic country have been made, trachoma can and has been eliminated.

Although trachoma has been eliminated from the western world, Chlamydial infection also represents the most common form of curable sexually transmitted disease (STD) in the UK and the United States. Since 1995, there has been a sustained increase in STDs with over 448 million new infections each year (WHO, 2005). Indeed, genital chlamydial infections in the UK have almost reached epidemic proportions with ~215,000 cases reported in 2010 (Health Protection Agency, 2010). This represents the most reported

sexually transmitted disease, overtaking genital warts, and is the first time that a bacterial infection has held such a title.

This thesis describes the development and application of a potentially powerful technology – proteomics – to the study of *C. trachomatis*. Before describing the technology and how it has been used in this area to date, it is pertinent to review certain aspects of the biology of *Chlamydia*, in order to give a more rounded perspective on the studies. Thus, the following sections describe the classification of these organisms, a synopsis of what is known about their physiology and developmental cycles, and an overview of their pathogenic properties, before going on to discuss areas of more immediate relevance such as chlamydial genomics and proteomics.

1.2 The genus *Chlamydia*

The virus-like obligate intracellular life cycle of chlamydiae and factors such as their inability to grow on conventional media has caused much confusion in the terminology that surrounds them and their grouping taxonomically. The genus *Chlamydia* is defined by many distinct characteristics and these are shown in **Table 1.1**.

Table 1.1. Defining characteristics of the genus *Chlamydia* (Moulder, 1984; Everett *et al*.,1999).

- Obligate intracellular habitat
- Unique developmental cycle
- Gram negative envelope without peptidoglycan
- Genus specific lipopolysaccharide
- Patches of hexagonally arrayed cylindrical projections
- Utilisation of host ATP for protein synthesis
- Small genome
- Glycogen producing

1.2.1 Classification

Chlamydia has been known by a number of names including *Ehrlichia, Chlamdoza, Rickettsiaformis, Bedsonia and Colesiota* and by terms such as PLV (psittacosis-lymphogranuloma venereum) and TRIC (Trachoma-inclusion-conjunctivitis). In 1966, Moulder (1966) reported the bacterial nature of chlamydiae. Shortly after, Page (1966) established the genus *Chlamydia* and later created two species in the genus *Chlamydia*: *C. trachomatis* & *C. psittaci* (Page, 1968), representing the group A (inclusions of chlamydiae containing glycogen) and group B (non-glycogen containing inclusions of chlamydiae) groupings respectively, of Gordon & Quan (1965). Based on observational differences of their life cycles and

metabolism, Storz and Page (1971) created the separate order *Chlamydiales,* moving from the order *Rickettsiales*. The development of DNA-based classification methods during the 1980s provided tools for distinguishing species (>70% homology) and genera (>20% homology). Using these data led to the creation of two additional species: *Chlamydia pneumoniae* in 1989 (Grayston,1989) and *Chlamydia percorum* in 1992 (Fukushi & Hirai, 1992), creating a total of four species, *C. trachomatis*, *C. psittaci*, *C. pneumoniae* and *C. percorum.*

In 1999, a new taxonomy was proposed (Everett *et al*., 1999) introducing more genera and species based on phylogenetic relationships. This requires members of the order *Chlamydiales* to be obligate intracellular bacteria with a chlamydia-like developmental cycle of replication and have >80% rDNA sequence identity with *Chlamydiales* 16S rRNA genes and/or 23S rRNA genes. Included in this order are the families *Chlamydiaceae*, *Skimkaniacea* and *Parachlamydiacea*. To be identified as a member of the *Chlamydiaceae* family, new *chlamydia*-like isolates require >90% identity with the 16S rRNA gene of the strain B577 from *Chlamydiaceae*. This increased the number of species in the family *Chlamydiaceae* to nine and grouped these species into two genera, *Chlamydia* and *Chlamydophila.* Two new species, *Chlamydia muridarum* and *Chlamydia suis* joined *Chlamydia trachomatis* and *Chlamydophila* acquired the current species *Chlamydia psittaci*, *Chlamydia pneumoniae* and *Chlamydia percorum* to form *Chlamydophila psittaci*, *Chlamydophila pneumoniae* and *Chlamydophila percorum*, respectively. For both genera, new species are required to be ≥ 97% identical. Three new *Chlamydophila* species are derived from *C. psittaci*: *Chlamydophila abortus*, *Chlamydophila caviae* and *Chlamydophila felis*.

Uptake of the new taxonomy has not been generally accepted by the scientific community and it has been suggested that it causes confusion. A diagram of the relationship between the old and the new system is represented in **Figure 1.1**. Discussions within this thesis will refer to the emended taxonomy.

Prior to the emended nomenclature, *C. trachomatis* was divided into three biovars (biological variants), closely related trachoma, Lympho-Granuloma Venereum (LGV) and the antigenically distinct murine biovar. Under the emended nomenclature, only strains from human biovars of trachoma and LGV are retained in *Chlamydia trachomatis.* These have been further subdivided into 19 serovars (Batteiger, 1996). Serovars A-K, Ba, Da, Ia, Ja are known as the trachoma biovar and are largely confined to infections of mucosal epithelial surfaces. Serovars A to C cause the hyperendemic ocular disease trachoma and Serovars D to K are associated with sexually transmitted genitor-urinary tract infections, conjunctivitis and neonatal pneumonia. The LGV biovar comprises of serovars L1, L2, L2a, L2b and L3, which are more invasive than disease caused by the urogenital serovars (D-K). LGV serovars infect predominantly monocytes and macrophages and pass through the epithelial surface to regional lymph nodes, often resulting in systemic infection.

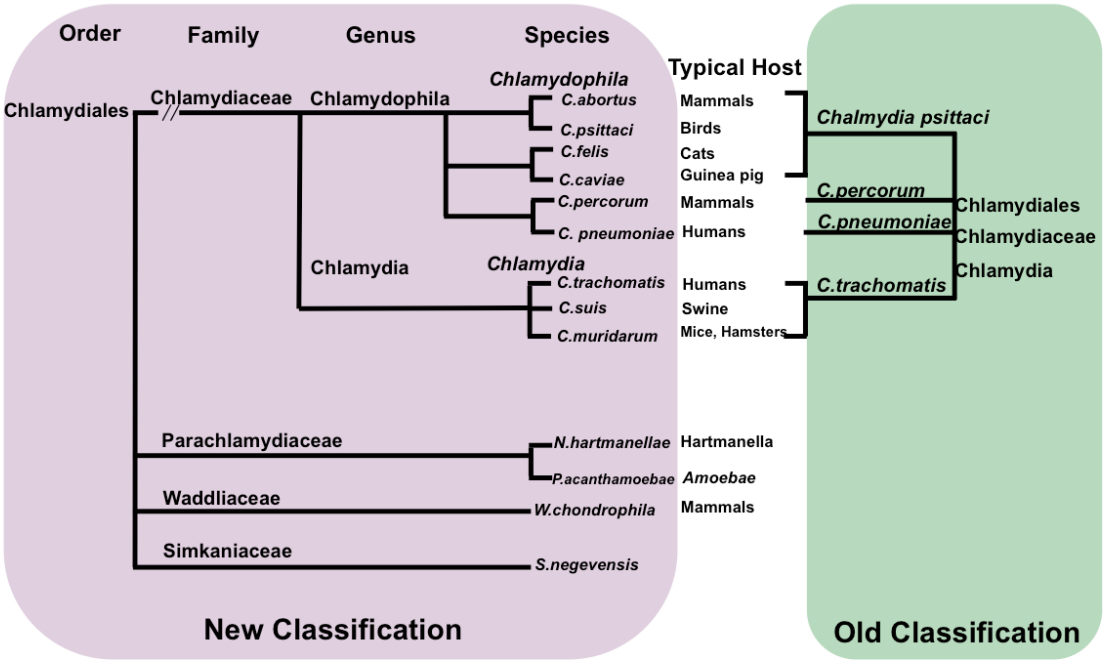


Figure 1.1. Comparison of the old and new classification nomenclature for the order Chlamydiales

The current classification of *C. trachomatis* is based on serological differentiation of antigenic epitopes on the Major Outer Membrane Protein (MOMP), encoded by the gene *ompA* (Wang and Grayston, 1991a; Wang and Grayston, 1991b; Wang *et al*., 1985). According to the amino acid sequence homology of these epitopes, these serovars have been placed in the following serogroups or class: B class (B, Ba, D, Da, E, L1, L2 and L2a), C class (A, C, H, I, Ia, J, Ja, K, and L3) and intermediate class (F and G) (Yuan *et al*.,1989).

The perceived limited diversity of chlamydiae as closely related obligate intracellular bacteria causing a wide range of disease in human and animal hosts has changed in recent years with the isolation of several 'chlamydia-like' organisms. These include endosymbionts of free-living amoebae (Michel *et al.,* 1994; Fritsche, 1993), which have been grouped in the family *Parachlamydiaceae*; a bacterium growing in 'cytoplasmic phagosomes' contaminating a cell culture (Kahane, 1993), which has been grouped in the family *Skimikaniacea;* and an isolate from an aborted bovine foetus (Dilbeck *et al*.,1990; Kocan *et al*., 1990) which was placed in a newly proposed family *Waddliaceae,* within the order *Chlamydiales* (Rurangirwa *et al*., 1999). The families' *Parachlamydiaceae, Simkaniaceae and Waddliaceae* are sister taxons to *Chlamydiaceae* because they have a Chlamydia-like cycle of replication and their ribosomal genes are 80-90% identical to ribosomal genes in the *Chlamydiaceae.* The type genera are *Parachlamydia* and *Simkania*, respectively, and new members of these genera should have 16S rRNA or 23S rRNA genes that are 95% identical to the type species *Parachlamydia*, *Acanthamoebae* and *Skimkania negevenisis,* respectively.

1.3 Physiology and the developmental cycle

Chlamydiales differ from the other main order of intracellular bacteria, the *Rickettsiales*, in their characteristic unique biphasic growth cycle, which has been extensively described (Ward, 1988; Moulder, 1991; reviewed by AbdelRahman and Belland, 2005). A representation of the developmental cycle of *Chlamydia* is shown in **Figure 1.2**. The cycle starts with the attachment of a small infectious, extracellular form, the elementary body (EB) to the host cell membrane. Through an invagination of the host cell membrane, the EB is ingested by the cell and becomes enclosed within an inclusion. These small EBs subsequently undergo differentiation into the larger obligate intracellular replicating form, the reticulate body (RB). These are considered to be metabolically active, but non-infectious; in contrast, infectious EBs are considered metabolically inactive. Cell division then occurs prior to the second differentiation stage, where RBs are transformed to EBs, which are subsequently released from the cell, and so the infectious cycle continues.

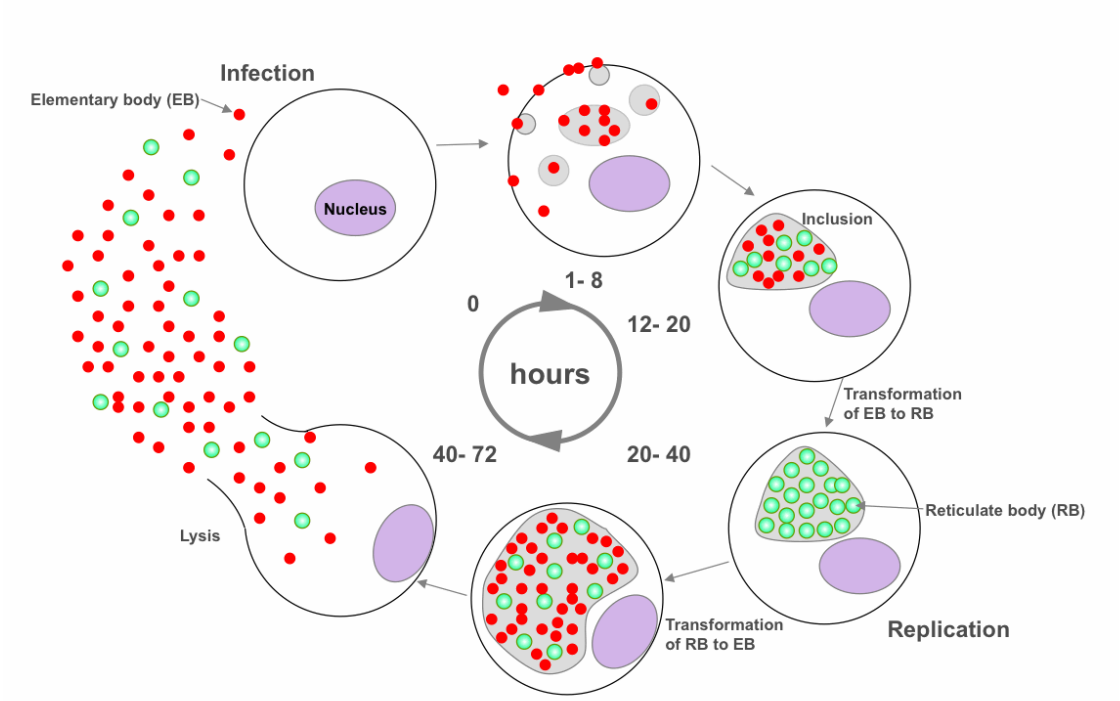


Figure 1.2. Graphical representation of the developmental cycle of *C. trachomatis*

1.3.1 Attachment

The first and most critical stage in the chlamydial development cycle is attachment of EBs to the host cell surface. The intracellular nature of chlamydial development and its specificity for its restricted host and cell types argues for the existence of a specific adhesion mechanism, although currently the receptors are not clearly defined. Isoelectric focusing of whole cells has shown that EBs of both the LGV and trachoma serovars of *C. trachomatis* carry a negative surface charge at physiological pH (Kraaipoel and Duin, 1979); similarly, the host cell also carries a net negative surface charge. Penetration of this electrostatic barrier is

therefore necessary to facilitate attachment of EBs to the host cell. Evidence supporting this requirement is provided by positively charged macromolecules such as DEAE-dextran or poly-L-lysine, which enhance attachment and inclusion formation of the trachoma, but interestingly, are not required for LGV strains (Kuo *et al*., 1976; Kuo *et al*., 1973). Additionally, for in vitro studies, it is necessary to centrifuge isolates from the trachoma biovar onto the surface of host cells such as Buffalo Green Monkey Kidney (BGMK) cells to achieve adequate infectivity, whereas, this is not required to achieve good infectivity of the LGV biovar. Although it is unclear why LGV strains are more efficient at attaching to the host cell, one possible explanation is that while there are no marked differences in net surface charge between the two biovars, there may be differences in charge at the local or molecular levels.

There are most likely multiple adhesins on the EB surface for interaction with the host cell (Campbell et al., 2006). The Major Outer Membrane Protein (MOMP) has been proposed as a potential adhesin (Su *et al*., 1990) and two proteins of 18 and 32 kDa in *C. trachomatis* have been shown to bind to host cells, depending upon serotype (Wenman and Meuser, 1986). Heparin, which inhibits the attachment of *Chlamydia* to host cells, has also been shown to affect these proteins. MOMP and these two smaller proteins have been identified as glycoproteins (Swanson and Kuo, 1990) with N-linked glycans (Swanson and Kuo, 1991a; Swanson and Kuo, 1991b). These structures, which rarely occur in bacteria, have been demonstrated to play an essential role in attachment and entry (infectivity) to the host cell (Kuo *et al*.,1996). OmcB from *C. pneumoniae*, a major component of the outer membrane complex, has also been shown using a yeast adhesion display system to adhere to HEp-2 epithelial cells. Further, blocking of the protein OmcB on the surface of *C. pneumoniae* EBs using anti-OmcB antibodies, reduced infection of HEp-2 cells by 60% (Moelleken and Hegemann, 2008). Similarly, the polymorphic membrane protein PmpD, has also been proposed to function as an adhesin (Wehrl *et al*., 2004).

1.3.2 Entry

Following attachment, uptake occurs such that the EBs are ingested by the host cell. Ingestion can occur via two routes: (i) the microfilament independent uptake of chalmydiae into clathrin-coated vesicles, or (ii) by the microfilament and energy-dependent process of phagocytosis or via both mechanisms (reviewed by Ward, 1988). This is dependent upon important factors such as strain, host cell and method of inoculation. Escalante-Ochoa, (2000) showed that, although *Chlamydophila psittaci* can use the microfilament-dependent and -independent entry pathways in both cell types, internalization and development of *Chlamydophila* in fibroblast cells mainly concerned processes mediated by microfilaments. In contrast, in epithelial cells, the internalisation mechanisms predominantly involved microtubule motor proteins.

A number of studies have demonstrated that the cytoskeleton of the host cell can be remodelled by microbial pathogens to facilitate their efficient uptake by eukaryotic cells. For example, enteropathogenic *Escherichia coli (E. coli)* recruit actin to form pedestal-like structures for attachment (Francis *et al*., 1991) and *Salmonella* induces membrane ruffles for internalisation (Finlay *et al*., 1991; Francis *et al*., 1993). A similar entry mechanism was demonstrated by Carabeo *et al*. (2002) in *C. trachomatis.* Attachment of EBs to the host cell elicited a localised concentration of actin, resulting in distinct microvillar reorganisation through

the cell surface, and the formation of pedestal-like structures at the site of attachment. Interestingly, a mutant cell line unable to support internalisation by *C. trachomatis* serovar L2 did not induce microvillar hypertrophy. In contrast, serovar D was able to be internalised, producing the pedestal-like structures. It is uncertain whether this cell line is defective in a specific receptor as it has not yet been fully characterised (Carabeo and Hackstadt, 2001). However, other studies have demonstrated that serovars L2 and D do not compete for attachment, indicating that different receptors may be involved (Davis and Wyrick, 1997). Although there maybe distinct receptors for different serovars, similar signals appear to be relayed to the host as both serovars D and L2 elicit microvillar hypertrophy in normal cell lines. These results are consistent with those involving receptor-mediated phagocytosis of *C. trachomatis* EBs by epithelial cells. A key protein involved in this cytoskeletal remodelling is the protein TARP (translocated actin recruiting phosphoprotein), a predicted type III effector protein that is present in all chlamydial species. Once translocated into the host cell, TARP induces rapid formation and polymerisation of actin filaments promoting EB entry (Clifton *et al*., 2004; Jewett *et al*., 2006). However, a consensus exists that *Chalmydia* is likely to use multiple mechanisms for entry (Moulder, 1991).

For survival within the host cell, chlamydiae at a very early stage (1-3hrs) must avoid the host cell's defence mechanisms. This is accomplished in part by separating themselves from the endocytic vesicular trafficking pathway. The bacteria ensure that no endocytic markers are accumulated on or within the chlamydial inclusion membrane (CIM) and hence they inhibit fusion with host cell lysosomes. While avoiding fusion with lysosomes, the chlamydial inclusion expands by fusion with sphingomyelin-containing lipids derived from the Golgi apparatus (Hackstadt *et al*., 1995; Hackstadt *et al*., 1997; Scidmore *et al*., 1996). Additionally, other glycerophospholipids and lipids from the Golgi, mitochondria, endoplasmic reticulum and lipid droplets are modified and incorporated into the inclusion membrane (Wylie *et al*., 1997; Cocchiaro *et al*., 2008). By intercepting this lipid traffic, chlamydiae are able to mimick the phospholipid composition of the host cell. This confirms observations that the phospholipid composition of chlamydiae, are closer to those found in eukaryotic cells than in prokaryotes (Wylie *et al*., 1997).

There are several encoded proteins that are associated with the CIM (Bannantine *et al*., 1998a; Bannantine *et al*., 1998b; Rockey *et al*., 1995; Belland *et al*., 2003). These and the discovery of a contact-dependent type III secretion (TTSS) pathway, similar to those found in other bacterial systems (Hsia *et al*., 1997; Stephens *et al*.,1998; reviewed by Beeckman and Vanrompay, 2010), represent an important development in the understanding of the host-pathogen relationship. The probable function of the type III secretion system is to modify host processes that may be necessary for host invasion, remodelling of the inclusion membrane, or affecting regulatory or host pathways. Already in the late 70s and early 80s, Gregory *et al*. (1979) and Matsumoto *et al*. (1976) observed the outer cell wall of the EB as having a hexagonal array of cylindrical projections of approximately 5 nm in size with 'rosettes' or fine holes. It was shown that one end of each projection was anchored into the cytoplasmic membrane while the other end of the projection protruded beyond the cell wall into the cytoplasm. These were later suggested to represent visual confirmation of a TTSS (Fields *et al*., 2003).

Rockey and co-workers have isolated and characterised several of the CIM associated proteins (Bannantine *et al*., 1998a; Bannantine *et al*., 1998b; Rockey *et al*., 1995; Scidmore-carlson *et al*., 1999). These proteins, Inc proteins, have been shown to be candidates for secretion by the TTSS (Subtil *et al*., 2001) and it is implied that they are secreted within the inclusion or through the CIM mediating communications with the cytosol. One particular protein, IncA, located on the cytoplasmic side of the CIM, undergoes phosphorylation by host cell kinases (Rockey *et al*., 1997) and has been proposed to be involved in subverting the signal transduction pathways in the host cell, to the advantage of the pathogen (Bavoil *et al*., 2000). However, a recent study demonstrates another role for IncA. IncA-laden fibres of *C. trachomatis* are used to create secondary inclusions (**Figure 1.3**), which are subsequently used to transport chlamydiae to these secondary inclusions and so actively expand the environment for development (Suchland *et al*., 2005). The Inc proteins are characterized by a hydrophobic bilobed motif containing 50-80 amino acid residues (reviewed by Kostryukova *et al*., 2008). Interestingly, although only about 10 have been shown to be localised to the inclusion membrane, Bannantine *et al*. (2000) identified through screening all predicted protein sequences within genomes of *C. trachomatis* and *C. Pneumoniae*, over 40 such sequences. As such, these chlamydial specific proteins are likely to play a vital role in the CIM development.

Expansion of this inclusion displaces the host cell nucleus and within two hours EBs are localised to the peri-nuclear region of the host cell. Dynein, part of the microtubule-organising centre has been implicated in the localisation of EBs to the peri-nuclear region because it co-localises with chlamydial early inclusions. Clausen *et al.* (1997) demonstrated that sodium vanadate, an inhibitor of dynein and tyrosine kinases, is detrimental to chlamydial development.

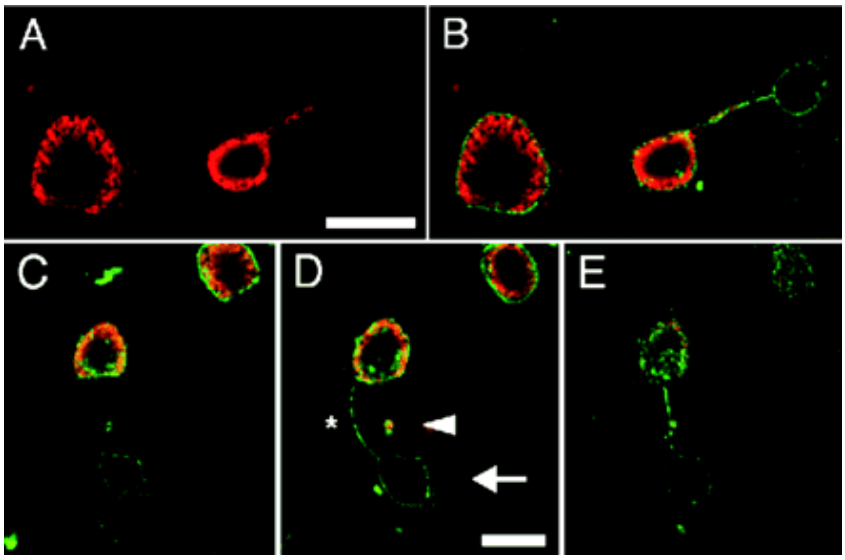


Figure 1.3. Chlamydial developmental forms within fibers linking primary and secondary inclusions.
C. trachomatis G/UW-57-infected HeLa cells (MOI, 0.1) were fixed 28 h PI and labeled with anti-IncA (green) and anti-MOMP (red). Confocal microscopy was then used to identify and document the presence of chlamydial developmental forms within IncA-laden fibers. (A and B) Two infected cells are shown, one

containing an inclusion lacking a fiber and one containing primary and secondary inclusions connected by a fiber. (A) The MOMP labeling identifies RBs that are found within the primary inclusion and within a fiber pointing to the right. (B) IncA and MOMP labeling within the same cells, demonstrating that the RBs outside the primary inclusion are present within an IncA-laden fiber. (C to E) A serial confocal Z-section showing an inclusion and associated fibers. (D) The secondary inclusion (arrow) is connected to the primary inclusion with at least two IncA-laden fibers (asterisk), and an RB (arrowhead) is found within one fiber. Scale bar, 10 μm. Reproduced with permission of the American Society for Microbiology (License number 2841450471563).

1.3.3 Differentiation, replication and lysis

Shortly after ingestion, the main differentiation process is initiated. The processes that trigger this intracellular interconversion of EBs to RBs are unknown. One proposal is that although protein synthesis in extracellular EBs is undetectable, initial translation of early phase proteins may be directed from stable transcripts present in infectious EB forms. Further, control of protein synthesis is through the existence of a series of promoters that are functional at specific times during the developmental life cycle (Plaunt and Hatch, 1988). At approximately 9 hrs the process of differentiation from EBs to RBs reaches its maximum. EBs lose their rigidity through the reduction of disulphide-linked MOMP to its monomeric form (Hatch *et al*., 1986), and are transformed to the RB phenotype, increasing in size from 0.3 μm to 1.0 μm and becoming flexible and fragile in nature (Hackstadt *et al*., 1985). Infection, ingestion and initiation of EB reorganisation occurs in minutes and reorganisation of EB to RB is completed within 8-24hrs.

Division of the RB occurs by binary fission at approximately 2-3 hours per generation and is complete in *C. trachomatis* by 24 hrs post infection, at which time, some RBs have commenced conversion to EBs while others continue dividing. The factors that initiate conversion of RBs to EBs are unclear and conflicting. Evidence suggests that growth of chlamydiae within the inclusion depends on their interaction with the CIM and consequently on TTSS activity in the area of contact. Contact-dependent replication of RB may ultimately lead to overcrowding and detachment resulting in TTSS inactivation. This may be part of a signalling cascade leading to the initiation of late differentiation (Hackstadt *et al*., 1997; Bavoil and Hsia, 1998; Bavoil *et al*., 2000). Although this may conflict with the previously proposed temporally regulated developmental cycle hypothesis (Plaunt and Hatch, 1988), it is consistent with the lack of synchronicity observed in chlamydial biology, i.e. whilst the development within the inclusion is relatively synchronous early on, it becomes asynchronous at later times.

At around 18-22 hrs post infection, RBs reorganise into EBs. Through several divisions of intermediate forms they undergo a reduction in size, genomic DNA condenses with the aid of histone-like proteins and sulphydryls of the bacterial cysteine-rich outer membrane proteins are crosslinked forming a rigid cell wall (Hatch *et al*., 1986). Eventually, infectious EBs and amorphous non-infectious intermediate forms are released either by lysis (*C. Psittaci*) or the inclusion is extruded by reverse endocytosis (*C. trachomatis and C. pneumoniae*). The resulting inclusions may contain 100 - 500 progeny that are free to infect more cells.

1.4 Human diseases induced by *Chlamydia*

1.4.1 Chlamydia trachomatis infections

1.4.1.2 Ocular infections – Trachoma

Trachoma, derived from the ancient Greek word meaning 'rough eye', is one of the oldest known human diseases and is the leading infectious cause of ocular morbidity. The disease is a chronic keratoconjunctivitis caused by the ocular serovars A, B, Ba and C of *Chlamydia trachomatis*.

Once a worldwide disease, trachoma is now confined to developing countries. The World Health Organisation (WHO) estimates that despite long-established control efforts, more than 500 million people are at high risk of infection (Polack *et al*., 2005). Recent estimates are that over 40 million people are infected, and about 1.3 million are blind in Africa, the Middle East, Central and Southeast Asia, and Latin America (Burton and Mabey, 2009). Generally, trachoma is associated with poor personal and environmental hygiene with transmission mainly by contact, although eye-seeking flies and gnats are also considered a major factor in the spread of the disease. It occurs most frequently in children with those in endemic areas typically being infected before 2 years of age (Dawson *et al*., 1976).

The disease begins with a mucopurulent conjunctivitis, developing into a follicular keratoconjunctivitis. Repeated episodes of infection cause superficial vascularization of the cornea (pannus) and conjunctival scarring which increase with the severity and duration of the disease. Severe scarring of the conjunctiva distorts the eyelid, a condition called entropion. This causes the eyelashes to turn into the eye (trichiasis) causing abrasion of the orb of the eye resulting in ulceration, secondary infection with other bacteria and ultimately scarring causing opacity and blindness (Jones, 1975).

1.4.1.2.3 Ocular infections – Paratrachoma

Clinically indistinguishable from trachoma, the genital strains of *C. trachomatis* serovars D to K ocassionally cause a chronic follicular conjunctivitis. This type of infection, termed paratrachoma or adult inclusion conjunctivitis follows a similar course of disease to trachoma, including follicular conjunctivitis with pannus and often conjunctival scarring. However, they are not associated with chronic disease, are not sight-threatening and are rarely associated with stable transmission cycles within a given community. Occurring generally in sexually active young adults (50% of men with paratrachoma also have a genital infection) (Darougar *et al*., 1972), it is usually transferred by the hands from the infected site to the eyes. It has an incubation period of 1 to 2 weeks, is usually self-limiting and is mono-ocular.

1.4.1.3 Genital tract infections

Chlamydia trachomatis causes more cases of sexually transmitted diseases (STDs) than any other bacterial pathogen, and make gential infections caused by *Chlamydia* a major public health problem. In the UK it is the most common Sexually Transmitted Infection (STI), with 215,501 diagnoses in Genito-Urinary Medicine clinics in 2010 (Health Protection Agency, 2010), and in the United States it is the most commonly reported notifiable disease (Centre for Disease Control and prevention, 2010). Highest rates are seen in young people, especially men and women under 24 years. Of the *C. trachomatis* genital tract infecting serovars D to K, serovars D to G are the most common (Oriel and Ridgway, 1982; Bax *et al*., 2011).

1.4.1.4.1 Genital tract infections in men

Non-gonococcal urethritis (NGU) is caused by *C. trachomatis* in 30 – 50% of cases (Horner *et al*., 2005). These can be asymptomatic, but in symptomatic patients it manifests itself as a mucopurulent discharge with dysuria. Post-gonococcal urethritis (PGU) is defined as persistent NGU following treatment of a gonoccocal urethritis infection. This is attributed to a mixed infection of *N. gonorrhoeae* and an organism responsible for NGU such as *C .trachomatis* (Oriel and Ridgway, 1982). There is also supporting evidence for the reactivation of latent genital chlamydial infection by concurrent gonorrhoea infections (Batteiger *et al*., 1989).

C. trachomatis is responsible for 30 - 60% of cases of epididymitis in men less than 30 - 35 years of age, while *N. gonorrhoeae* is the second most common organism in such cases (Oriel, 1992). Inflammation of the reproductive tubes that carry sperm from the testis (epididymitis) presents as pain in the affected testicle; when severe, the condition may be accompanied by abdominal pain, fever and malaise. In chronic epididymitis, blockage of the epididymis may occur, causing infertility due to reduced sperm numbers. Occasionally, a scrotal mass is observed, mimicking a testicular malignancy (Ward *et al*., 1999). Although proctitis is generally attributed to infection with LGV, other chlamydial serovars have been isolated from anoreceptive intercourse-practising homosexual men (Munday and Taylor-Robinson, 1983).

1.4.1.4.2 Genital tract infections of women

Cervicitis

The two most regularly identified pathogens that may cause inflammation of the cervix (cervicitis) are gonococci and *C. trachomatis*. Presentation of cervicitis includes: pain passing urine, soreness, cervical discharge, swelling and erythema. Occasionally, lymphoid follicles similar to those found in trachoma are observed. Approximately 80 percent of infections are asymptomatic; however, if left untreated, infection can spread causing pelvic inflammatory disease (PID).

It has also been recognised that chlamydial and other infections may play a role in the aetiology of cervical and other cancers (Quirk and Kupinski, 2001). In particular, large serological studies have suggested evidence linking *C. trachomatis* infection with an associated risk of developing cervical cancer (Wallin *et al*., 2002; Lehtinen *et al*., 2011).

Endometritis and Salpingitis

Forty percent of women with untreated chlamydia go on to develop Pelvic Inflammatory Disease (PID), a general term referring to infection of the upper genital tract. Sequelae include, infection of the fallopian tube (salpingitis), uterine lining (endometritis), ovaries and surrounding tissues. Clinical manifestations range from asymptomatic endometritis to severe symptomatic salpingitis with mucopurulent discharge (Paavonen, 1998). Acute and chronic salpingitis leads to fibrosis and scarring creating major complications. These include chronic pelvic pain, damage to the fallopian tubes, affecting egg transport leading to ectopic pregnancy or causing blockage preventing transport and fertilization of the egg, leading to infertility (Brunham *et al*., 1986; Kosseim and Brunham, 1986; Barrett and Taylor, 2005).

1.4.1.5 Lymphogranuloma venereum (LGV)

The lympho-granuloma venereum (LGV) biovar comprises serovars L1 to L3. Although more uncommon, they are more invasive than those caused by the urogenital serovars (D-K) and transmission is almost exclusively by sexual contact. LGV is endemic in the West Indies, Africa, India, Southeast Asia, and the Caribbean. Although uncommon in Europe and North America, a new LGV variant designated L2b was thought to be responsible for outbreaks in the Netherlands, neighbouring European countries, the United Kingdom and the United States, causing proctitis rather than genital ulceration and the typical inguinal buboes (Nieuwenhuis *et al*., 2003; Spaargaren *et al*., 2005a; Spaargaren *et al*., 2005b). Observed variations in the gene used to genotype *Chlamydia* strains, *ompA,* supported speculation of a new variant representing a newly emerging infection. Recently, the genomes of both the L2 and the 'epidemic' LGV isolate (L2b) have been sequenced. Comparison of these genomes found no additional genes in the L2b strain, however, there was some evidence of gene loss and regions of heightened variation, which have previously been important sites for inter-strain variation. It has therefore been suggested that this is unlikely to be a newly emerged strain, but is an old strain with relatively new clinical manifestations (Thomson *et al*., 2008). More recently, deep genome sequencing of a clinical isolate revealed a recombinant of the L2 and D strain. Regions of genetic exchange in this strain (L_2c) included the toxin gene, which interestingly, after ~20 passages in laboratory culture was lost. Deletions and insertions were also observed for the genes *ftsK*, *tarp*, *hctB,* whose functions are associated with replication, chlamydial host cell entry and DNA compaction in the later stages of the chlamydial developmental cycle. These findings are likely to have important implications for the evolution and emergence of new chlamydial strains, highlighting the importance of applying new high-throughput sequencing technologies to understand the phenotypes of such clinical recombinants and their influence on pathogenicity (Somboonna *et al*., 2011).

LGV has a specific tropism for lymphoid cells and spreads through the inguinal lymph nodes becoming systemic. There are three stages of infection. The first stage normally begins after an incubation period of 1 - 3 weeks with a small painless lesion that appears at the site of infection, normally the penis or the vagina. This may be accompanied by headache and fever. After 2 to 6 months, progression to the second stage is indicated by lymphadenopathy. The lymph nodes become enlarged 'buboes' that eventually rupture and drain causing variable courses of disease. If left untreated, the third stage of infection can lead to fistulas, stictures, genital elephantiasis, frozen pelvis, and infertility (Mabey and Peeling, 2002).

1.4.1.6 Neonatal infections

Chlamydial exposure through the birth canal leads to 60% of infants born acquiring infection (reviewed by Schachter and Grossman, 1981). About one third of those infants exposed develop neonatal conjunctivitis, while approximately one in six develops pneumonia.

Neonatal conjunctivitis caused by *C. trachomatis* has an incubation period of around 10-14 days causing inflammation and mucopurulent discharge. The condition is not considered to be sight-threatening and is self-limiting, resolving itself within a few months without treatment. Scarring of the cornea is uncommon, although in more severe cases some keratitis and micropannus can develop. Infection sites can also include the nasopharyx, middle ear, trachea, lungs, rectum and vagina with no obvious disease pathology. In untreated cases, 10 – 20% of neonates will go on to develop *C. trachomatis* pneumonia (Beem and Saxon, 1977). Most infants do not require hospital treatment, but in severe cases hospitalization may be required.

1.4.2 *Chlamydophila* infections in humans

1.4.2.1 *Chlamydophila pneumoniae*

Chlamydophila pneumoniae, originally called the TWAR strain from the name of the original two isolates. (Taiwan (TW-183), isolated from an ocular site and an acute respiratory isolate designated AR-39) is now considered a separate species of *Chlamydia* (Grayston *et al*., 1989) and are described within the genus *Chlamydophila* under the revised nomenclature (Everett *et al*., 1999).

Infecting the mucosal surfaces of the respiratory tract, *C. pneumoniae* displays a wide range of disease. It is one of the most common human pathogens and causes acute infections such as pneumonia, bronchitis, sinusitis and pharyngitis. It is estimated that *C. pneumoniae* is the cause of 10% of community-acquired pneumonia cases and 5% of bronchitis and sinusitis cases in the United States (Kuo *et al*., 1995a). Infection is most prevalent in children from 5 -14 years of age and 50 to 70% of the adult population worldwide is seropositive for this pathogen (Grayston, 1992). However, since the antibody response is time-limited to 3-5 years, infection may occur several times during a persons lifetime (Kuo *et al*., 1995a).

Respiratory infections caused by *C. pneumoniae* are often mild with no symptoms or only symptoms of extended cough (Miyashita *et al*., 2003) and in most cases they probably remain undiagnosed. Nevertheless, some infections cause severe illness and may cause pneumonia and bronchitis. *C. pneuomoniae*, has so far only been shown to infect humans and transmission appears to be via person- -person, although this route appears to be relatively inefficient (Grayston *et al*., 1993).

Chlamydial infections that are asymptomatic and thereby untreated may become persistent and lead to chronic conditions. The association of *C. pneumoniae* with chronic human diseases was first shown in sero-epidemiological studies, which demonstrated the association of antibodies to *C. pneumoniae* with acute myocardial infarction (Saikku *et al*., 1988; Arcari *et al*., 2005). Further, links with other chronic diseases such as asthma, chronic bronchitis and chronic obstructive pulmonary disease have also been suggested to be associated with *C. pneumoniae* infection, based upon sero-epidemiological evidence (Hahn *et al*., 1991; Azzouzi *et al*., 2005). An association of *C. pneumoniae* with coronary artery disease has also been supported by detection of either antigens by immunohistochemistry or nucleic acids of the organism by the polymerase chain reaction in the affected tissues (Kuo *et al*., 1993a; Kuo *et a*l.,1993b; Kuo *et al*., 1995b; Ouchi *et al*., 1998; Leinonen and Saikku, 2002; Belland *et al*., 2004).

1.4.2.3 *Chlamydophila psittaci*

Chlamydophila psittaci, previously *Chalmydia psittaci,* is the causative agent of psittacosis (parrot fever). Although the disease was first found in the psittacine, the natural reservoir for *C. psittaci* occurs in a wide range of domesticated and feral birds, but is also incidentally transmissible to humans. Historically, psittacosis was used to describe the infection in psittacine birds and ornithosis in feral birds. Both have been shown to cause the same infections (Page, 1966) and so the use of the term Chlamydiosis for chlamydial infections in all species including birds, animals and humans has become universally accepted.

Chlamydial infection in birds is both a major economic problem and an occupational health hazard for those working in the poultry and pet bird industries. Infection in birds occurs primarily in the gastro-intestinal tract with transmission of infection by the shedding of chlamydiae in the faeces or by discharge from the respiratory tract. Exposure to dead or living infected birds or droppings also provides a method of transmission among all the aforementioned species. Person- -person transmission can occur, but is rare.

The most common human *C. psittaci* infections cause acute respiratory tract infections that may have systemic manifestations. Presentation is extremely variable - some patients having only a mild cough while others can develop severe or fatal pneumonia. The incubation period of the disease is highly variable between 1 to 4 weeks with presentations of fever, chills and headache in both mild and severe disease states. The alternative more severe disease state has similarities to typhoid, general feverish state without respiratory involvement and bradycardia. Primary diagnosis is often made after confirming the existence of bird exposure (Schachter, 1988). *C. psittaci* diseases can show surprising persistence, lasting up to ten years with very few symptoms (Meyer and Eddie, 1951).

1.4.2.4 Other human infections

There are several reports of human abortion, or severe respiratory disease in non-pregnant humans caused by the ovine abortion strain *Chlamydophila abortus,* previously from the taxon *C. psittaci* (Mare, 1994). Generally, reported cases are during the lambing season when there is a higher risk of exposure to chlamydiae from aborting ovine foetuses. Chronic conjunctivitis caused by infection with feline *C. felis* has also been reported (Hartley *et al*., 2001).

1.4.2.5 Non-human infections

Chlamydia has been isolated from a large number of feral and domesticated mammals, as well as many avian species and frogs (reviewed by Storz, 1988). Diseases include ocular disorders, pneumonias, arthritis, abortion, intestinal infection, meningitis, mastitis and hepatitis (reviewed by Storz, 1988). Of particular importance is *Chlamydophila abortus,* a major cause of abortion in sheep, cattle and goats. In Britain it is estimated to cost the sheep industry in excess of £20 million per annum (Aitken *et al*., 1990).

1.4.2.6 Persistent infections

Although not clearly understood, *Chlamydia* can persist in a viable, but non-cultivable growth stage resulting in a long-term relationship with the host cell (reviewed by Beatty *et al*., 1994a; Wyrick, 2010). For instance, the ability of *C. pneumoniae* to cause persistent respiratory infections in humans is well-documented (Hammerschlag *et al*., 1992; Miyashita *et al*., 1996). Distinguishing between this persistent state and a re-infection causing chronic disease presents difficulties. Nonetheless, multiple lines of evidence support persistent infections both in vitro and in vivo.

Abnormal chlamydial development in vitro has been induced by a number of different methods including antibiotic treatment, low nutrient induction, cytokine exposure (particularly IFN-γ), induction by phage, during monocyte infections and when maintained in continuous culture (reviewed by Hogan *et al*., 2004). These result in persistent states which show many similarities to each other such as the development of small inclusions, loss of infectivity, enlarged aberrant RB that do not undergo binary fission and inhibition of differentiation at different stages, i.e. EB to RB or RB to EB. When inhibitory growth pressure is removed, the normal developmental cycle is often resumed. The persistent state of this well-adapted pathogen may represent an important mechanism for intracellular survival (Hogan *et al*., 2004). It is likely that differences in persistent forms between strains may have subtly evolved to enhance the pathogens survival to specific environmental cues and different host niches.

Fehlner-Gardiner (2002) and Caldwell (2003) provided an explanation for specific tissue tropisms or niches based upon indole-rescuable phenotypes in *C. trachomatis*. Genital serovars (D - K and L1 and L3), but not ocular (A - C) serovars growing in tryptophan-deficient medium, are able to synthesize tryptophan

from exogenous indole and therefore recover their normal infective cycle. Interestingly, tryptophan depletion by induction of the tryptophan degrading enzyme indoleamine 2,3-dioxygenase (IDO) by IFN-γ has been shown to be the major mechanism for IFN-γ-mediated persistence in *C. trachomatis serovar* A and *C. pneumoniae* A-03 in aortic smooth muscle cells (Beatty *et al*., 1994b; Pantoja *et al*., 2000). These observations together suggest the persistence of a chlamydial infection in vivo maybe governed by the interplay between IFN-γ mediated tryptophan depletion and other factors such as the availability of exogenous indole, the strains ability to synthesis tryptophan, the susceptibility of specific strains to IFN-γ or other cytokines and IDO induction levels. Adding to this complexity, it is likely that there will be many other processes in addition to tryptophan catabolism, such as iron depletion, the nitric oxide synthase effector pathway, that could also be attributable to IFN-γ. Taking into account the wide host range of *Chlamydia*, the importance of each mechanism and their interplay, it is likely that persistent states will vary between species and reflect their requirements for survival within a specific niche.

1.5 Diagnosis of chlamydial infections

Confirmation of a chlamydial infection depends upon obtaining a clinical sample for a suitable laboratory diagnostic test. Historically, intrusive procedures that may not be acceptable to people who are asymptomatic have been necessary to obtain the required cervical and urethral specimens. The development of new detection methods has facilitated the use of urine samples, which are more acceptable.

1.5.1 Cell culture

For many years, the method of choice for confirming a chlamydial infection was the inoculation of clinical material into cell culture. A variety of cell lines including McCoy cells, mouse L929, BHK21 (Baby hamster kidney), HeLa and BGMK (Buffalo Green Monkey Kidney) cells have been used (Hobson *et al*., 1982). Infected cells are examined for the presence of the characteristic chlamydial inclusion bodies by either staining with Iodine, staining with Giemsa stain, fluorochrome-labelled poly- or mono-clonal antibody or by enzyme immunohistochemistry.

1.5.2 Immunofluorescence

Direct detection of chlamydial elementary bodies using immunofluorescence provided the first commercially available test. This test was based upon a fluorescein-conjugated, species specific monoclonal antibody to the major outer membrane protein of *C. trachomatis*. When compared to tests carried out in cell culture, good results were obtained with values of 96.3% sensitivity and 99.5% specificity (Uyeda *et al*., 1984). Other manufacturers subsequently produced similar tests based upon monoclonal antibodies of lipopolysaccharide (LPS) extending the methodology to the detection of all *Chlamydophila* species. Although the test is relatively fast, it requires careful examination by skilled microscopists and is subjective.

1.5.4 Immunoassays

Introduction of a range of enzyme-linked immunosorbant assays (ELISA) based upon chlamydial antigens and cognate monoclonal antibodies provided a less subjective automatable test. For example, NovaLisa™ is a diagnostic ELISA assay for the qualitative measurement of IgA, IgG and IgM antibodies against *C. trachomatis*.

1.5.5 Nucleic acid amplification based tests

Nucleic acid amplification-based methods are now of prime importance for the diagnosis of chlamydial infections (Lisby *et al*., 1999) and have been considered the most important advance in chlamydial detection since cell culture (Stary *et al*., 1998).

The polymerase chain reaction (PCR) or ligase chain reaction (LCR), are highly sensitive and specific methods for the detection of *Chlamydia* using multiple-copy gene targets, e.g. those on the cryptic chlamydial plasmid. Detection of *C. trachomatis* in genital specimens of infected symptomatic and asymptomatic men and women has been increased by 30% using these types of amplification technologies (Dille *et al*.,1993; Schachter *et al*.,1994; Ostergaard *et al*.,1990). The true sensitivity of PCR and LCR is of the order of 90 to 97% (Cheng *et al*., 2001). Improved performance in terms of reproducibility and suitability for automation has enabled testing in populations of low prevalence, thereby providing a method that informs infection control strategies. However, new rRNA-based tests have been reported to offer further increases in sensitivity, allowing the detection of infections previously missed by PCR (Yang *et al*., 2007).

1.5.5.1 Serological detection of antibody

Serological tests used for the detection of current or prior exposure to chlamydial infections, or for comparisons between infected sample populations, are clearly useful. There are wide variations between various tests and their correlation to anti-chlamydial titre. So, while a single serum sample showing raised anti-chlamydial titre maybe useful for diagnosing entrenched infections (Taylor-Robinson, 1992), it may not indicate an active infection, as elevated levels can exist in the absence of the pathogen. The considered Gold Standard of serological diagnosis of *C. trachomatis* infections has been the serotype-specific micro-immunofluorescence test (MIF) (Wang and Grayston, 1991b), providing classification of *C. trachomatis* into 15 serotypes. However, the observation of cross reactivity with *C. pneumoniae* leading to false positives (Gijsen *et al*., 2001) in combination with other issues such as lack of objectivity, reproducibility has limited the suitability of this assay for routine use. An evaluation of two species-specific enzyme immunoassays (EIA) against MIF and were found to be less laborious, more specific and less expensive (Bax *et al*., 2003). Although, serological tests fall short of providing diagnoses on an individual basis, their high negative predictive value may be of use in identifying patients where *C. trachomatis* is unlikely to play a role in infection.

1.6 Chlamydial genomics

The restricted and obligate intracellular nature of chlamydiae and until recently, the lack of any gene transfer system (Wang *et al*., 2011), has hindered the understanding of chlamydiae at both the cellular and molecular level. Therefore, the sequencing of several chlamydial genomes has been very important in furthering *Chlamydia* research.

Fortunately, the size of the genomes of chlamydiae are relatively small (~ 1 Mb) and so sequencing of these genomes is easily achieved. At the initiation of these studies, eight genomes representing five strains of chlamydiaceae had been sequenced, annotated and published (Stephens *et al*.,1998; Kalman *et al*., 1999; Read *et al*., 2000; Read *et al*., 2003; Shirai *et al*., 2000; Carlson *et al*., 2005; Thomson *et al*., 2005). However, recent advances in high-throughput genome sequencing technologies now facilitate the rapid sequencing of many isolates from different strains. Therefore, for the practical purposes of this thesis, an overview of these sequences representing a single strain and serovar is shown in **Table 1.2**. Two isolates of *C. trachomatis* serovar B have also been sequenced, but are not contained within the table since they are yet to be published. Draft sequences of these genomes can be obtained at the following URL:http://www.sanger.ac.uk/sequencing/Chlamydia/trachomatis/CTA_CTB/). *C. trachomatis,* serovar A, D and *C. muridarum* also contained a plasmid, and in *C. caviae* and *C. pneumoniae* AR39 a bacteriophage. These were also sequenced and are associated with the relevant genome sequences. The primary description of these genome sequences has provided thorough and valuable information that has expanded our previous perspectives of chlamydiae.

Table 1.2 Genome sizes of sequenced chlamydiaceae

Genome	Size (Mb)	Plasmid/Phage	Reference
C. trachomatis, D/UW-3/CX	1.042	7493 bp plasmid	Stephens *et al*., 1998
C. pneumoniae, CWL029	1.230		Kalman *et al*., 1999
C. pneumoniae, AR39	1.234	4524 bp phage	Read *et al*., 2000
C. pneumoniae, J138	1.226		Shirai *et al*., 2000
C. muridarum MoPn	1.080	7501 bp plasmid	Read *et al*., 2000
C. caviae (GPIC)	1.173	7966 bp phage	Read *et al*., 2003
C. trachomatis, A/Har-13	1.051	7510 bp plasmid	Carlson *et al.,* 2005
C. abortus	1.144		Thomson *et al*., 2005
C. trachomatis, L2/434/Bu	1.039	7499 bp plasmid	Thomson *et al*., 2008
C.trachomatis L2/UCH-1/proctitis	1.039	7499 bp plasmid	Thomson *et al*., 2008
C. trachomatis E/11023	1.043		Jeffrey *et al*., 2010
C. trachomatis F/70	1.048		Jeffrey *et al*., 2010
C. trachomatis G/9301	1.043		Jeffrey *et al*., 2010
C. trachomatis J/6276	1.043		Jeffrey *et al*., 2010

1.6.1 Genome comparison

Different strains of chlamydiae share similarities in terms of their unique biphasic developmental cycle and morphology. However, as previously discussed, they show differences in tissue tropism, infectivity

and inclusion morphology. The varying levels of DNA homology between strains, suggests that comparative analyses of chlamydial genomes may reveal differences. These differences can be used to formulate hypotheses to explain the observed differences in pathogenicity and tissue/host specificity.

Seven hundred and ninety eight genes were conserved in all the published chlamydiae sequences prior to the sequencing of the *C. abortus* genome in 2005, representing the possible minimal set of genes required for the development and intracellular survival of *Chlamydia*. One hundred and eighty three of these genes could not be found in any of the 70 other microbial organisms published in the TIGR database (Read *et al*., 2003).

Comparative analysis of the *C. caviae* genome (Read *et al*., 2003) showed that 68 out of 1009 *C. caviae* genes were not found in any of the other chlamydial genome sequences up to this date. Similarly, comparative analysis of *C. pneumoniae* contained 168 genes of unknown function (Read *et al*., 2003). These niche-specific genes are not likely to be required for primary functions and are more likely to be necessary for virulence and survival of chlamydiae in specific sites or hosts. *C. caviae* possesses an almost complete tryptophan biosynthesis operon, indicating the ability of *C. caviae,* to synthesise tryptophan from the early precursor anthranilate. In comparison, *C. trachomatis* only possess a limited set of these genes (Stephens *et al*.,1998). Ocular-gential and LGV serovars are able to synthesise tryptophan from indole, whereas the ocular serovars A to C contain a truncated *trpA* gene and serovar B lacks the *trpA* operon and hence are unable to synthesis tryptophan (Shaw *et al*., 2000b), confirming the findings of Fehlner-Gardinier *et al*. (2002) and Caldwell *et al*. (2003). These niche-specific genes are also absent in the sequenced strains of *C. pneumoniae* suggesting they are also unable to synthesise tryptophan (Kalman *et al*., 1999). Differences in tryptophan requirements and sensitivity to IFN-γ –mediated growth inhibition (Beatty *et al*., 1994b) are likely to reflect some of the strain differences observed in host specificity, virulence, persistence and transmission of chlamydiae.

A tox gene, similar to cytotoxic genes from enterobacteria, is also present in *C. caviae* with orthologues in *C. muridarum*. It has been suggested that these products may be secreted by the type III secretion system, in order to inhibit actin polymerisation (Read *et al*., 2003). Truncated ORFs of these genes were found in both the L2 and D serovars, but no orthologues were identified in *C. pneumoniae* (Read *et al*., 2003; Thomson *et al*., 2008). Again, the absence or presence of these niche-specific genes may account for the different sites of infection between chlamydiae as they may influence trafficking and ultimately affect the degree of systemic infection (Belland *et al*., 2001).

Eighty percent of the predicted proteins from *C. pneumoniae* have a corresponding orthologue in *C. trachomatis* (Kalman *et al*., 1999). A major difference between these two species is the increased number of genes encoding for the polymorphic membrane proteins (*pmps*) similar to those found in *C. psittaci*, increasing from nine *pmp* genes in *C. trachomatis* to 21 in *C. pneumoniae*. Several of the *pmp* genes contain frame shifts that vary between isolates (Grimwood and Stephens, 1999). The *C. pneumoniae* genomes are more than 99.9% identical and the small differences are mainly found in *pmps* and the polymorphic protein

(ppp) family of *C. pneumoniae* (Daugaard *et al*., 2001; Rocha *et al*., 2002). *Pmp* genes represent a major area of variability between strains. The fact that chlamydiae with such a small genome has maintained these paralogs, suggests that they play a pivotal role in structural, functional or antigenic polymorphism.

The size of all the *C. trachomatis* genomes sequenced to date are similar in size. Comparison of the *C. trachomatis* serovars D/UW-3/CX, A/Har-13 and L2/434/Bu (Stephens *et al*., 1998; Carlson *et al*., 2005; Thomson *et al*., 2008), have 846 genes common to all three genomes. Differences in the coding sequences between these genomes could be accounted for by *in-silico* prediction differences or as a result of functional loss, indicating that differences in disease aetiology was not attributable to gene acquisition (Thomson *et al*., 2008).

1.6.2 Highlights of chlamydial genome sequencing studies

For the first time, genomic data has provided a putative list of gene products that informs us of the metabolic capabilities of chlamydiaceae. The functional assignment of these genes cannot however be taken for granted. Assignments are made on the basis of sequence homology, the levels of which can be somewhat arbitrary. Chlamydiaceae lineage is deeply separated from that of other eubacteria and so many of their proteins show only low levels of homology with known proteins. Nonetheless, when supported by biochemical characterisation, confident functional assignments can be made. The vast number of genes and their products makes discussion of each gene product in turn an impractical task. Therefore, the following discussion focuses on key subsets of these genes.

In terms of central metabolism, the pathways that appear to be present in *C. trachomatis* suggest that it is an aerobic organism utilising glutamate as the primary carbon source with glucose and oxoglutarate playing supplementary roles during different stages of the development cycle. Genes encoding the proteins for an intact glycolytic pathway were identified, although controversy concerning the presence of the predicted enzyme fructose-1,6-diphosphate aldolase has led to the proposal of an alternative route via the pentose-phosphate pathway (also called the hexose monophosphate shunt) to circumvent this enzyme (Stephens *et al*., 1998). An incomplete tricarboxylic acid (TCA) cycle was identified. In light of the absence of the genes encoding the enzymes citrate synthase, aconitase, and isocitrate dehydrogenase, it has been suggested that the pathway operates via an aspartate shunt. No genes encoding glyoxylate-bypass enzymes were identified, indicating that chlamydiae are unable to use fatty acids or acetate as carbon sources (Stephens *et al*., 1998). Consistent with the observation of accumulation and storage of glycogen within chlamydiae, a complete glycogen synthesis and degradation system was identified, supporting a role for glucose as a carbon source at different stages of the developmental cycle. Genes encoding essential functions in aerobic respiration were also well represented in the chlamydial genome (Stephens *et al*., 1998).

Traditionally, chlamydiae have been considered ‘energy parasites’ obtaining ATP from their host cells (Moulder, 1991; Hatch *et al*., 1982). Two ATP transporting proteins with sequence identity to those from *Rickettsia prowazekii* were identified in both the *C. trachomatis* and *C. pneumoniae* genomes supporting this accepted hypothesis (Andersson, 1998). It was therefore surprising when genes encoding a wide range of

ATPases as well as phosphoglycerate kinase, pyruvate kinase, and succinate thiokinase were also identified via genomics. This finding suggests that chalmydia may have the capability to produce ATP themselves and hence may not be strict ATP auxotrophs (Stephens *et al*., 1998). However, it remains to be seen if these compounds are actually expressed at the protein level and if so when. The ability to synthesise ATP autonomously may be important when *Chlamydia* is unable to obtain ATP from the host cell such as in the early and late stages of development (Hatch *et al*., 1982).

Several groups of chlamydial encoded proteins can be considered of special interest. These include (i) proteins involved in making peptidoglycan (ii) type III secretion proteins; (iii) inclusion membrane proteins (incs); and (iv) the polymorphic membrane proteins (pmps) previously discussed.

It has previously been proposed that chlamydiae lack peptidoglycan because muramic acid, one of the major components has not been biochemically detected in any significant amounts (Fox *et al*., 1990). The presence of nearly a full set of genes involved in peptidoglycan synthesis was therefore unexpected. Peptidoglycan has also been suggested to play a role in the division of RB. This, combined with the sensitivity of *Chlamydia* to cyclo serine and beta-lactam antibiotics such as penicillin (Chopra *et al*., 1998), strongly supports the hypothesis that chlamydiae synthesise peptidoglycan or a peptidoglycan–like component (Stephens *et al*., 1998).

Orthologous genes that are predicted to encode a type III secretion system were also found in their entirety, first in *C. caviae* and subsequently in *C. trachomatis* (Hsia *et al*., 1997; Stephens *et al*., 1998). The TTSS is found in many gram-negative bacteria and their genes are typically linked with 'pathogenicity islands' with a relatively high A+T content. The genes encoding the chlamydial TTSSs have an A+T content similar to the rest of the genome and are found in three loci on the chromosome (Stephens *et al*., 1998). The TTSS in *Chalmydia* is likely to play a role in modifying the host cell processes that may be necessary for host cell invasion, restructuring of the inclusion membrane, or affecting host cell regulatory pathways (reviewed by Beeckman and Vanrompay, 2010). As previously noted, surface projections on both RB and EB have been observed using electron microscopy (Matsumoto, 1982). Originally thought to be involved in nutrient uptake before the discovery of the TTSS genes, these projections are now considered to be type III needles (Bavoil and Hsia, 1998; Hatch, 1998).

Homologous proteins associated with the chlamydial inclusion membrane termed inclusion membrane proteins (incs), have been identified in all sequenced chlamydiae, but have not been found in any other sequenced organism. The first of these were identified in *C. psittaci* and termed incA, B and C (Rockey *et al*.,1995; Bannantine *et al*., 1998a). Several incs have now been identified and all of these have a characteristic bilobed hydrophobic region. In total, 33 genes encoding proteins with this hydrophobicity pattern have been identified in the *C. trachomatis* genome and 93 in the *C. pneumoniae* CWL029 genome (reviewed by Vandahl *et al*., 2004; Kostryukova *et al*., 2008). The likely role of this class of protein is in inclusion membrane remodelling and transport. This was demonstrated by Suchland *et al,* (2005), who

showed that IncA was involved in the expansion of the inclusion membrane by creating secondary inclusions and transporting chlamydiae to those inclusions.

1.6.3 Genetic transformation

Towards the end of these studies, a genetic transformation system in *C. trachomatis* was developed and reported, representing a major advance for the field of *Chlamydia* research (Wang *et al*., 2011). The developed system used a constructed shuttle vector based on the 7.5 kb chlamydial plasmid of *C. trachomatis* L2/434/BU and the *E.coli* plasmid pBR325, with transformation achieved by simple calcium chloride treatment of EBs. Selection of transformants was achieved by taking advantage of the penicillin-induced RB phenotype, which forms non-dividing aberrant RBs, where further development to the EB form is arrested. Genetically stable, penicillin resistant transformants are selected from the aberrant RBs by the isolation of penicillin resistant EBs over successive passages. The author's demonstrate this genetic transformation system by producing a penicillin resistant *C. trachomatis* strain, expressing the green fluorescent protein within chlamydial inclusions. They also substantiate the associated link that the chlamydial plasmid is necessary for the synthesis and accumulation of glycogen in chlamydiae. By transforming the chlamydial plasmid into a non-glycogen producing, plasmid-free strain of *C. trachomatis* L2, glycogen production and accumulation was restored, confirming a role for the plasmid in glycogen biosynthesis. If this transformation system can be applied routinely, it will represent an important milestone and has the potential to advance our understanding of chlamydial pathogenesis substantially.

1.7 Transcriptional profiling of *Chlamydia*

Comparative sequence analysis of genomes can help formulate hypotheses based upon the presence or absence of a particular gene or group of genes. However, it is likely that a substantial fraction of a chlamydial genome is differentially regulated and co-ordinately transcribed providing critical functional and regulatory roles and mediating the phenotypic changes observed during the developmental cycle (Nicholson *et al*., 2003). The development of microarray hybridisation techniques (reviewed by Lander, 1999; Brown and Botstein, 1999) provides a technology able to measure the relative levels of mRNA between two cellular states. Advances in microarray technology (alternatively termed transcriptional profiling) have made this a popular technique for global gene expression analysis.

Two elegant studies have used transcriptional profiling to analyse two *C. trachomatis* strains, serovar D and L2 (Nicholson *et al*., 2003; Belland *et al*., 2003). These studies have provided powerful new insights into the temporal expression of genes that control the different stages of the developmental cycle and those that determine the nature of the host-pathogen interaction.

3.2% of the chlamydiae genome (29 genes) was transcriptionally active as early as 1 hour post-infection (PI). By 3 h PI, an additional 200 genes were transcriptionally active. This corresponded to

microscopic observations of the differentiation of EB to RB by chromatin decondensation. At 8 h PI there was significant amounts of intense transcriptional activity, which was maintained through the period of the developmental cycle when RB replicate by binary fission. During this period (16-24 h PI), almost every gene was transcribed highlighting the fact that chlamydiae, with its small genome, has very little facultative capacity. During the late stages of development, a subset of 28 genes were specifically expressed (Belland *et al*., 2003). Since genes were classified according to the moment their transcript was observed, and owing to some asynchronicity in the cycle, the comparison of gene transcripts expressed very early or late in the cycle will be the most informative.

1.7.1 Immediate - Early genes

Among the Immediate -Early genes, seven had been previously described. These included the first early stage protein to be cloned and sequenced called the early upstream ORF gene (Wichlan and Hatch, 1993), a family of inc genes (*inc*D, E, F and G) and the chaperonin genes *gro*EL and *gro*ES (Shaw *et al*., 2000a). Newly identified immediate early genes included ADP/ATP translocase, nucleotide phosphate transporter, oligopeptide permease, a D-alanine/glycine permease, malate dehydrogenase, methionine aminopeptidases and nucleoside phosphohydrolase. These components are involved in translocation and interconversion of metabolites within the bacterial cell. At 1 h PI two further inc proteins were transcribed. The number of inc-like proteins predicted to be expressed reiterates the important role that these proteins must play in modifying the chlamydial inclusion membrane to support growth and survival.

A novel predicted 162 kDa transcript encoded by CT147 (CT denotes the *gene* prefix *Chlamydia trachomatis,* serovar D) with a significant level of homology to the early endosomal antigen 1 (EEA1) was confirmed by immunofluorescence and radiolabelling to colocalise with the inclusion membrane. EEA1 is involved in endosomal trafficking and fusion in mammalian cells (Christoforidis *et al*., 1999; Mu *et al*., 1995; Simonsen *et al*., 1998). The transcript of CT147 was first detected at 1 h PI, reaching a maximum expression at 8 h PI. Expression of the protein was first detected at 8 h PI using immunofluorescence where it was localised to the periphery of the chlamydiae inclusion and detected at 16 h PI using immunoprecipitation. The temporal difference in detection of protein expression was attributed to the different sensitivities of the two assays. Between 24 and 40 h PI, the protein disappears with the concomitant appearance of several lower molecular mass immunoreactive species, indicating post-translational modification of CT147 by either a chlamydial or a host protease (Belland *et al*., 2003).

1.7.2 Late genes

Twenty six genes were expressed in the late stages of development, including previously characterised genes such as those encoding the histone-like proteins HctA and HctB, which mediate chromosomal condensation in the differentiation of RB to EB (Brickman *et al*., 1993; Barry *et al*., 1993). The secondary differentiation process (RB to EB) is characterised by the formation of a highly disulphide cross-linked outer membrane (OM) complex. Two late genes encoding for cysteine rich OM proteins (*omcA* and *omcB*) and the major outer membrane protein (OmpA) form this OM complex. It is proposed that these extensive intra- and inter- disulphide linkages are formed by two thioredoxin disulfide isomerases (CT780 and CT783), also

expressed late in the cycle. The expression of two highly conserved membrane thio-proteases may also have a proteolytic function in the formation and maturation of this OM complex. These two proteases show homology to adenoviral proteases that have been shown to play a role in the maturation of virus particles (Greber, 1998).

The application of microarray technology to study the temporal gene expression of chlamydiae has proved to be invaluable. The identification of a number expressed genes with novel functions, has provided fresh insights into the biology of the chalmydiae and may even have provided new potential targets for therapeutic agents.

To date, however, it has been difficult to correlate the patterns of mRNA expression with the corresponding patterns of protein expression. This is due in large part to the technical difficulties associated with studying the latter. These difficulties, together with recent proteomic advances that overcome some of them, are described in the next section.

1.8 Proteome analysis

The completion of a genome sequence is frequently insufficient to determine the biological function of a gene. Measurement of the genes at the mRNA level can provide valuable information, as demonstrated by Belland *et al*., (2003). However, no simple correlation exists between changes in mRNA expression levels (transcriptomics) and those at the protein level (proteomics), with several studies reporting disparity between the relative expression of mRNAs and the corresponding expressed protein (Anderson and Seilhamer, 1997; Gygi *et al* 1999a; Cox *et al*., 2005). This is unsurprising in view of the fact that many regulatory processes impact on gene expression after transcription. For example, mRNAs may be translated with different efficiencies and may also have different half-lives. Similarly, proteins undergo different rates of degradation and may also require post-translational modifications in order to be active. However, by exploiting new technologies to measure some of these parameters more precisely, closer correlation between mRNA and protein expressions are observed (Schwanhausser *et al*., 2011).

The original dogma of one gene producing one protein, we know, no longer stands true. In eukaryotic cells, it has been suggested that each gene can specify an average of 6 to 8 proteins, although the precise number is uncertain (Strohman, 1994). The mechanisms responsible for this diversity include: alternative splice variants, post-translational modifications, proteolytic cleavage, differences in protein conformation, and changes in oligomeric state (protein-protein interactions). Different forms of a protein may reside in one or several cellular locations, each performing a specific biological function. Collectively, therefore, there are compelling reasons for undertaking direct analysis of proteins, even if the task is markedly more difficult than working at the mRNA level.

Proteomics can be defined as the systematic, large-scale analysis of proteins, protein-protein interactions and their post-translational modifications. As such, the ideal proteome study would provide the absolute quantitative measurement of every protein; its isoforms, modifications and complexes within a given sample. Although not routine in the majority of proteomic laboratories, the application of new proteomic technologies are beginning to achieve proteome coverage comparable to those of deep transcriptome sequencing technologies. For example, Nagaraj *et al*. (2011) quantified 9230 proteins encoded by 9207 genes in HeLa cells, matching nearly all those transcripts identified by deep transcriptome sequencing of the same sample.

There are two essential steps in a proteomic analysis: the separation of the proteins in a sample derived from tissue or cells and the identification of the proteins in that mixture. Historically, the high-resolution separation of complex mixtures of proteins has been performed using two-dimensional gel electrophoresis (2-DGE), in which hundreds or even thousands of proteins are separated orthogonally, according to their charges (pI) and molecular masses using polyacrylamide gels (O'Farrell, 1975). Once separated, the proteins in the gel are visualised by staining (e.g. with coomassie brilliant blue, silver stains or fluorescent dyes such as Sypro Ruby), and quantitated by image analysis to pinpoint the spots of interest. Spots of interest are then picked for protein identification. These days, the approaches used for the subsequent identification of these proteins are generally based on biological mass spectrometry. The current main approaches of protein separation, protein identification, modern mass spectrometry (MS) instrumentation and advancements in the context of proteomics will be discussed later; however, to maintain continuity, previous chlamydiae proteomic studies are first discussed.

1.8.1 Proteome analysis of Chlamydiae

1.8.1.1 Early *Chlamydia* proteome studies

Early proteome studies focused on the Chlamydial Outer Membrane Complex (COMC) from EBs. Batteiger (1985), compared lymphogranuloma venereum and ocular strains of *C. trachomatis* and identified three major outer membrane proteins (MOMP) within this complex, a protein of variable molecular mass, 42 – 45 kDa, (OmpA), a 60 kDa protein (OmcB) and a 12 kDa protein (OmcC). The 60 kDa protein in *C. trachomatis* L2 was found to be more basic and was only observed after analysis by non-equilibrium pH gradient electrophoresis. This provided the first evidence of structural differences between the LGV and ocular strains, supporting their classification into separate biovars.

Moroni *et al*. (1996) compared the COMC from different *C. trachomatis* serovars, as well as from *C. pneumoniae* and *C. caviae* using 2-DGE. OmcB from *C. trachomatis* L2 was resolved in the gels, but migrated one pH unit more basic than OmcB of *C. trachomatis* F, and two pH units more basic than that of *C. trachomatis* D. No further proteins were identified in these studies.

1.8.1.2 More recent *C. trachomatis* proteome studies

The first proteomic study of whole cell lysates of *C. trachomatis* was aimed at identifying the synthesis of early proteins from *C. trachomatis* serovar L2 using pulse-label experiments in combination with 2-DGE (Lundemose *et al*., 1990). Seven proteins were observed at 2 to 8 h PI. Four of these proteins were labelled at 2 to 4 h PI, three of which were identified using colocalisation with proteins detected by immunoblotting with known antibodies. These three proteins of 75, 62, and 45 kDa were identified as the S1 ribosomal protein, the GroEL-like protein, and DnaK-like protein, respectively. The remaining four proteins were not identified. Early transcription of the *groEL* gene was recently confirmed by transcriptional profiling (Shaw *et al*., 2000a; Slepenkin *et al*., 2003). Interestingly, the transcript for GroES, a 20 kDa protein that has been shown to be co-transcribed with GroEL in *C. psittaci* (Morrison *et al*., 1989) was not detected. This was probably attributable to its small size and hence, the reduced number of amino acids available for label incorporation, a limitation of these types of radiolabelling experiments.

Improved resolution of 2D gels attributed to immobilised pH gradient (IPG) strips enabled Bini *et al*. (1996) to generate one of the first reference maps for *C. trachomatis.* Approximately 600 spots were separated in silver stained gels. A combination of immunoblotting with known antibodies and N-terminal protein sequencing was used to identify nine known proteins. Seven sequences were obtained from yet uncharacterised proteins. These analyses were carried out prior to the availability of chlamydial genome sequences and hence information and technology for the identification of proteins was limited.

1.8.1.3 *C. pneumoniae* proteome studies

The first comprehensive map of *Chlamydia* following the publication of genome sequences for the chlamydiae was that of elementary bodies separated by 2-DGE (Vandahl *et al*., 2001). One hundred and sixty seven different proteins were identified by mass spectrometry from samples labelled with [^{35}S] methionine/cysteine, constituting 15% of the genome. Samples were pooled from different stages of the developmental cycle aimed at increasing proteome coverage. This qualitative study provided important new findings, including: the first report indicating the expression of the type III secretion system in EBs; 8 new pmps; confirmation of 31 out of 167 previously hypothetical proteins; validation of a high number of proteins involved in metabolism, transcription and translation (Vandahl *et al*., 2001).

A quantitative proteomic study using [^{35}S] pulse-labeling in combination with 2-DGE, examined the global protein expression profile of *C. pneumoniae* from 24 to 48 h PI during the transition from reticulate to elementary bodies. Of the 35 proteins identified in this study, 31 of these proteins increased in expression during the transition from RB to EB and only 4 proteins showed an observed decrease. These included proteins associated with amino acid and cofactor biosynthesis, maintenance of cytoplasmic function, modification of the bacterial cell surface and energy metabolism. The results of this study infer that metabolic pathways may be involved in the transition from RB to EB (Mukhopadhyay *et al*., 2006).

1.8.1.4 The chlamydial secretome

Orthologues of type I-IV secretion systems similar to those of other Gram-negative bacteria have been identified in the chlamydiae (Hsia *et al*., 1997; Stephens *et al*., 1998; Kalman *et al*., 1999), and the expression of several proteins from the type III secretion apparatus were observed in proteome studies of *C. pneumoniae* (Vandahl *et al*., 2001*)* and *C. trachomatis* (Shaw *et al*., 2002b*)*. However, the identification of chlamydial secreted proteins from the cytoplasm of host cells has been complicated by the fragility of the inclusion and RBs. Shaw *et al*. (2002b) compared the protein content of whole lysates of infected cells to purified chlamydiae using 2-DGE to identify chlamydial proteins that are found outside the bacterium. Using this subtractive proteomic approach resulted in the identification of two proteins, CT858 of *C. trachomatis* and Cpn1016 of *C. pneumoniae*. These were orthologues of the protein known as the *Chlamydia* protease-like activity factor (CPAF), which has been shown to down-regulate host cell transcription factors required for MHC class I and II presentation and subsequently confirmed to be secreted (Zhong *et al*., 2001a).

1.8.1.5 Further proteomic studies on the outer membrane complex

In addition to earlier studies, further proteome analysis of the COMC identified several proteins including a predicted membrane component of the TTSS, YscC, indicating that the apparatus is assembled in EBs and is membrane-associated in *Chalmydia* (Vandahl *et al*., 2002). Other type three III secretion components identified in the *C. pneumoniae* study were not observed in the COMC (Vandahl *et al*., 2001). In addition to the major outer membrane proteins (OmpA, OmcB, OmcC), other major constituents of the COMC included the Pmps, as characterised by Vandahl *et al*, (2002) using 2-DGE. Pmps have been confirmed to be part of the family of exported Gram-negative bacterial proteins designated autotransporters (Wehrl *et al*., 2004). Further confirmation of Pmps in the COMC came from Birkelund *et al*. (2009), identifying 7 out of the 9 *C. trachomatis* Pmps in the COMC (PmpB, C, D, E, F, G and H), using combined fractional diagonal chromatography (COFRADIC). In addition to previously identified proteins of the COMC, they also reported a putative membrane protein (CTL0541), a putative exported protein (CTL0887), a hypothetical protein (CTL0626) and the low calcium response protein D (SctV), a bacterial inner membrane component of the type III secretion system. Further *in-silico* analysis of the hypothetical protein CTL0626, identified a conserved domain motif for a carbohydrate-selective porin, OrpB. This motif was also found to be present in other sequenced chlamydial genomes. Further evidence of OrpB as a component of the COMC was provided using immunblotting and immunofluorescence microscopy. The COMC was more recently characterised by Liu *et al*. (2010) using differential proteomics, confirming the COMC proteins reported by Birkelund *et al*. (2009). The authors also identified an additional three previously unreported proteins, CTL0493 (Omp85), CTL0645 and a peptidoglycan associated lipoprotein (Pal), as probable components of the COMC.

1.8.1.6 Proteomic comparison of the serovars of *C. trachomatis*

2-DGE reference maps of EBs from *C. trachomatis* A, D and L2 were published in 2002 (Shaw *et al*., 2002a). The findings were similar to those reported in the study of *C. pneumoniae* with ~16% of the ORFs identified in the *C. trachomatis* serovars. From the 144 protein species identified in all serovars, 55 migrated

differently in serovars D and L2, 52 differed between A and L2 whereas only 26 differed between A and D. This reflects the greater similarity between the serovars A and D than between LGV, serovar L2 and A/D. Twenty five ORFs expressing hypothetical proteins were identified in *C. trachomatis* D including CT579, the orthologue of the chlamydial-like protease factor (CPAF) identified in the COMC study. Serovar specific differences in the pmp proteins were observed and may reflect differences in mechanisms of host cell attachment and/or chlamydial virulence between serovars. In particular, the detected N-terminal fragment of PmpF identified in L2 were not observed in the same area on *C. trachomatis* A or D gels indicating serovar-specific differences in the PmpF amino acid sequence, expression levels or processing (Shaw *et al*., 2002a). The majority of the differences observed between serovars are likely to be attributable to the substitution of charged amino acid with non-charged (or *vice versa*) and hence may not have major biological implications (Vandahl *et al*., 2004).

1.8.1.7 Summary of chlamydiae proteomic studies

Table 1.3 presents a summary of the chlamydiae proteomic studies discussed in this Chapter.

Table 1.3 Summary of proteomic studies of Chlamydiae

Study type	Sample type	Technique used	Chlamydial species	Serovar	Reference
Qualitative	COMC	2-DGE	*C. trachomatis*	L1, L2, L3 A, B, Ba, C to K	Batteiger *et al*., 1985
Quantitative	EB, RB	Pulse-labeling, 2-DGE	*C. trachomatis*	L2	Lundemose *et al*., 1990
Qualitative	COMC	Pulse-labeling, 2-DGE	*C. trachomatis* *C. pneumoniae* *C. psittaci*	L2, D, F IOL-207 6BC	Moroni *et al*., 1996
Qualitative	EB	2-DGE	*C. trachomatis*	L2	Bini *et al*., 1996
Quantitative	EB, RB (IFN-γ treatment)	Pulse-labeling, 2-DGE	*C. trachomatis*	A, L2	Shaw *et al*., 1999
Qualiitative	EB	1-DGE	*C. trachomatis*	L2	Mygind *et al*., 2000
Qualitative	EB	Pulse-labeling, 2-DGE	*C. pneumoniae*	VR1310	Vandahl *et al*., 2001
Qualitative	EB	2-DGE	*C. pneumoniae*	VR1310	Vandahl *et al*., 2002
Qualitative	EB	Pulse-labeling, 2-DGE	*C. trachomatis*	A, D, L2	Shaw *et al*., 2002a
Qualitative	Secreted proteins	Pulse-labeling, 2-DGE	*C. trachomatis* *C. pneumoniae*	A, D, L2 VR1310	Shaw *et al*., 2002b
Quantitative	EB, RB	Pulse-labeling, 2-DGE	*C. pneumoniae*	VR1452	Mukhopadhyay *et al*., 2006
Qualitative	COMC	COFRADIC	*C. trachomatis*	L2	Birkelund *et al*., 2009
Qualitative	COMC	LC-MS/MS	*C. trachomatis*	L2	Liu *et al*., 2010

1.9 Proteomic techniques

The complex nature of the proteome has presented researchers with major challenges in terms of its analysis, even for relatively simple expressed proteomes, such as those of microbes (O'Connor *et al*., 2000). These major challenges include suitable protein separation strategies, the determination of protein expression levels and the subsequent identification of those proteins of interest. The following section focuses on these major challenges and recent methods to resolve them using present day proteomics.

1.9.1 Protein separation strategies

1.9.1.1 Two-dimensional polyacrylamide gel electrophoresis

The classical method for quantitative and qualitative expression proteomics has combined high-resolution separation of proteins using two dimensional polyacrylamide gel electrophoresis (2-DGE) (O'Farrell, 1975; Klose, 1975; Gorg *et al*., 1988) with mass spectrometry (MS) or tandem mass spectrometry (MS/MS) approaches for the identification of the protein spots. As mentioned previously, 2-DGE can separate hundreds or even thousands of proteins orthogonally according to their charge (pI) in the first dimension and by their molecular masses in the second dimension.

In the first dimension, termed isoelectric focussing, proteins migrate through a pH gradient formed within a polyacrylamide gel until they reach their isoelectric point (the point at which their charge is the same as the surrounding pH). With the development of immobilised pH gradient (IPG) gels (Bjellqvist *et al*., 1982), in which the pH gradient is maintained by acrylamido buffers that are co-polymerised and so 'immobilised' into the gel, the overall sample to -sample reproducibility has greatly improved (Gorg *et al*.,1988).

In the second dimension, IPG strips containing the focussed proteins are soaked in a denaturing solution containing sodium dodecylsulphate (SDS). The strong negatively charged SDS binds to all of the proteins within the strip making them essentially have the same charge. This strip is placed at the cathode of a second polyacrylamide gel. When an electric field is applied to this second dimension gel, proteins migrate from the cathode (IPG strip) towards the anode. Although all the proteins move in the direction of the anode, smaller proteins move faster through the gel than larger ones, hence, proteins are separated according to their size. This second dimension separation is essentially the same as conventional one dimensional polyacrylamide gel electrophoresis (1-DGE), which is also used in many proteomic approaches for protein separation.

After separation, the proteins need to be visualised using a protein specific stain. There are a number of different stains available with a range of sensitivities and dynamic ranges. The most commonly used staining techniques are coomassie blue, silver staining, and fluorescent stains such as Sypro Ruby. For a detailed review on staining techniques, see Patton (2002). All proteins appear as spots on the gel indicating their location. The intensity of the spot can subsequently be used to quantify the amount of protein between

samples separated on different gels. This can usually be achieved using commercial gel image quantitation packages such as PDQuest (BioRad, Hercules, CA).

Because even the best 2-DGE can separate no more than 1500 proteins, the dynamic range of protein expression that can be measured is limited to only the most abundant proteins of a crude protein mixture. Relative to DNA microarrays, the number of induced proteins observed in 2-DGE can be underestimated by 2 to 4 fold (Eymann *et al*., 2002; Hommais *et al*., 2001; Conway and Schoolnik, 2003). It is also now evident that there are many other challenges faced when utilizing this technology with proteins that have extreme physico-chemical properties such as very acidic, very basic, very small, very large, or low abundance proteins (Gygi *et al*., 2000).

1.9.1.2 Multidimensional chromatography

Faced with the limitations of 2-DGE technology, there has been a major thrust to alternative technologies. One approach that has generated a notable amount of interest is multidimensional liquid chromatography (Link *et al*., 1999; Opiteck *et al*., 1997).

Multidimensional chromatography allows separation of complex mixtures by using multiple columns with different stationary phases. These columns are coupled orthogonally, which means that fractions from the first column can be selectively transferred to other columns for additional separation using an alternative physio-chemical principle. This enables separation of complex mixtures that cannot be separated using a single column. This technology coupled on-line to MS/MS, provides a powerful method for resolving and identifying peptides/proteins from complex mixtures and thus has also been termed multidimensional protein identification technology (MudPIT). An example of this was developed in the Yates laboratory using strong cation exchange (SCX) and reverse phase nanocapillary columns arranged in series to provide high resolution separation and concentration of tryptic peptides derived from protein samples, prior to their identification by MS/MS. In this study they were able to resolve and identify 1,484 proteins from 5540 peptides in bakers yeast (*S. cerevisiae*), in comparison to around 300 previously identified by 2-DGE, thus significantly improving proteome coverage. Included in these proteins were low abundance proteins, proteins with extreme pI and molecular mass, and integral membrane proteins (Washburn *et al*., 2001; Wolters *et al*., 2001). By combining two or even more orthogonal separations based on different physiochemical properties, samples of increasing complexity, over wide dynamic ranges can be resolved (Garbis *et al*., 2011).

1.9.1.3 GeLC-MS/MS

Another technically simpler alternative to both 2-DGE and MudPIT has been termed GeLC-MS/MS. Like MudPIT, significant gains in proteome coverage have been found using this technology by essentially boosting peak and load capacity. GeLC-MS/MS uses conventional 1-DGE for protein separation. Slices of the gel, containing fractionated proteins, are then subjected to in-gel digestion with a site-specific endo-protease such as trypsin, prior to further fractionation with nanocapillary LC columns and on-line peptide identification via tandem MS (Schirle *et al*., 2003). Although, the good separation ability of GeLC-MS/MS has been applied to relatively simple mixtures, it has also been demonstrated to be capable of separating

relatively complex mixtures by resolving and identifying 1289 plasmodium proteins (Lasonder *et al*., 2002). The technique has found particular application in qualitative proteomics, alternatively referred to as 'shotgun'proteomics' (Skipp and O'Connor, 2011).

Although not a true quantitative technique, several empirical indications can be used to estimate the relative abundance of a protein in a mixture using this technique. In general, the higher the amount of protein, the greater the MS signal intensity of their corresponding peptide ions. Also the number of sequenced peptides recovered for each protein and the number of spectra obtained for each peptide are proportional to the abundance of the protein in question. These are refered to as 'spectral counting' or 'exponentially modified peptide abundance index' (Rappsilber *et al*., 2002; Liu *et al*., 2004; Ishihama *et al*., 2005; Lu *et al*., 2007).

1.10 Protein Identification

In the past decade, a major limitation has been the lack of sensitive methods to unambiguously identify the separated proteins of interest. A number of methods for protein identification have been explored during this time and these are discussed below.

1.10.1 Co-localisation

In co-localisation, unknown proteins are identified by comparison to previously identified proteins for identification. For example, in 2-DGE, the molecular weights and pI's of previously identified proteins are compared to those of unknown proteins analysed under identical conditions. Proteins that migrate to the same position as the previously identified known proteins on the gel are considered to be the same proteins. Co-localisation is clearly limited in its use for large-scale protein identification experiments, but, in small focused analyses, it can play an important role.

1.10.2 Protein sequencing by the Edman degradation technique

Ideally, the ultimate protein identification would consist of the entire sequence of a protein including its post-translational modifications. Although modern N-terminal protein sequenators using Edman degradation (Edman and Begg, 1967) are able to assign the identity of a protein with sufficient confidence when >10 amino acid residues are obtained, they are limited to a maximum analysis of around 70 amino acid residues per peptide/protein and have high sample concentration requirements to achieve such coverage. Additionally, 80% of eukaryotic proteins cannot be N-terminally sequenced due to post-translational modifications at the N-terminus (Nokihara, 1998). To circumvent this problem of short read-lengths and N-terminal blocking, internal peptides can be sequenced following digestion of proteins with an endo-protease or by chemical cleavage and then peptide separation. Proteomic studies require sensitive, high throughput identification strategies whereas these types of approaches require significant (pmol) amounts of protein and are costly. Further, the determination of each amino acid residue using an N-terminal sequencer takes around 23 minutes Thus, for the 10 amino acids required for a confident assignment, approximately 4 hours per

peptide would be required. Sequence analysis via the C- terminus is even more difficult and requires more starting material.

1.10.3 Protein Mass and pI

Experimentally determined mass and/or the pIs of proteins from 2-DGE have been compared to theoretical values of proteins for identification. However, the masses and p*I*s of proteins determined by 2-DGE are of low accuracy making assignment of proteins based on these measurements impossible. In addition, post-translational modifications that are not indicated in the theoretical sequence will not be taken into account in the molecular mass and pI calculation, which further complicates identification. However, the use of mass as an identification feature in proteomics is appealing. Developments in the area of biological mass spectrometry (MS) now allow the measurement of intact proteins and their proteolytic fragments to a very high accuracy (Kelleher *et al*., 1999; Biemann and Papayannopoulos, 1994; Shevchenko *et al*., 1996) There has therefore been a major push to devise new techniques to exploit these improvements in mass spectrometry technology.

1.11 Biological Mass spectrometry

Advances in biological MS, together with similar advances in bioinformatics, have enabled the rapid, unambiguous identification of proteins that are present in the sub-picomolar range. MS is now the method of choice for protein identification and also forms the basis of many new approaches for the simultaneous quantitation of proteins within a biological sample. For clarity in the explanation of these MS-based approaches, it is first relevant to discuss the elements and types of instrumentation used in biological mass spectrometry.

A mass spectrometer consists of three essential elements:
1. An ionization source, which converts molecules in either solution or solid form into gas phase ions.
2. One or more mass analysers, to separate the gas phase ions according to their mass-to-charge ratios (*m/z*).
3. An ion detector to count the ions emerging from the last mass analyser.
Most mass spectrometers use the same type of ion detector, but use different ion sources and mass analysers. Conventionally a mass spectrometer is named after the type of its ion source and the type of mass analyser.

1.11.1 Ion sources

One of the most important advances in biological mass spectrometry has been the innovation of robust techniques for ionising biomolecules such as peptides, thereby allowing their efficient introduction into mass analysers for analysis. The two key techniques matrix-assisted laser desorption ionisation (MALDI) and electrospray ionisation (ESI) have revolutionised such analyses, allowing the analysis of proteins in excess of 1 MDa, and have been combined with a variety of mass analysers.

1.11.1.1 Matrix-assisted laser desorption ionization (MALDI)

MALDI (Karas and Hillenkamp, 1988; Beavis and Chait, 1996) is achieved in two steps. In the first step, the compound to be analysed is mixed in a solvent containing small organic molecules, called a matrix, and that has an absorbtion wavelength that closely matches that of the MALDI source laser. This mixture is then dried onto a surface to form a heterogeneous layer containing analyte embedded in matrix crystals. The second step involves ablation of bulk portions of this crystalline matrix by intense, short pulses of laser. The irradiation by the laser causes rapid heating of the crystals, which in turn causes local sublimation and expansion of the matrix into the gas phase. Ionisation reactions can occur at any time during this process, but the origin of ions in MALDI is still not fully understood (Zenobi and Knochenmuss, 1998). The rationale for the matrix is that it has a different absorbance wavelength to the analyte thereby minimising laser-mediated fragmentation of the latter. The analyte embedded in this matrix co-desorbs, achieving ionisation without directly receiving the laser light energy.

Matrix selection and optimization of the sample preparation are the most important factors in this type of analysis. The choice of matrix is based on the laser wavelength used and the class of compound to be analysed. Currently there are two main MALDI sources that deliver energy on different wavelengths. UV is more common and delivers the energy electronically using either an N_2 or frequency tripled Nd:YAG laser with a wavelength of 337 nm or 355 nm respectively. IR-MALDI delivers energy vibrationally to the matrix using wavelengths around 3 μm. Common UV-MALDI matrices used for peptide analysis include α-Cyano-4-hydroxycinnamic acid and 2,5-Dihydroxybenzoic acid.

The MALDI technique is relatively insensitive to contamination (salts, buffers, detergents, etc) compared to other ionisation techniques and produces mainly singly charged molecular species with very few multiply charged or fragment ions, producing simpler mass spectra. Historically, certain types of source / mass analyser have been used in combination. Although MALDI, a pulsed ion source, has been linked to other types of analyser, the most important has been the time-of-flight (TOF) analyser. The introduction of the MALDI source has led to considerable development and innovation in the area of TOF technology.

1.11.1.2 Electrospray ionisation

The success of the continuous ion source, electrospray (ESI), started when Fenn *et al.* (1989) showed that multiply charged ions could be obtained from proteins, allowing their molecular masses to be determined on instruments, whose mass range was limited to 2000 Da. At the beginning, ESI was thought to only have application for the analysis of proteins, but later, was extended not just to the analysis of polymers and biopolymers, but also to the analysis of small polar molecules. The source was easy to couple to high-performance liquid chromatography (HPLC), nano- and capillary- LC and capillary electrophoresis, and became the source of choice for such applications.

In ESI, gas phase ions are generated by applying a potential to a flowing liquid, containing both the analyte and solvent molecules. A strong electric field is applied to a liquid passing through a capillary tube (normally at 0.2 – 10 μl/min). The electric field is obtained by applying a potential difference (2 – 5kV)

between this capillary and the counter-electrode (**Figure 1.4**). This electric field at the spray tip creates a charge separation by, in positive ion mode, attracting anions to the capillary tip and repelling cations to the solution surface. The field pulls the surface towards the sampling aperture, but is opposed by the surface tension of the liquid. These forces balance forming a Taylor cone from which a spray appears. The solvent contained in the droplets evaporates causing them to shrink to the point where coulombic forces cause them to divide into smaller droplets. This process continues, producing smaller and smaller droplets, until the field strength on their surface becomes great enough to cause ions to be desorbed from the surface. The detection limits that can be achieved with ESI have improved dramatically by reducing the flow rates to the nanolitre/minute range using microspray (Andren *et al*., 1994) and nanospray inlets (Wilm and Mann, 1996). Andren *et al*. (1994) were able to detect 320 zeptomole/µL of a neuropeptide using a microspray inlet. ESI has typically been used in conjunction with quadrupole or ion-trap mass analysers to produce information by tandem mass spectrometry (MS/MS). However, the use of continuous ion sources, such as ESI with orthogonal acceleration time of flight mass (OaTOF) analysers is now commonplace.

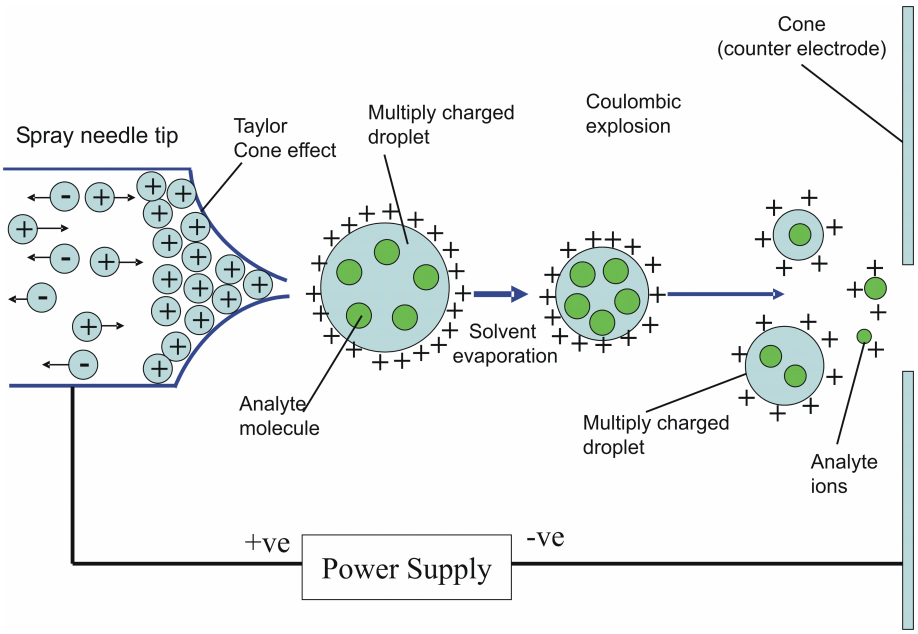


Figure 1.4. Representation of the Electrospray ionisation (ESI) process.

1.11.2 Mass analysers

Once the ions have been generated, they need to be separated according to their *m/z* ratio. Mass analysers use time (e.g., TOF) or an electric or magnetic field (e.g., quadrupole or ion trap) to separate ions of a particular *m/z* before their detection by the ion detector. There are many types of analyser. However, only four will be covered: quadrupole mass filters (Q), TOF analysers, Fourier-transform ion cyclotron resonance and Orbitrap, the first two of which, have been used during the course of the research described in this thesis.

1.11.2.1 Quadrupole mass filters

Quadrupole analysers (Paul and Steinwegen, 1953; Ferguson *et al*., 1965), consist of four parallel rods (**Figure 1.5**) with circular or, ideally, hyperbolic section. Oscillating and constant voltages are applied to these rods to generate a field that allows ions of a particular *m/z* to pass down between the rods, that is, to have a stable trajectory through the quadrupole to an ion detector. By scanning through a range of voltages, the field can be altered to select for ions of different *m/z* values. Ions that have unstable trajectories discharge themselves against the rod and are not detected.

As a scanning instrument, the quadrupole is more amenable to continuous ion sources such as ESI. However, although the stream of ions is continuous, only ions within a specified *m/z* window are allowed through; the remainder are wasted resulting in a low duty cycle (i.e. the number of ions detected in one scan). Quadrupoles are considered to be low-resolution instruments and are generally operated at unit resolution, i.e. a resolution that is sufficient to separate two peaks one mass unit apart.

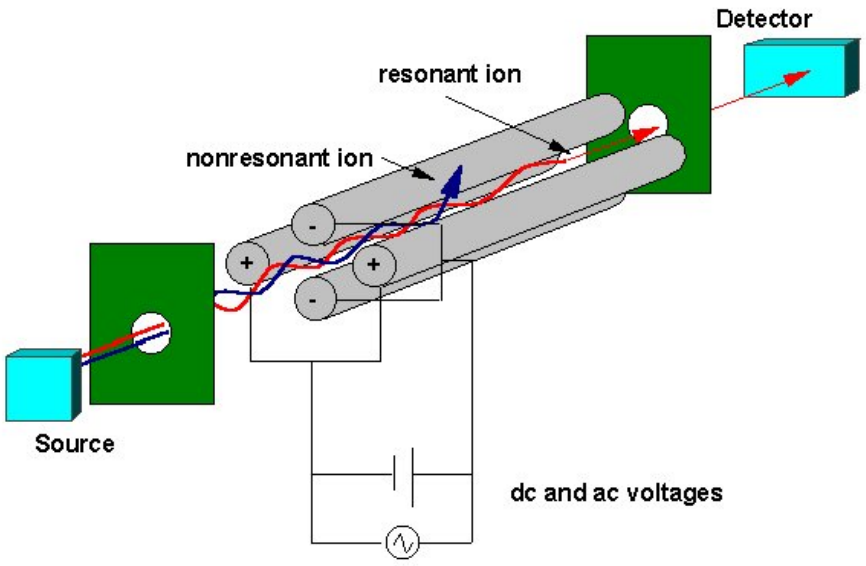


Figure 1.5. Layout of a quadrupole mass filter. The red arrow represents an ion with a stable trajectory, blue an unstable trajectory of the ion. Reproduced with permission from www.chem.vt.edu/chem-ed/ms/graphics/quad-sch.gif.

1.11.2.2 Time of Flight mass analysers

Stephens in 1946 (Stephens, 1946) described the principle of the time of flight analyser. Since the end of the 1980s there has been renewed interest, in part, due to advances in electronics to handle the dataflow, but mainly because of the development of pulsed ion sources like MALDI, to which TOF analysers have been particularly suited.

A time-of-flight analyser uses the differences in flight time through a field-free drift region to separate ions of different *m/z* values. Ions are produced in pulses and an electric field accelerates them into a field-free region at very low pressure (~10^{-7} Torr), with a kinetic energy qV, where q is the ion charge and V is the applied accelerating voltage. Since the ions kinetic energy is ½ mv^2, lighter ions have a higher velocity than heavier ions and reach the detector at the end of the drift region sooner. Their *m/z* value can be calculated using the following equation.

$$m/z = 2eV_{obs}\left(t/d\right)^2$$

e = charge of an electron, V_{obs} = accelerating potential, d=length of flight tube and t = time taken to traverse the flight tube.

Initially, one of the limiting factors in the development of TOF analysers was poor resolution. This could be attributed in part to Wiley and MacLaren's (1955) observation that ions of a particular *m/z* would reach the detector with a spread in arrival times. This is due to the effects of uncertainty in the time of ion formation, location in the extraction field and differences in initial kinetic energy, resulting in reduced resolution. Using a pulsed two-grid ion source, Wiley and McLaren developed an approach termed delayed extraction, which compensates for temporal, spatial and differences in kinetic energy distribution providing improvements in resolution (Wiley and MacLaren, 1955). In brief, all the ions are given the same initial kinetic energy by the extraction pulse and then drift along the field free region where they are separated, so that all ions of the same *m/z* arrive at the detector simultaneously. However, because the pulse energy is not felt by all ions with equal intensity, a kinetic energy spread for each *m/z* exists. This lowers the resolution by creating a time-of-flight spread for each *m/z*. Differences in these kinetic energies can be compensated by incorporation of a reflectron at the end of the flight tube. This is in essence an electrostatic mirror and consists of a series of electric fields that deflect ions back along the flight tube, resulting in a re-focusing of ions with the same *m/z* value on the ion detector. 'The development of the reflectron represents one of the most important advances in increasing resolution of the TOF analyser' (Mamyrin, 2001). Some modern commercial TOF analysers boast resolutions of up to 60,000.

The requirement of TOF analysers for discrete packets of ions (pulses of ions, i.e. from MALDI source) prohibited the use of continuous ion sources such as ESI, until the introduction of the orthogonal acceleration Time-of-Flight analyser (OaTOF) (Guilhaus *et al*., 2000). In an OaTOF analyser, an accelerating potential is applied at right angles to the continuous ion beam generated from the ion source. The ion beam is then chopped using a pulsed voltage supply coupled to the orthogonal accelerator to provide repetitive pulses at a frequency of a few kilohertz. These packets of ions are then allowed to drift into the flight tube where the ions separate according to their flight time.

In a TOF analyser, since all ion times (*m/z*) are measured on the micro-second time scale, the duty cycle is much shorter than scanning instruments, such as the quadrupole and hence offer increased sensitivity at full scan. The development of the OaTOF analyser has provided the advantage of being able to couple liquid chromatography, capillary electrophoresis, etc. through ESI sources.

Accurate mass measurement, coupled with sufficient resolution, greatly restricts the enormous number of possible molecular formulas that might be represented by a particular molecular mass. The technological advances of TOF in terms of resolution and hence mass accuracy make TOF analysers a powerful tool for proteomic analysis. Mass measurement accuracy of 10 ppm allows useful measurements of molecular formulas, although 1 to 2 ppm is preferable. TOF mass spectrometers are now providing better than 2 ppm mass accuracy routinely.

1.11.2.3 Fourier transform ion cyclotron resonance

The Fourier-transform ion cyclotron resonance (FT-ICR) analyser (Comisarow and Marshall, 1974a; Comisarow and Marshall, 1974b) is also an analyser that has been used for proteomic analysis. The analyser has extremely high sensitivity and resolution (up to 1,000,000) with mass accuracy that can exceed 1 ppm.

It has been proven to be particularly useful for the analysis of complex mixtures, i.e., tryptic digests of proteins, where it has been shown that the accurate mass of certain single peptides measured by FT-ICR, along with easily obtainable constraints, can be used to identify proteins unambiguously by sequence database searching (Goodlett *et al*., 2000).

In FT-ICR (reviewed by Marshall, 1998), ions are injected into a penning trap. These ions are then excited using a resonant excitation pulse where they begin to cycle (orbit). Once this excitation pulse is removed, the ions continue in their orbit. Within the trap, detector plates are located at fixed positions. As the ions move near these plates, they induce an image (electrical signal) on the detector plates. This image current will oscillate at the ions resonant frequency and can be amplified, digitised and recorded. Since ions with the same *m/z* will rotate at the same frequency, a fourier transformation on the signal can be used to deconvolute the data and obtain a mass spectrum.

FT-MS has the advantage of improved sensitivity as well as much higher resolution and thus precision. However, there are a number of disadvantages that has limited their wide acceptance in proteomic laboratories. For example, as the speed of front-end separations increases, the analysis time is reduced, reducing sensitivity and space-charge-related mass shifts limit dynamic range and mass accuracy. This has provided impetus for research into new types of
high resolution mass analyzers.

1.11.2.4 Orbitrap

A fundamentally new type of mass analyser, the Orbitrap (Makarov, 2000) has recently been introduced, offering high-resolution (typically 60,000 – 240,000 fwhm), sub- ppm mass accuracy (Olsen *et al*., 2005), high sensitivity, increased dynamic range and reduced running costs. (commercially available since 2005). Ions are electrostatically trapped in an orbit around a central, spindle-shaped electrode. They perform two kinds of movements in parallel: First, they cycle in an orbit around the central electrode. Second, they also move back and forth along the axis of the central electrode. Thus, the ion movement resembles a ring that oscillates along the axis of the spindle. This oscillation generates an image current in detector plates that is recorded. The frequencies of these image currents are dependant upon the *m/z* of the ions and mass spectra are obtained by Fourier transformation of the recorded image currents. As such, these instruments are routinely used within proteomic laboratories (Scigelova and Makarov, 2006).

1.11.3 Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) or multistage mass spectrometry (MS^n) allows more structural information to be obtained on a particular ionic species. This is particularly the case if soft ionisation techniques, which produce very little fragmentation, such as MALDI or ESI are used.

In the most common MS/MS experiment, a first analyser is used to isolate a precursor ion, which then undergoes fragmentation to yield product ions (also known as daughter ions), which are measured in a second spectrometer. It is possible to increase the number of stages. For example, one can fragment an ion of a particular mass to produce product ions and then select one of those product ions for further fragmentation and measure the resulting fragment ions. This would be termed MS/MSMS or MS^3. The number of steps can be increased further to yield an MS^n experiment (where *n* refers to the number of generations of the ions being analysed).

Tandem mass spectrometry can be achieved in two ways: in space by the coupling of two physically distinct instruments, or in time by performing an appropriate sequence of events in an ion storage device. Conventionally, for tandem in space methods, two transmissive mass analysers with an ion manipulation stage between them have been used. The most common instrument of this type is the triple quadrupole mass spectrometer (QqQ). The coupling of a quadrupole analyser to a TOF analyser, commonly known as Q-TOF, provided a major advancement for the area of proteomics allowing analysis using high resolution tandem mass spectrometry of peptides in complex mixtures using both MALDI and ESI ionisation (Morris *et al*., 1996; Shevchenko *et al*., 2000). However, obtaining higher order MS^n spectra requires *n* analysers to be combined in series and as such is difficult to implement because of the technical challenges and potential cost.

Tandem-in-time methods such as the ion trap, ICR and FT-ICR are able to achieve MS^n successively by analysing, reacting and reanalysing in the same instrument. A significant difference between an ion trap and FT-ICR is that in the former, ions are expelled from the trap to be analysed. In the Fourier transform

mass spectrometer, they can be observed non-destructively and are therefore measured at each step in the sequential fragmentation process. Ion traps have also been coupled with TOF analysers to remove the limitation of low *m/z* detection associated with ion trap instruments.

1.11.3.1 Collision Induced Dissociation fragmentation

In MS/MS and MS^n studies, a precursor ion is selected and generally fragmented in a collision cell generating product ions before the mass spectrum is acquired. Using this fragmentation data, structural information can be obtained for different types of molecules such as peptides, proteins, sugars and small molecules. One of the most common methods of fragmentation currently used in proteomics is collision induced dissociation (CID), also known as, collisionally activated decomposition (CAD) (Hayes and Gross, 1990; McLuckey, 1992). In CID, a precursor ion undergoes repeated collisions with a collision gas at a pressure, bringing the ion into an excited state. Once the fragmentation threshold is reached, the ion undergoes unimolecular decomposition forming product ions. The types of product ions generated are dependent upon the energy used and the precursor ion. At lower energies, neutral losses such as H_2O, CO, CO_2, etc are observed. At higher energies more structurally significant product ions are obtained and often result in cleavage of the molecule at characteristic positions such as those found in peptide fragmentation. The fragmentation of protonated peptides follows a defined nomenclature as shown in **Figure 1.6**. Product ions resulting from the backbone cleavage of the αC-C, the C-N amide linkage, or N-αC bond are called *a*-, *b*-, and *c*-type ions, if the charge is retained on the amino-terminal fragment, or *x*-, *y*-, and *z*-type ions, respectively, if the charge is retained on the C-terminal fragment ion. The product ions are numbered according to their positions from their respective terminal end. In general, the most commonly observed product ions are *b*- and *y*-type ions in low-energy CID. In high-energy CID conditions, *d*-, *v*-, and *w*-type ions corresponding to side chain cleavages may also be formed. Although there are alternative types of fragmentation techniques, CID was used during the course of the studies presented in this thesis.

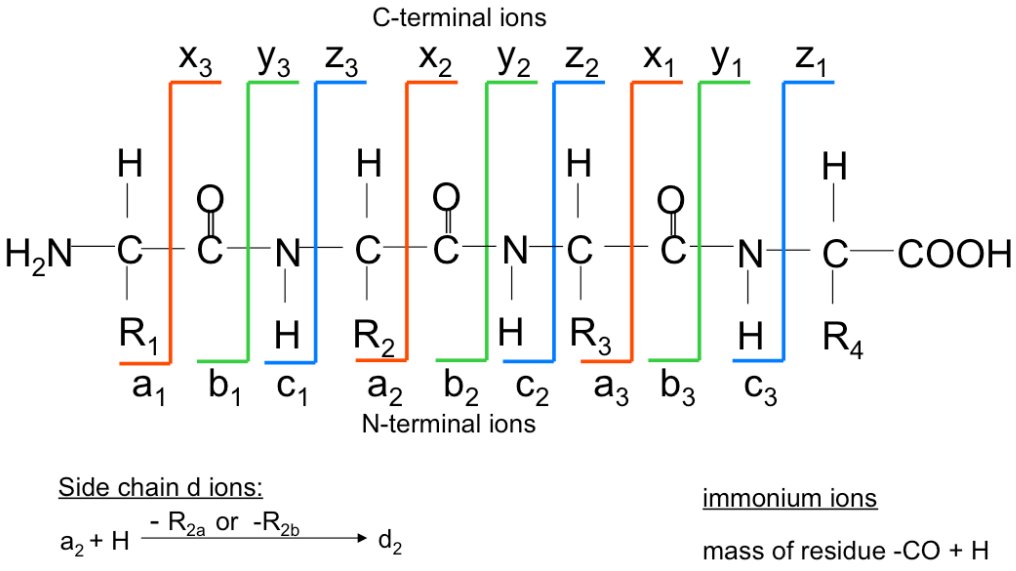


Figure 1.6. The nomenclature for ions derived from backbone fragmentation of a peptide by CID (Biemann, 1988).

There are four different types of scan in MS/MS and these are outlined below and in **Figure 1.7**. Even though these are based on scanning modes of a triple quadrupole, most can be applied to other instruments with very little modification.

1. Product ion scan - In this case, the precursor ion is focussed in the first mass spectrometer, fragmented in the collision cell and the resulting product ions measured in the second spectrometer.

2. Precursor ion scan - In this case the second mass spectrometer is held to measure the occurrence of a particular fragment ion and the first mass spectrometer is scanned. This results is a spectrum of precursor ions that arise from that particular product ion.

3. Neutral loss scan - In this case the first mass spectrometer is scanned as in (2) but this time the second mass spectrometer is also scanned but at a defined mass offset to produce a spectrum of precursor ions that undergo a particular neutral loss.

4. Selected reaction monitoring – In this case, both the first and second analysers are focused on selected masses so that the masses in the first mass spectrometer are only selected if a fragment ion of the correct mass is also observed.

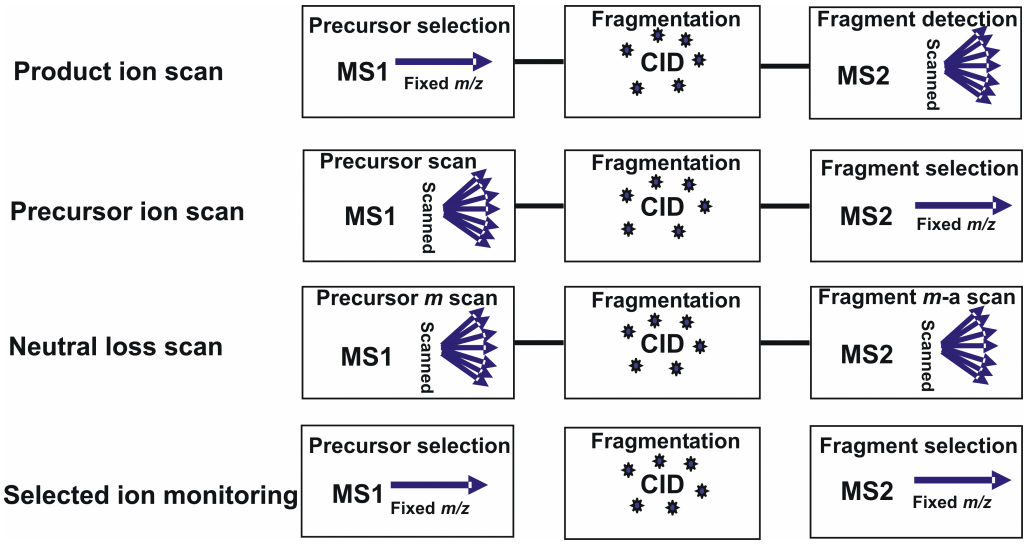


Figure 1.7. The four main scan modes in tandem mass spectrometry.

1.11.4 Peptide mass fingerprinting

MALDI-TOF-MS is high-throughput, sensitive and has traditionally been used as a primary screen for the identification of proteins from 2-D gels using a peptide mass fingerprinting approach (PMF). In PMF, a protein sample is digested with a site-specific protease, such as trypsin and the peptide masses determined by MS (Yates *et al*., 1993; James *et al*., 1993; Henzel *et al*., 1993). The measured peptide masses obtained provide a fingerprint of the protein under analysis. Using readily available genomic data, the masses of the measured proteolytic peptides are compared to predicted proteolytic peptides generated '*in- silico*' from the genomic data (**Figure 1.8**). In detail, a protein sequence entry from a genomic database is theoretically digested according to the specificity of the protease used for digestion, taking into account user-specified parameters such as number of missed cleavage sites and known modifications (e.g., cysteine modifications). Essentially, a theoretical mass spectrum is constructed for each protein within the database, statistically analysed and the best match to the submitted list of peptides obtained.

This type of approach is dependent on obtaining a sufficient number of peptides that match the protein being identified. However, the presence of unknown post-translational modifications or a reduced number of peptides as a consequence of low sample amounts, may result in some theoretical masses being unaccounted for, and hence reduce the statistical confidence of the identification. Other disadvantages are the requirement for the protein sequence to be present in the database of interest (i.e., a genome sequence is required) and that the presence of a mixture of proteins can significantly complicate the analysis and potentially compromise the results.

The comparison or scoring can be simple (i.e., number of matching peptide ions) or more complex. The problem with using the number of matches for assignment is that it does not show how well the theoretical match fits the experimental data (i.e., quality of match) or whether sufficient data has been provided to identify a hit above the background (i.e., significance match) (Eriksson and Fenyö, 2002). There are a number of scoring systems now available based upon a range of different scoring characteristics from heuristic to probalistic, all of which have their advantages and disadvantages. Several software packages have been developed, but the most commonly used are: MASCOT (Perkins *et al*.,1999), PepIdent (Gras *et al*.,1999), Profound (Zhang *et al*., 2000), SEQUEST (Eng *et al*., 1994), X!Tandem (Craig and Beavis. 2004) ProteinLynx (Waters, Manchester, UK) and more recently Andromeda (Cox *et al*., 2011). Peptide fragmentation patterns from MS/MS experiments may also be used in a similar approach to identify peptides. As described earlier, peptides fragment in a predictable manner, this property can be used to obtain theoretical fragmentation patterns for individual peptides, which can then be matched to the observed fragmentation patterns using similar matching and scoring algorithms.

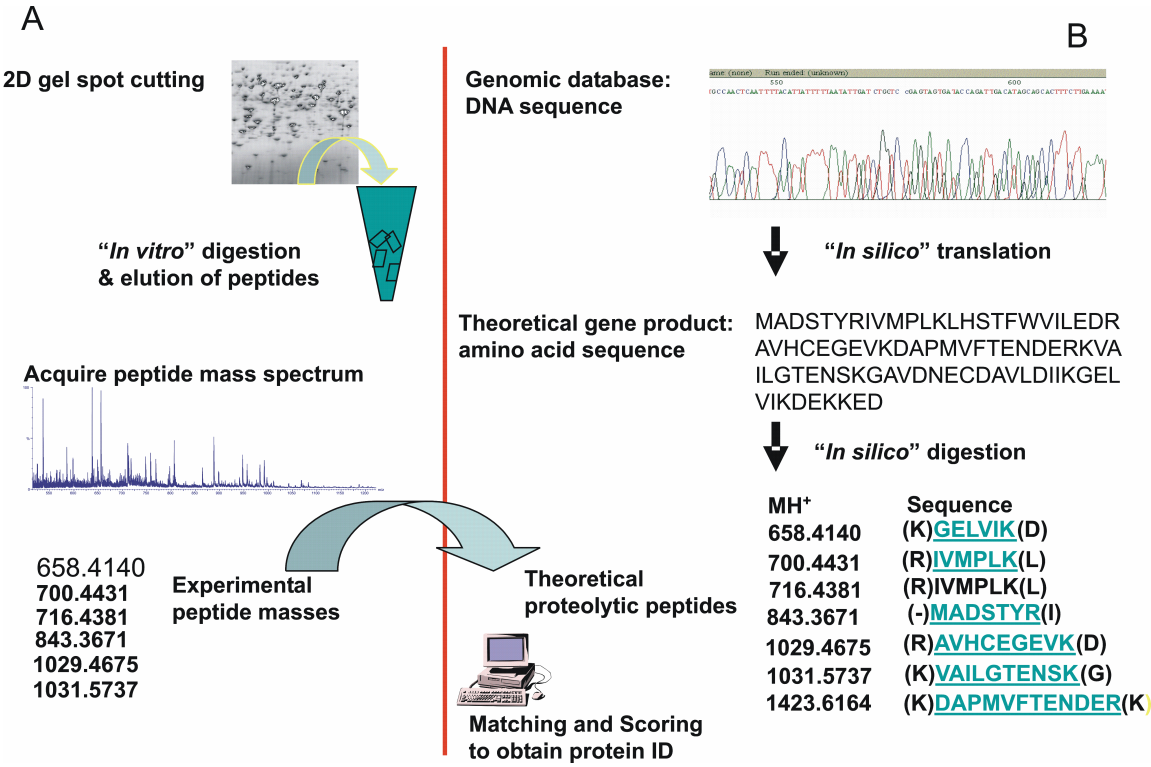


Figure 1.8. Representation of a peptide mass fingerprinting experiment. Workflow (A) outlines a process for generation of the experimental data. Workflow (B) outlines the '*in silico'* process for matching and scoring of the experimental data to a genomic database to obtain a protein ID.

1.11.5 LC-MS/MS

Peptide mass fingerprinting can allow the identification of a single protein or in some cases the components of a simple mixture. However, the analysis of complex mixtures or the identification of a protein from a single peptide is not possible. In contrast, combining liquid chromatography with tandem MS (LC-MS/MS), allows the identification of proteins from complex mixtures and can also provide identification of a protein from a single peptide. Universally, most LC-MS/MS experiments are accomplished by coupling on-line reversed phase high performance liquid chromatography (RP-HPLC) to an MS/MS instrument. Protein digests can be loaded onto the RP-HPLC column, separated and fragmentation information for each peptide obtained by MS/MS. Selection criteria for the acquisition of fragmentation data can be defined. The speed and complexity of these types of separations often prevents the collection of MS/MS data of all components within a mixture. A method called data-dependent-acquisition (DDA) allows criteria such as intensity and charge state to be used in an intelligent way to interrogate samples. As part of these criteria, exclusion lists can be used to eliminate previously analysed ions for a defined period of time, thereby forcing the pursuit of low abundance precursors and hence increasing coverage of the sample (Davis *et al*., 2001 and Spahr *et al*, 2001). The replacement of large bore columns by nano and capillary columns for LC-MS/MS, has vastly increased sensitivity by reducing the volume of solvent sprayed (between 50 and 400 nl/min) and effectively concentrating the sample as it enters the source, allowing the identification of proteins at the low femtomole level (Emmett and Caprioli *et al*., 1994; Natsume *et al*., 2002).

1.11.6 Qualitative analysis

In MS/MS, a specific peptide ion is selected and fragmented. Theoretically, this makes the complexity of the original digest irrelevant. However, in practice, one-dimensional peptide chromatography often does not provide sufficient peak capacity to separate peptides from complex mixtures to allow current mass spectrometers to 'keep up' with MS/MS demands (Michalski *et al*., 2011). To address this problem, various different combinations of protein and peptide separation schemes have been explored involving two- or three-dimensional chromatography and/or one-dimensional gel electrophoresis as discussed earlier in this Chapter. Alternative MS strategies have also been employed. Utleg *et al*. (2003), utilised gas phase fractionation to assign 128 previously unknown proteins, from a total of 139 from the human proteasome by repeated experiments using narrow but overlapping *m/z* ranges. Using a combination of these different fractionation strategies, including 1D-GE, peptide isoelectric focusing and gas phase fractionation, essentially complete proteome coverage of the yeast, *Saccharomyces cerevisiae* has been achieved (de Godoy *et al*., 2008). Nonetheless, the combinatorial nature of using multiple fractionation steps requires extensive MS instrument time (~1000 hrs) and are likely to prohibit the use of such strategies on a routine basis, even for relatively small proteomes. However, more recently, one-dimensional LC separations using long columns combined with orbitrap analysers, have afforded proteome coverage on a microarray scale within a single analytical run (Iwasaki *et al*., 2010; Nagaraj *et al*., 2012). These 'single-shot' approaches are very attractive since they reduce analysis time, minimise sample requirements, and potentially improve reproducibility between analytical runs. While at present they may not be suitable for in-depth characterisation of a proteome, e.g. isoforms or sites of post-translational modifications, with advances in both separation and MS technologies, 'single-shot' approaches could be applied to more complex proteomes.

1.11.7 Quantitative approaches

The concentration of an analyte and its relationship to the measured signal intensity depends upon a number of factors that are difficult to control and are not fully understood (Aebersold and Mann, 2003). As a result, without internal standards, mass spectrometers are poorly quantitative. To provide quantification, whilst utilising the speed and sensitivity of LC-MS/MS, strategies based upon stable isotopes have been employed. The difference in mass between pairs of chemically identical analytes of different stable isotope composition can be measured in a mass spectrometer. The measured ratio of the signal intensities between these pairs indicates the abundance ratio of the two analytes. It is on this basis that a number of strategies centred on stable isotope tags have been introduced.

Metabolic labelling of samples using isotopically labelled amino acids provides certain advantages. Stable isotope labelling with amino acids in cell culture, or SILAC (reviewed by Ong *et al*., 2003), utilises the incorporation of essential amino acids such as [^{13}C]-labelled arginine and/or lysine into proteins within a particular cell state. This can then be compared to a cell state where the proteins contain native [^{12}C] arginine and a differential expression ratio obtained. The early combination of samples from cells in the two states eliminates errors due to subsequent handling steps, such as the use of multi-step purification strategies. As

discussed in the previous section, de Godoy *et al*. (2008), made the landmark advance of essentially identifying the entire yeast proteome. Further, this study was a comparison of haploid and diploid yeast proteomes using SILAC, quantifying 4,399 proteins, demonstrating that relative quantitation data can also be obtained for complete proteomes.

As elegantly demonstrated above, SILAC is well suited to cells grown in culture; however, the application of the approach to intact multicellular organisms presents additional complexities. Nonetheless, the approach has been extended to nematode worms (Krijgsveld *et al*., 2003), human tissue (Geiger *et al*., 2010a) and even mice (Krüger *et al*., 2008). But, for many organisms, particularly animals, difficulties associated with isotope incorporation means that *in vivo* labelling is not an option (Beynon and Pratt, 2005). In such cases, post-isolation chemical isotope tagging of proteins, is currently the most commonly used labelling method. A number of isotope tagging chemistries have been described, although probably the most commonly used are ICAT (Gygi *et al*., 1999b) and iTRAQ (Ross *et al*., 2004; Choe *et al*., 2007), both of which are available from ABSciex (formerly available from Applied Biosystems).

The ICAT reagent contains a biotin affinity tag, a linker that contains stable isotopes and a thiol-specific reactive group. The method relies on tagging cysteine residues at the protein level and isolating peptides containing these tagged residues by affinity chromatography, after proteolytic digestion. This reduces the number of peptides isolated from the pool of peptides, thus providing the advantage of reducing sample complexity. Large-scale evaluation of the reproducibility of the ICAT approach, resulted in a median coefficient of variation of 18.6%. However, the technique was biased towards acidic proteins (pI<7), under represented small proteins (< 10kDa) and surprisingly showed no superiority over 2D-PAGE for the analysis of hydrophobic proteins (Molloy *et al*., 2005).

In the iTRAQ tagging method, the dependence on cysteine containing peptides is eliminated through the use of amine-specific tags, thus potentially allowing the tagging of most tryptic peptides (Ross *et al*., 2004; Choe *et al*., 2007). There are currently eight possible tags, which permit multiplexing of up to eight samples in a single experiment. The tags have an identical mass as a result of differences in other parts of the iTRAQ tag structure (**Figure 1.9**). Relative quantification is achieved using MS/MS via eight strong diagnostic product ions, *m/z* =113, 114, 115, 116, 117, 118, 119 and 121 Da, produced during fragmentation of the labelled peptide. A comparative study looking for markers of endometrial cancer using both iTRAQ and ICAT suggests that data obtained by the two approaches are complementary (Desouza *et al*., 2005). ICAT analyses were more selective and provided better detection of lower abundance peptides and proteins, conversely, iTRAQ analysis identified a larger percentage of abundant proteins by a number of multiple peptides, providing improved statistical confidence in the quantification data. Other potential advantages of the iTRAQ approach include, retaining post-translational modifications, which may be otherwise lost if using ICAT. Peptide lysates generated by digestion with trypsin are labelled via their N-termini using one of the eight different iTRAQ tags. Tagged peptides are selected and fragmented in the mass spectrometer, releasing a reporter ion whose intensity reflects the quantity of the peptide. Comparison of reporter ion ratios allows

differential expression analysis to be performed. The peptide backbone fragment ions are used to identify the peptide.

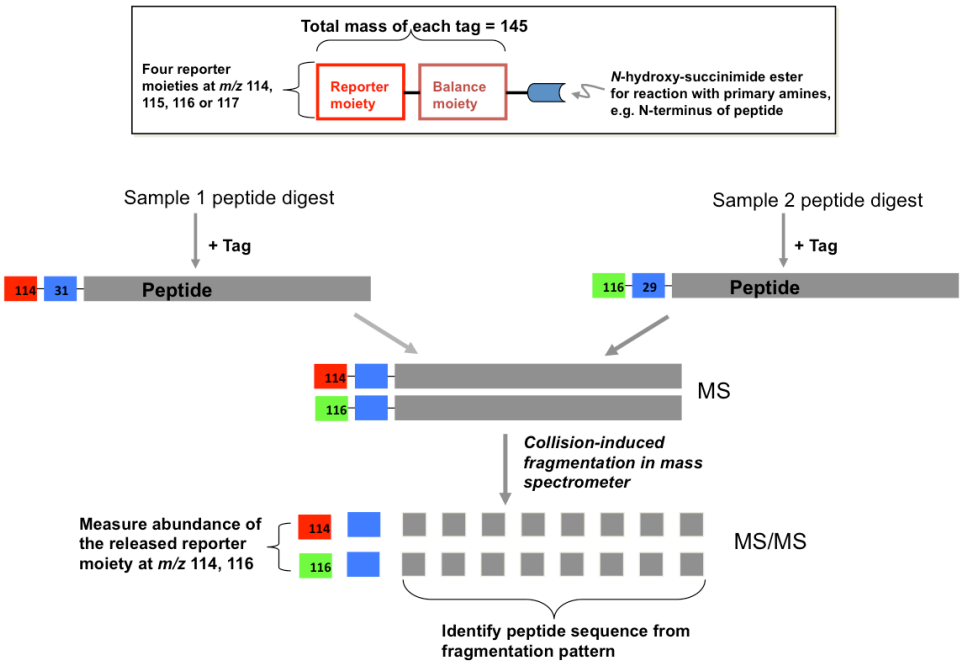


Figure 1.9. Structure of the iTRAQ tag and workflow for a 2-plex experiment (after Ross *et al*., 2004).

The quantification approaches discussed so far have been concerned with the relative amounts of a protein between two cellular states. A targeted approach termed AQUA, described by Gerber *et al.,* (2003), allows the measurement of targeted proteins in absolute amounts. An internal peptide standard is synthesised with incorporated stable isotopes, based upon a predicted tryptic peptide corresponding to the target protein of interest. In the method described by Gerber *et al*., protein lysates are separated by 1-DGE and bands corresponding to the molecular mass of the protein of interest are excised (**Figure 1.10**). The gel band is subjected to in-gel tryptic digestion in the presence of known amounts of the internal peptide standard and an LC-MRM experiment performed. By comparing the peak area of a specific MS/MS fragment ion from the heavily labelled internal peptide, to the corresponding fragment ion of the native peptide, the absolute concentration of the protein of interest can be calculated (discussed in detail in Chapter 3). Although still relatively low throughput, multiplexing strategies and the elimination of the 1-DGE separation step, has allowed precise measurement of 10's to 100's of proteins in a single assay, with high sensitivity. (Kuzyk *et al*., 2009). The approach is very attractive for validation studies and is more routinely being used as a better alternative to immune-based assays such as ELISA.

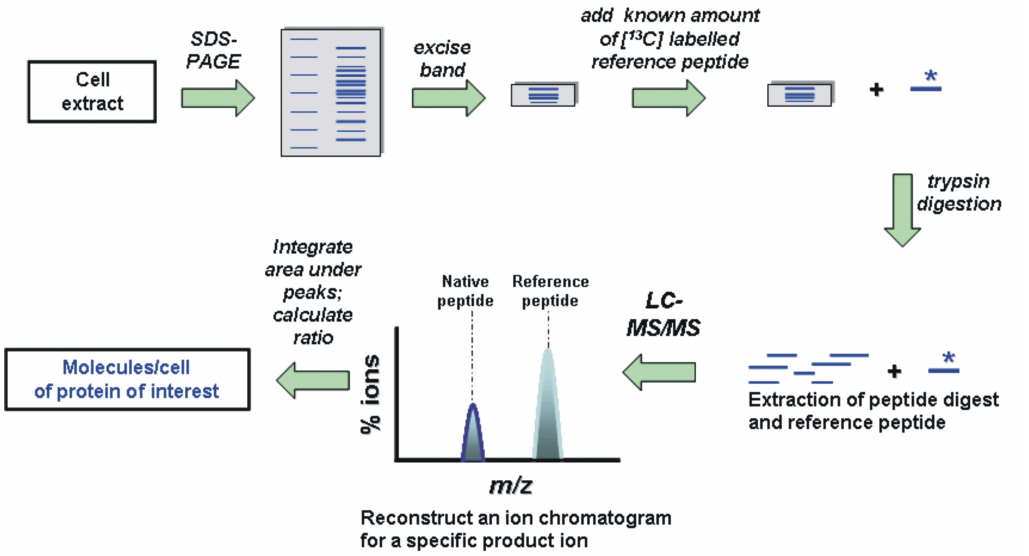


Figure 1.10. AQUA Workflow (after Gerber *et al*., 2003).

'Label-free' strategies for quantitation are increasingly being used since they require no labeling and can therefore be applied to almost any biological sample (reviewed by Neilson *et al*., 2011). These strategies take advantage of the accurate mass capabilities of high-resolution mass spectrometers and highly reproducible chromatography. Nearly all label-free approaches involve the integration of peptide ion abundances into chromatographic peak areas from LC-MS or LC-MS/MS runs. The integrated peaks for each peptide ion can then be aligned and compared across multiple LC-MS or LC-MS/MS experiments. However, for label-free acquisition using LC-MS/MS, there is a trade-off between acquisition of MS survey scans for precise peptide quantitation and the number of MS/MS scans required for peptide assignment. This balance can often be difficult to achieve, especially in complex mixtures, leading to precursor ions detected in MS, but with no corresponding MS/MS fragmentation spectra for identification. One way of resolving this is to perform multiple runs of the same sample, performing LC-MS for quantitation and then a separate LC-MS/MS experiment for identification. Integrated peptide intensities can then be associated with their respective peptide identity using an Accurate Mass Retention Time (AMRT) pair (Conrads *et al*., 2000; Page *et al*., 2004; Silva *et al*., 2005). For each peptide ion within an analytical run, an Accurate Mass Retention Time (AMRT) pair is generated consisting of a precise mass and the LC elution time of the peptide, providing a unique identifier. Using these AMRT pairs, the intensites of the same AMRTs can be compared across many LC-MS and LC-MS/MS chromatograms, providing both a relative measure of peptide abundance and peptide identity.

A label-free approach that allows the simultaneous quantitation and identification of proteins within a single analytical run is based upon the MS acquisition strategy LC-MS^{E} (Silva *et al*., 2005; Silva *et al*., 2006a; Silva *et al*., 2006b). Data is acquired using a Q-ToF MS, where the quadrupole is operated in RF mode (the quadrupole is only used to focus, but not select the ions). As the peptides are separated by RP chromatography and electrosprayed into the MS, the collision cell is alternated between low and elevated

collision energies at ~1 second intervals to obtain ion intensities from both precursor (low energy) and fragment ions (elevated energy) simultaneously. Using sophisticated software, each fragment ion is aligned with its corresponding precursor ion by exploiting subtle differences in the retention time maxima of each eluting precursor. Since there is no disruption to the MS signal by quadrupole switching, the precursor ion intensity can be used to compare ion abundances between different analytical runs, whilst the elevated fragmentation data is used to identify each peptide precursor. Using this data-independent approach, improvements in protein and proteome coverage are observed, increasing confidence in peptide assignment. This mode of acquisition has also recently been applied using alternative MS platforms (Geiger *et al*., 2010b).

An extension of this approach allows the MS peptide intensities identifying a particular protein to be used to determine their absolute concentration within a sample. Silva *et al*. (2006), showed that the average MS signal response for the top three most intense peptides from an internal protein standard, could be used as a universal response factor (counts/mol) that can be applied to the average of the top 3 most intense peptides of any other protein measured within that sample. The authors successfully used these absolute concentrations to determining the stoichiometry of functional complexes within *Escherichia coli.* Modifications of this Top3 strategy have also been successfully employed. Malmstrom *et al*. (2009), determined the absolute quantity of 19 proteins using targeted LC-MRM and heavily labelled peptide isotopes. These 19 anchor proteins were used as internal standards to apply the Top3 strategy, obtaining estimates of absolute protein abundance for 51% of the human pathogen *Leptospira interrogans,* ranging from 40,000 copies/cell to >10 copies/cell.

1.12 Aims and objectives

1. To establish, develop and apply both qualitative and quantitative proteomic techniques for the proteomic analysis of the intracellular bacterial pathogen *C. trachomatis* L2 in both isolated EB and RB forms.

2. To compare using qualitative proteomic techniques the protein expression profiles of EBs and RBs from *C. trachomatis* L2, to provide an insight into the biology of chlamydiae during its unique developmental cycle. The dataset generated will also provide a baseline for further quantitative studies.

3. To compare the protein expression profiles of EBs and RBs from *C. trachomatis* using quantitative proteomic techniques developed and established in the earlier part of the work in this thesis. Complementing existing transcriptomic data (Belland *et al*., 2003), the quantitative data generated will further increase our biological understanding of chlamydiae.

CHAPTER 2

MATERIALS AND METHODS

2.0 MATERIALS AND METHODS

2.1 Materials

The following HPLC grade solvents were obtained from Sigma-Aldrich (Poole, UK): acetonitrile, methanol. The following chemicals and reagents were purchased from Sigma-Aldrich: formic acid, acetic acid, trifluoroacetic acid, sodium chloride, potassium chloride, sodium dihydrogen phosphate (NaH_2PO_4), sodium monohydrogen phosphate (Na_2HPO_4), ammonium bicarbonate (NH_4HCO_3), triethylammonium bicarbonate, DTT (dithiothreitol), iodoacetamide, α-cyano-4-hydoxycinnamic acid, sodium dodecyl sulphate (SDS), copper sulphate ($CuSO_4.5H_2O$), bicinchoninic acid. Immobilised pH gradient (IPG) strips, pH 4-7 (Amersham-Pharmacia Biotech), ExcelGel, 12-14% SDS-acrylamide precast gels (Amersham-Pharmacia Biotech), 4-12% NuPAGE precast gels (Invitrogen, Paisley, UK), Sypro Ruby protein stain (Biorad, Hercules, CA), RapiGest™ (Waters corporation, Milford, USA).

Standard protein and peptides [Glu^1]-fibrinopeptide B (ACTH fragment 18-39), L-(tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK)-treated bovine trypsin, bovine serum albumin (BSA), horse heart myoglobin were purchased from Sigma-Aldrich. Alcohol dehydrogenase reference digest was purchased from Waters corporation (Milford, USA). Proteomic grade trypsin was obtained from Promega (Southampton, UK). Unless otherwise stated, all other reagents were acquired from Fisher Scientific (Loughborough, UK) and Promega (Southampton, UK).

Custom synthetic peptides were purchased from Peptide Research Products (Southampton, UK). [^{13}C] labelled isotopic Leucine was obtained from Cambridge Isotopes (Cambridge, UK).

Water for HPLC and all buffer and reagent preparations was produced using an in-house MilliQ™ water purification system (Millipore, Billerica, MA)

2.2 General techniques

2.2.1 Bacterial strains

The strains of *Chlamydia trachomatis*, *Salmonella* Typhimurium, *Escherichia coli* K-12 used in this study are listed below:

Bacteria	Strain
Salmonella Typhimurium	SL1344
E. Coli K-12	MG1655
Chlamydia trachomatis	L2/343/Bu (VR902B)

2.2.2 Bacterial growth media

S. Typhimurium (SL1344) cells were cultured using Luria-Bertani (LB) medium consisting of 10 g tryptone, 5 g yeast extract, 5 g NaCl per litre of deionised water. The pH was adjusted to 7.2 using NaOH. LB plates were produced by supplementation of LB media with 1.5% Bacto-agar No.3.

E.coli K-12 were cultured using 2YT medium consisting of 16 g tryptone, 10 g yeast extract, 5 g NaCl per litre of deionised water. The pH was adjusted to 7.4 using NaOH. All solutions were autoclaved at 121ºC for 20 min prior to use.

2.2.3 Growth of *Salmonella* Typhimurium SL1344

S. Typhimurium SL1344 were cultured in LB medium overnight at 37^{o}C with shaking. The culture was diluted 1 in 40 into fresh pre-warmed LB media and grown at 37^{o}C with agitation. At an A_{600} of approximately 0.5, the culture flask was cooled by incubation in ice water for 20 min prior to harvesting of the cells by centrifugation at 4,000 x *g* at 4°C for 20 min, using a JA21 rotor in a Beckman Avanti centrifuge. The cells were washed with pre-cooled 0.9% (w/v) sodium chloride solution (~250 ml), re-pelleted and stored at -80ºC.

2.2.4 Growth of *Salmonella* Typhimurium under conditions of osmotic stress

Overnight cultures of *Salmonella* Typhimurium were diluted 1:40 into fresh pre-warmed LB media and grown with gentle agitation at 37ºC. Sodium chloride in LB was added to a final concentration of 0.5 M. The cultures were incubated for a further 60 min at 37ºC with gentle agitation. When the cells had reached mid-exponential growth (A_{600} = 0.5), the culture flask was cooled for 20 min in ice water, prior to harvesting of the cells by centrifugation at 4,000 x *g* at 4°C for 20 min, using a JA21 rotor in a Beckman Avanti centrifuge. The cells were washed with pre-cooled 0.9% (w/v) sodium chloride solution (~250 mls) re-pelleted and stored at -80ºC.

2.2.5 Growth and preparation of *E. coli* K-12 (MG1655) for AQUA analysis

Overnight cultures of *E. coli* K-12 (MG1655) were diluted 1:6 into fresh pre-warmed 2YT broth and grown aerobically with gentle agitation at 37ºC to an absorbance at 600nm (A_{600}) of 1.0. The culture was diluted 1:6 into pre-warmed 2YT broth and 35 ml samples dispensed into 250 ml flasks and incubated at 37ºC with gentle agitation. At each time point, a 35 ml sample was transferred to a Falcon tube whereupon 500 µl of culture was removed and added to 500 µl of 60 mM sodium azide for determination of cell numbers using A_{600} and haemocytometer measurements, as described below. The remaining cells were harvested by centrifugation at 12,000 x *g* for 5 min at 4ºC. Pellets were re-suspended in 2 X Final Sample Buffer (as detailed in *section 2.4.2*) in the ratio of 100 µl per 0.1 A_{600}, incubated at 70ºC for 3 min and stored at -20ºC until use.

2.2.5.1 Cell counting

The absorbance at 600 nm (A_{600}) of 1 ml of culture containing sodium azide was measured against a suitable blank using a UV spectrophotometer (Hitachi, UK). For haemocytometer measurements, bacterial cells were counted using a standard counting chamber (Depth 0.1 mm, 1/400 mm^2, Hawksley, UK) in conjunction with a Carl Zeiss phase contrast binocular microscope fitted with a x60 objective lens. 100 µl of the diluted cells containing sodium azide were added to the counting chamber, covered with a cover slip and counted by visual inspection.

2.2.6 Growth of *C. trachomatis*

C. trachomatis L2/434/Bu (VR902B) was obtained from the American Type Culture Collection and cultured at 37ºC in Buffalo Green Monkey Kidney Cells (BGMK) in Dulbecco Minimal Essential Medium (MEM) (Sigma-Aldrich, Poole, UK) supplemented with 1 µg/ml of cycloheximide (Sigma-Aldrich) and 25 µg/ml of gentamicin (Sigma-Aldrich) as described by Pickett *et al*. (2005). For each purified preparation of EBs or RBs, *Chlamydia* were cultured in 14 x T-175 cm^2 flasks per preparation. Cells were harvested at 15 h post-infection for RBs and 48 h post-infection for EBs. Growth of *C. trachomatis* was performed in collaboration with Dr. Joanne Spencer and Mrs Leslie Cutcliffe at Southampton General Hospital.

2.2.7 Purification of EBs and RBs

Infected monolayers were detached with phosphate buffered saline (PBS) containing 0.125% (w/v) trypsin, 0.02% (w/v) EDTA. Individual cell suspensions were pooled and the total cells pelleted in Dulbecco MEM supplemented with 10% foetal calfs serum by centrifugation (Allegra 6R, Beckman Coulter, High Wycombe, UK) at 3,000 x *g* for 10 min. The infected cell pellet was re-suspended in 6 mls of PBS:H_2O (1:10) and homogenised for 6 min using a Dounce homogeniser to break open cells and release RBs and EBs. The homogenate was centrifuged at 250 x *g* to pellet the cell debris. The supernatant, which contained partially purified *Chlamydia,* was mixed with an equal volume of PBS.

Further purification of partially pure RBs and EBs was achieved using two cycles of density gradient centrifugation. The partially pure mixture was layered onto 35 mls of 20% (v/v) Urografin 370 (Schering Healthcare, UK) in PBS and centrifuged using an Optima L-80 XP Ultracentrifuge (Beckman Coulter) at 100,000 x *g* for 2 h in a pre-chilled Beckman SW28 rotor. After centrifugation, the supernatant was discarded and the cell pellet containing EBs and/or RBs, were resuspended in 2 mls of PBS. The chlamydial pellet was further purified, by layering onto a discontinuous urografin gradient consisting of 34%, 44% and 54% Urografin 370 in PBS layers. The gradient was prepared as follows: Three percentages of Urografin were prepared, 34% (11.9 ml PBS + 6.1 ml Urografin), 44% (6.2 ml PBS + 4.8 ml Urografin) and 54% (3.7 ml PBS + 4.3 ml Urografin). To an ultra clear centrifuge tube (Beckman Coulter), 18 mls of 34% Urografin was added. Using a syringe with a large metal needle, 10 mls of 44% Urografin was added by placing the needle tip to the bottom of the 34% Urografin and adding slowly. A clear interface between the gradients was observed. This was then repeated with 7 ml of the 55% Urografin. The 2 mls of resuspended EBs and/or RBs were gently layered onto the gradient cushion and centrifuged at 100,000 x *g* for 2 h in a Beckman SW28

rotor. After centrifugation, EBs banded at the 44%/54% interface and RBs banded at the 34%/44% interface. The collected bands were centrifuged at 35,000 x *g* for 30 min in a Beckman 55.2 rotor and pelleted. Each pellet was resuspended in ~ 500 µl of PBS and stored in aliquots at -80ºC. Purification of EBs and RBs was performed in collaboration with Dr. Joanne Spencer and Mrs Leslie Cutcliffe at Southampton General Hospital.

2.2.8 Host-free Protein Synthesis

Urografin purified RBs were immediately incubated in a host-free reaction mixture (100 mM Tris-HCl pH 7.0, 50 mM KCl, 5 mM MgCl2, 7.8mM creatine phosphate, 1.5 mg of phosphocreatine kinase per ml, 19 unlabelled amino acids, 1 mM ATP and 10 mCi of L-[^{35}S] methionine) as described (Stephens *et al*., 1998). After incubation at 37ºC, samples were solubilised in SDS-PAGE final sample buffer and SDS-PAGE was performed using the discontinuous buffer system method (Laemmli, 1970) with 10% acrylamide gels (acrylamide:bisacrylamide 38.5:1 w/w). Gels were stained and prepared for autoradiography by treatment with 1 M sodium salicylate - 50% methanol for 30 min at room temperature, then dried under vacuum and exposed to Kodak XAR-5 film at -70ºC. This procedure was performed courtesy of Mrs Leslie Cutcliffe at Southampton General Hospital.

2.2.9 Estimation of protein concentration

2.2.9.1 Bicinchoninic acid protein assay

Protein concentration was determined using the bicinchoninic acid (BCA) method (Smith *et al*., 1985). Protein standards were prepared using bovine serum albumin in the range 100 to 1000 µg/ml in the same buffer as the sample. 200 µl of 4% (w/v) $CuSO_4.5H_2O$ was added to 10 ml of bicinchoninic acid (BCA). 200 µl of BCA working reagent was added to 20 µl of each protein standard and 20 µl of each sample contained in a 300 µl 96 well microtitre plate. Samples and standards were measured in duplicate. After incubation at room temperature for 30 min, plates were read at 570 nm using a Dynex plate reader and analysed with Revelation 3.2 software (Dynex Technologies Limited, Worthing, UK)

2.2.9.2 Bradford protein assay

Protein concentration was determined using the Bradford protein assay (Bradford, 1976). Protein standards were prepared using bovine serum albumin in the range 50 µg/ml to 500 µg/ml in the same buffer as the sample. A working dye reagent was prepared by diluting 1 part dye reagent concentrate (BioRad, Hercules, CA) with 4 parts MilliQ H_2O (v/v). 200 µl of the diluted Bradford dye reagent was added to 10 µl of each protein standard and 10 µl of each sample contained in a 300 µl 96 well microtitre plate. Samples and standards were measured in duplicate. After incubation at room temperature for 10 min, plates were read at 570 nm using a Dynex plate reader and analysed with Revelation 3.2 software (Dynex Technologies Limited, Worthing, UK)

2.2.10 Genome quantification by real-time qPCR

A single copy of the *omcB* gene is located on the *C. trachomatis* L2 chromosome. The absolute number of genomes in both EB and RB preparations were accurately determined by performing 5′-exonuclease (TaqMan) assays with unlabelled primers and carboxyfluorescein/carboxytetramethylrhodamine (FAM/TAMRA) dual-labelled probes based on the *omcB* gene as previously described (Pickett *et al.,* 2005). 5 µl of sample was added to 20 µl reaction mixture containing forward primer (300 nM), reverse primer (300 nM), probe (100 nM) and TaqMan Universal PCR Master Mix (Applied Biosystems). Real-time PCR cycles were performed in an ABI PRISM 7700 Sequence Detection System (Applied Biosystems) according to the manufacturer's instructions. Genome determination was performed courtesy of Mrs Leslie Cutliffe at Southampton General Hospital.

2.3 Protein preparation and separation techniques

2.3.1 Two-dimensional gel electrophoresis

2.4.1.1 Preparation of EB and RB whole cell lysates

EB and RB cell pellets were centrifuged at 6,000 x *g* for 10 min in a bench-top centrifuge (Heraeus, Hanau, Germany) and the supernatant removed and discarded. Cell pellets were resuspended and extracted using lysis buffer (2% (v/v) Triton X-100, 2.5% (v/v)) IPGphor buffer, (range: pH 4-7), 2.5% (v/v) β-mercaptoethanol. Extracts were then treated with the Plus-one 2-D Clean-Up Kit (Amersham Biosciences, Little Chalfont, Bucks, UK) according to the manufacturer's instructions. Samples were snap-frozen in liquid N_2 and stored at -80ºC.

2.3.1.2 Immobiline DryStrip gel rehydration

Approximately 50 µg of EB or RB protein extracts were mixed with rehydration solution (containing per 5 ml: 8 M urea, 0.1 g CHAPS, 25 µl pH 4-7 IPG buffer, 15 mg dithiothreitol, and a few grains of Orange G) to give a final sample volume of 400 µl. The sample-buffer solution was added to an 18 cm IPG strip holder. An 18 cm pH 4-7 IPG gel strip "Immobilised DryStrip" was placed into the sample-buffer facing gel-side down. The lid was placed on the IPG strip holder and placed into the IPG-phor unit (Amersham Biosciences, Little Chalfont, Bucks, UK), in accordance with the manufacturer's instructions. The unit was programmed to allow rehydration of the strips for 20 h before isoelectric focusing (IEF) was performed as follows: 1 h at 300 V, 2 h at 500 V, 1 h at 1000 V, 2 h at 2000 V, 3 h at 3500 V followed by a sixth step at 5000 V for 24 h.

2.3.1.3 Equilibration of Immobiline DryStrips

The IPG strip of focused proteins was transferred to a glass tube ~ 10 ml volume (Fisher Scientific) and 10 ml of equilibration buffer 1 added (containing per 10 ml: 0.05 M Tris-HCl (pH 6.8), 6 M urea, 30% (v/v) glycerol, 1% (w/v) SDS, 25 mg dithiothreitol). The tube was sealed and incubated at room temperature with rocking for 10 min. After this time, the solution was discarded and replaced with equilibration buffer 2

(containing the same as buffer 1, but replacing the 25 mg of dithiothreitol with 0.45 g of iodoacetamide per 10 ml and a few grains of Bromophenol Blue to aid monitoring of the second dimension electrophoresis). After a further incubation of 10 min at room temperature with rocking, the IPG strip was removed and drained on filter paper for 15 min each side to remove excess liquid.

2.3.1.4 Second dimension polyacrylamide gel electrophoresis

The Multiphor II Electrophoresis unit was prepared and maintained at 15ºC in accordance with the manufacturer's instructions. The equilibrated IPG strip was placed gel side down onto a 12 -14% polyacrylamide gel (ExcelGel XL, Amersham Pharmacia Biotech) as detailed in the manufacturer's instructions. Electrophoresis was performed at 20 mA for 45 min and then at 40 mA until the dye front had reached the bottom of the gel. 2-D gel electrophoresis of EBs and RBs was performed in collaboration with Dr. Joanne Spencer at Southampton General Hospital.

2.3.1.5 Staining

Following electrophoresis, proteins were visualised using the fluorescent protein stain Sypro Ruby. The gel was washed with a small volume of fixing solution (containing 7% (v/v) acetic acid in 10% (v/v) methanol) to remove excess mineral oil. Once washed, the gel was placed in 300 ml of fixing solution and incubated at room temperature with gentle rocking for 30 min. After 30 min the fixing solution was replaced with 300 ml of Sypro Ruby stain and incubated for 16 h at room temperature with gentle rocking in a covered container to eliminate light. The Sypro Ruby stain was replaced with MilliQ H_2O and the gel incubated for a further 3 h in the container at room temperature with rocking.

2.3.1.6 Imaging

The Sypro Ruby stained gel was washed with MilliQ H_2O and placed gel facing up onto the transilluminator of a VersaDoc 3000 Imager (BioRad, Hercules, CA, USA). The imager was controlled through the 2-D gel image analysis software PDQuest (BioRad). The gel was imaged using a Sypro Ruby filter (462 nm excitation and 610 nm emission wavelengths) using an exposure time of between 15 and 30 sec, depending upon the intensity of the protein spots.

2.3.1.7 Image analysis and Gel spot excision

Spot detection, alignment and selection of gel spots for excision, were performed using the 2-D gel analysis software PDQuest (BioRad). Gel spots were excised using a Biorad ProteomeWorks™ spot cutter integrated and controlled through PDQuest. Alignment of the spot cutter with the high resolution gel image acquired using the VersaDoc was obtained by acquiring a low-resolution image using the on-board spot cutter camera to locate landmark spots to align the spot cutter to the high resolution gel image. Because the spot cutter used epi-illumination for excitation as opposed to trans-illumination used to acquire the high-resolution image on the VersaDoc, the sensitivity of the spot cutter camera was insufficient for detection of fluorescent protein spots for alignment of the cutter. Gels were therefore stained using Colloidal Coomassie Brilliant Blue (Colloidal CBB, BioRad) to allow visualisation using white light trans-illumination. Sypro Ruby stained gels were transferred into 300 ml of colloidal CBB and incubated for 1 h at room temperature

with gentle rocking. After 1 h, the stain was replaced with MilliQ H_2O and incubated for a further 3 h at room temperature with rocking. Gels were placed onto the cutting platform (gel side facing upward) of the spot cutter and imaged with white light trans-illumination for 1 second. This image was used to align the spot cutter to the high-resolution image acquired using the VersaDoc according to the manufacturer's instructions. Spots were cut using a 1.5 mm cutting tip and dispensed into 50 µl of MilliQ H_2O in a 96 well microtitre plate. The cutting tip was washed between each spot cut using 50% (v/v) methanol. The 96 well microtitre plates were sealed and stored at +4ºC until in-gel digestion was performed.

2.3.2 One-dimensional SDS gel electrophoresis

Protein samples were solubilised using 2 x Final Sample Buffer (FSB) (0.125 M Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 200 mM dithiothreitol, 0.05% (w/v) bromophenol blue). Samples were heated at 70ºC for 10 min and allowed to cool to room temperature prior to loading; 10 µL of Precision plus All Blue prestained markers (BioRad) were used for size approximation. Separation was performed using 4–12% NuPAGE Novex Bis-Tris gels (Invitrogen) with NuPAGE MOPS running buffer (Invitrogen) with an XCell SureLock apparatus (Invitrogen) at 200 V for 50 minutes. Gels were stained with 50 ml of Colloidal Coomassie Brilliant Blue (BioRad) for 1 h at room temperature with rocking and destained using analytical grade water for 16 h at room temperature with rocking.

2.3.3 Manual In-gel digestion

2.3.3.1 Reduction, alkylation and digestion

In-gel digestion was performed following the method of Shevchenko *et al*. (1996). Gel bands were cut into 1 mm x 1 mm pieces using a sterile scalpel blade and placed in a 0.5 ml microcentrifuge tube with 150 µl of MilliQ water. The water was removed and replaced with ~ 300 µl acetonitrile (approximately 3-4 times the volume of the gel pieces). The gel pieces were allowed to dehydrate for approximately 10 min before removing the acetonitrile and replacing with ~ 50 µl of 10 mM DTT in 0.1 M NH_4HCO_3 (enough liquid to cover the gel pieces). The gel pieces were incubated for 30 min at 56ºC to reduce proteins and subsequently dehydrated with acetonitrile as described above and the supernatant discarded. Alkylation of the proteins was achieved by addition of 40 µl of 55 mM iodoacetamide in 0.1 M NH_4HCO_3 and incubation for 20 min at room temperature in the dark. Gel pieces were then washed with ~300 µl of 0.1 M NH_4HCO_3 for 15 minutes and the supernatant discarded. The gel pieces were dehydrated using acetonitrile and the supernatant discarded. Where the gel pieces still retained Coomassie Blue stain, the samples were further incubated overnight in 1:1 (v/v) 0.1M NH_4HCO_3:acetonitrile, prior to removal of the solution, dehydration with acetonitrile and the supernatant discarded. Protein digestion was accomplished by the addition of enough trypsin (1.5 ng/µl trypsin in 50 mM NH_4CO_3, 5 mM $CaCl_2$) to cover the gel pieces and incubated at 37ºC for 45 min. After incubation, 5-25 µl of the trypsin buffer without trypsin (50 mM NH_4CO_3, 5 mM $CaCl_2$) was added to keep the gel pieces wet during enzyme cleavage. The samples were then incubated for 16 h at 37ºC. For AQUA experiments 500 fmol of the relevant stable isotopic reference peptide was added in 50 mM NH_4CO_3 to the gel pieces at the trypsin addition stage.

2.3.3.2 Peptide Extraction

After 16 h digestion, 15 µl of 25 mM NH_4HCO_3 was added to the gel pieces and incubated at 37ºC for 15 minutes with vigorous shaking. The gel pieces were briefly centrifuged to bring them to the bottom of the tube, acetonitrile added (~ 2 times the volume of the gel pieces) and incubated for a further 15 min at 37ºC with vigorous shaking. The gel pieces were centrifuged at 9,000 x *g* in a microcentrifuge (Heraeus) for 1 min and the supernatant collected into a 0.5 ml microfuge tube. 50 µl of 5% (v/v) formic acid was added to the gel pieces and after incubation at 37ºC for 15 min, acetonitrile added (~2 times the volume of the gel pieces) and the samples incubated for a further 15 min at 37ºC. The gel pieces were centrifuged at 9,000 x *g* and the supernatants pooled. Supernatants were lyophilised *in vacuo* using an Eppendorf Concentrator plus (Eppendorf, Hamburg, Germany) and stored at +4ºC until use.

2.3.4 Automated in-gel digestion

Automated in-gel digestion and peptide extraction was performed using an automated MassPREP™ workstation (Waters, UK). Gel pieces in a 300 µl 96 well microtitre plate were subjected to the following steps:

Destain

Two Coomassie Blue destain steps were performed consisting of the addition of 50 µl of 100 mM NH_4HCO_3 and 50 µl acetonitrile followed by incubation for 10 min at 40ºC and then the supernatant removed and discarded.

Dehydration

50 µl of acetonitrile was added to the gel pieces, incubated for 5 min at 40ºC and then removed and discarded.

Reduction

50 µl of 10 mM dithiothreitol in 100 mM NH_4HCO_3 was added to the gel pieces incubated at 40ºC for 30 min and then removed and discarded.

Alkylation

50 µl of 55 mM iodoacetamide in 100 mM NH_4HCO_3 was added to the gel pieces, incubated at 40ºC for 20 min and then removed and discarded

Wash

50 µl of 100 mM NH_4HCO_3 was added to the gel pieces and incubated for 10 min at 40ºC. 50 µl of acetonitrile was then added, incubated for 5 min and then removed and discarded.

Dehydration

50 µl acetonitrile was added to the gel pieces and incubated for 5 min at 40ºC. The liquid was removed and replaced with a further 50 µl acetonitrile, removed and allowed to evaporate for 15 min.

Digestion

25 µl of trypsin at a concentration of 6 ng/µl in 50 mM NH_4HCO_3 was added to the gel pieces and incubated for 15 min at 37ºC. 10 µl of 50 mM NH_4HCO_3 was added and the gel pieces incubated for 5 h at 37ºC.

First Extraction

30 µl of extraction buffer was added to the gel pieces (2% acetonitrile (v/v) containing 1% formic acid (v/v)) and incubated for 30 min at 37ºC.

MALDI target spotting

1.6 µl of matrix (2 mg of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile:H_2O (v/v), containing 0.1% trifluoroacetic acid (v/v)) was added to a steel MALDI target plate (Waters). 2.0 µl of extracted peptide solution was added onto the target and allowed to air dry at room temperature for 15 min.

Peptide transfer

15 µl of the remaining peptide solution was transferred to a clean 200 µl 96 well microtitre plate (ABgene, UK)

Second Extraction

12 µl of extraction buffer and 12 µl of acetonitrile was added to the gel pieces and incubated for 30 min at 37ºC. 15 µl of supernatant was removed and combined with the previously extracted peptides. Microtitre plates were lyophilised *in vacuo* and stored at +4 ºC until use.

2.3.5 In-solution digestion

Proteins were solubilised in 50 mM NH_4CO_3 containing 0.1% Rapigest. Tris[2-carboxyethyl] phosphine (TCEP) was added to a final concentration of 5 mM and incubated for 60 min at 60°C. Iodoacetamide was then added to a final concentration of 55 mM and incubated at room temperature in the dark for 15 min. Digestion was performed using TPCK-modified trypsin (Sigma-Aldrich), at an enzyme:substrate ratio (w/w) of 1:25 and incubated for 16 h or overnight at 37ºC.

2.3.6 Amino acid analysis

Amino acid analysis was performed by Alta Biosciences, Birmingham, UK.

2.3.7 GeLC – MS/MS of EBs and RBs

One-dimensional SDS polyacrylamide gel electrophoresis coupled with Nano-liquid chromatography-tandem mass spectrometry (GeLC-MS/MS) was used as previously described (Schirle *et al*., 2003) to identify proteins from *Chlamydia trachomatis,* serovar L2.

Purified elementary and reticulate bodies were re-suspended in 200 µl of 50 mM NH_4HCO_3 containing 0.1% (v/v) RapiGest, incubated for 10 min on ice and sonicated for 10 min in a sonication bath containing chilled water (~4ºC). Samples were centrifuged at 9,000 x *g* for 15 min to remove cell debris, the supernatant removed and the protein concentration determined using the BCA assay as detailed in *section 2.2.9*. Preparations containing 130 µg of whole cell lysate were solubilised in 30 µl of 2 x Final Sample Buffer, fractionated using a NuPAGE 4-12% gradient SDS-polyacrylamide gel and stained as described in *section 2.3.2*. After visualisation with colloidal Coomassie Blue, each gel lane (length: 7cm) was excised, cut into 29 equal-sized pieces, and each band subjected to *in-situ* trypsin digestion using a modified automated method of Shevchenko *et al*., (1996), as described in *section 2.3.4*. MALDI-ToF MS as detailed in *section 2.5.2* was used to confirm the presence of proteolytic peptides from gel bands excised and digested from the top, middle and bottom of the gel. Samples were lyophilised *in-vacuo* using an Eppendorf Concentrator plus (Eppendorf) and resuspended in 30 µl of acetonitrile:water (v/v) containing 0.1% (v/v) formic acid. NanoLC-MS/MS was performed on each of the peptide extracts as described in *sections 2.3.9.5* and *2.6.1* and the data processed and searched against a protein translation of the *C. trachomatis*, serovar D genome and NCBI human genome as described in *section 2.7*.

2.3.8 MudPIT

Purified elementary and reticulate bodies were re-suspended in 200 µl of 50 mM NH_4HCO_3 containing 0.1% (v/v) RapiGest, incubated for 10 min on ice and sonicated for 10 min in a sonication bath containing chilled water (~4ºC). Samples were centrifuged at 9,000 x *g* for 15 min to remove cell debris, the supernatant removed and the protein concentration determined using the BCA assay as detailed in *section 2.2.9*. ~100 µg of protein lysate was reduced, alkylated and digested as detailed in *section 2.3.5*. The resulting peptide digest was adjusted to pH 2 with 1 mM HCl and incubated at 70ºC to precipitate the Rapigest. The solution was clarified by centrifugation at 9,000 x *g* prior to MudPIT analysis.

A workflow of the procedure used for MudPIT analysis and a plumbing schematic are shown in **Figures 2.1 and 2.2**, respectively. ~100 µg of a protein digest of either EBs or RBs were diluted 1:1 with 5 mM KH_2PO_4 pH 3.0 containing 10% (v/v) acetonitrile and 10 µl injected using a low volume autosampler (Waters corporation) onto a 5 mm x 0.35 mm i.d. Optipak SCX trap column (Waters corporation), connected to a StreamSelect 10-port valve (Waters corporation). With the valve set at position 1 (**Figure 2.2**), sample was introduced onto the column and washed with aqueous 1% (v/v) acetic acid using a CapLC, nanoLC system (Waters corporation) for 5 min at a flow rate of 10 µl/min. Uncharged peptides that did not bind to the SCX trap column were washed onto the in-series PepMap C18 RP trap column, 5 µm, 100Å, 300 µm i.d. x 1 mm (Dionex, Sunnyvale, CA, USA). The 10-port valve was then switched to valve position 2 (Figure 2.2), allowing elution of the uncharged peptides from the RP trap onto an Atlantis C18 analytical column, (3 µm particle size, 75 µm i.d. x 150 mm, Waters corporation), where the peptides were separated and introduced on-line into a Q-Tof Global Ultima mass spectrometer. Data directed acquisition experiments were performed (see *section 2.6.1*) using the following linear RP gradient at a flow rate of ~200 nl/min: 5% Solvent A (acetonitrile/water, 5:95 (v/v); 0.1% (v/v) formic acid) to 60% Solvent B (acetonitrile/water, 97:3

(v/v); 0.1% (v/v) formic acid) formed over 40 min, before a steeper 15 min gradient to 80% solvent B. The column was maintained at 80% solvent B to remove remaining material and then re-equilibrated over 10 min to the initial starting conditions, and the 10-port valve switched back to valve position 1. 10 column volumes of 25 mM KCl was injected onto the SCX trap column, eluting released peptides onto the RP trap column. To remove salts, the RP trap was washed for 25 min with 1% (v/v) acetic acid at a flow rate of 10 µl/min and then the 10-port valve switched to valve position 2 to allow analytical separation and analysis of the eluted peptide fraction. This process was repeated using the following concentrations of KCl; 25 mM, 35 mM, 50 mM, 60 mM, 75 mM, 100 mM, 200 mM. Data was processed and searched against a protein translation of the *C. trachomatis* serovar D, *C. trachomatis* L2/434/Bu and *human* NCBI genomes as described in *section 2.7.*

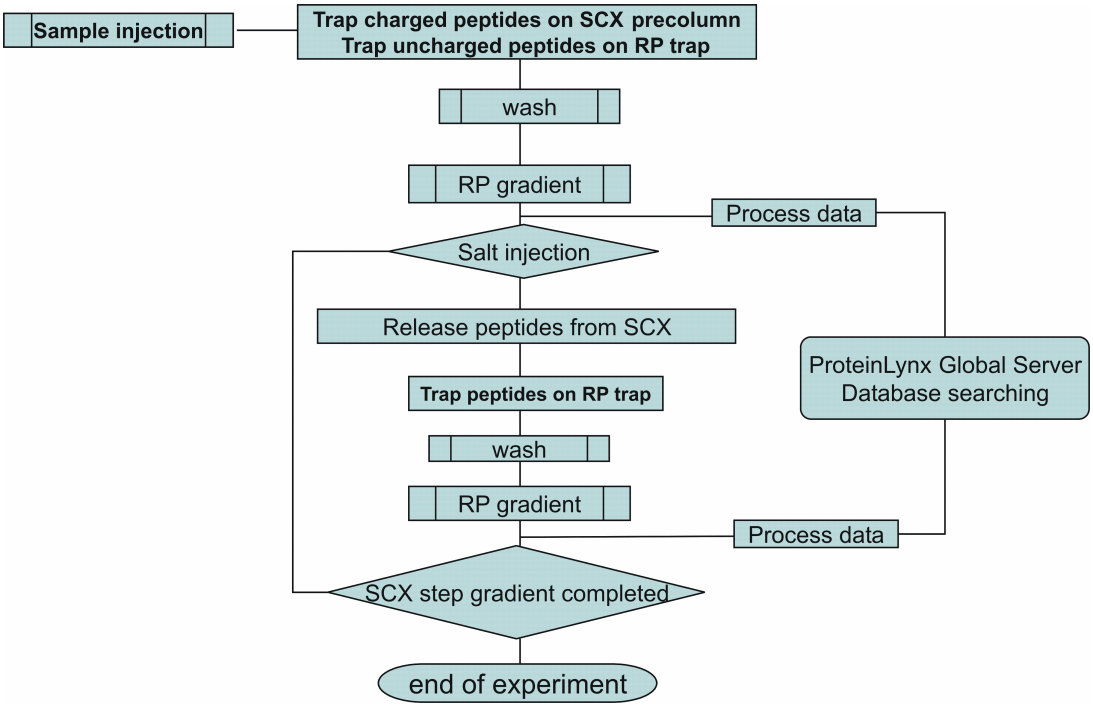


Figure 2.1. Experimental workflow of the proteomic analysis technique MudPIT used in this study**.**

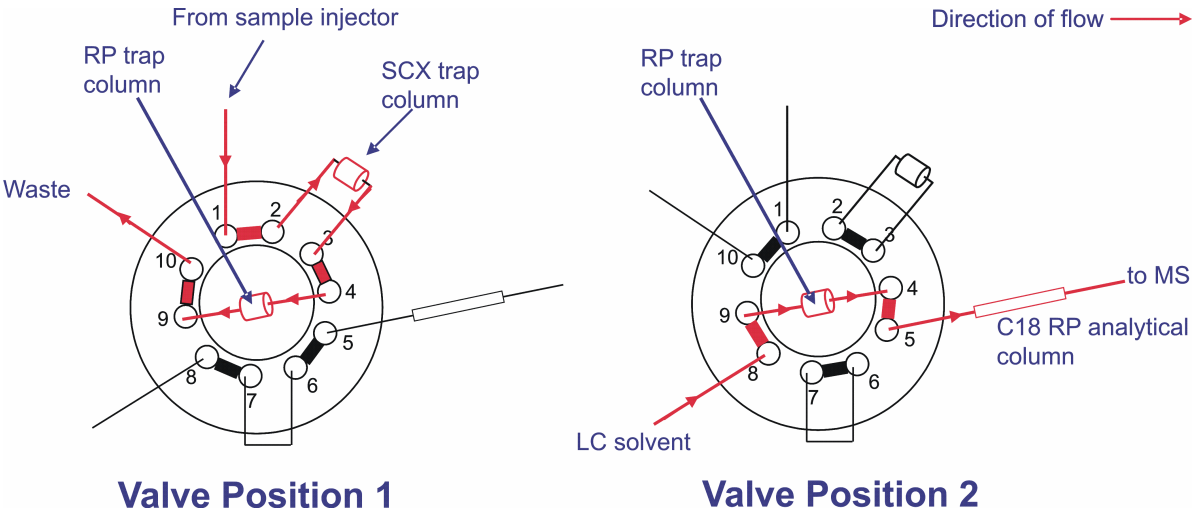


Figure 2.2. Plumbing schematic of the 10-port valve used for MudPIT analyses.

Initially, with the StreamSelect 10-port valve set to position 1, samples are introduced onto the SCX trap column via the sample injector using 1% acetic acid. Charged peptides bind to the SCX trap column and uncharged peptides pass through the SCX column onto the RP trap column. After a period of washing, the 10-port valve is switched to valve position 2. Peptides bound to the RP trap column are eluted onto the C18 RP analytical column and separated using a RP gradient developed using the LC system. When the analytical separation is complete, the 10-port valve is switched back to position 1. This process is repeated using increasing incremental concentrations of salt introduced via the sample injector to release fractions of peptides from the SCX column onto the analytical RP column.

2.3.9 iTRAQ

2.3.9.1 Preparation of EBs and RBs from *C. trachomatis for* iTRAQ labelling

400 µl of 0.5 M triethylammonium bicarbonate buffer containing 5 mM dithiothreitol and 0.1% (w/v) SDS was added to cell pellets of EBs (two biological replicates) or RBs (two biological replicates) and incubated on ice for 2 h. The resulting solution was transferred to a FastPrep lysis vessel containing 'Lysing matrix D' (Q-Biogene, CA). Samples were lysed using the Savant FastPrep system (Thermo Scientific, Langenselbold, Germany) for 3 cycles of 30 seconds at a speed setting of 6. The vessel was removed, chilled on ice for 3 min and the process repeated. The FastPrep vessel was then centrifuged in a bench-top centrifuge (Heraeus) at 9,000 x *g* for 5 min and the supernatant removed. A further 100 µl of dissolution buffer was added to the FastPrep vessel, vortexed, centrifuged for a further 4 min and the supernatant collected. The supernatants were pooled and the protein concentration determined using the Bradford protein assay (see *section 2.2.9*). Remaining supernatant was flash frozen using liquid nitrogen and stored at -80ºC.

2.3.9.2 Preparation of *S.* Typhimurium for iTRAQ analysis

Cell pellets, prepared as detailed in *sections 2.2.3* and *2.2.4* were re-suspended in 1.6 ml of 20 mM HEPES pH 8.0, transferred to a 5 ml sonication vessel and sonicated in a salt ice bath for 6 cycles (cycle: 15 sec ON followed by 1 min off) using an MSE Soniprep 150 (Sanyo) fitted with an exponential microprobe. The cell lysate was then centrifuged at 25,681 x *g* using a JA21 in a Beckman Avanti centrifuge for 20 min at 4ºC to remove cell debris and the supernatant collected. The protein concentration of the supernatant was determined using the BCA assay (*section 2.2.9*) prior to flash freezing the samples using liquid nitrogen and storage at -80ºC.

2.3.9.3 Preparation of samples for iTRAQ analysis from whole cell lysates of *Chlamydia trachomatis* and *Salmonella* Typhimurium

Samples for iTRAQ labelling were carried using the supplied kit and according to manufacturer's instructions, *Chemistry Reference Guide*. Part Number 4351918 Rev. A. 05/2004 (Applied Biosystems, Foster City, USA).

To each of up to four sample tubes containing ~100 µg of sample in 0.5 M triethylammonium bicarbonate buffer (where the final volume of dissolution buffer was no more than 34 µl), 1 µl of 2% (w/v) SDS (except for chlamydial protein lysates, since these already contained 0.1% (w/v) SDS) and 2 µl of the reducing agent, 50 mM TCEP (tris-(2-carboxyethyl) phosphine (TCEP) were added and vortexed. Samples were incubated at 60°C for 1 h. After incubation, 1 µl of the cysteine blocking reagent, 200 mM methyl methane-thiosulfonate (MMTS) in isopropanol was added, prior to incubation for a further 10 min at room temperature. Proteins were digested by adding 10 µl of 1 mg/ml trypsin containing 88.8 µg $CaCl_2$. Samples were incubated overnight at 37ºC.

After digestion, each vial of iTRAQ reagent was brought to room temperature and 70 µl of ethanol added to each reagent vial. Each vial was vortexed for 1 min to dissolve the iTRAQ reagent and centrifuged at 9,000 x *g* for 1 min. The contents of one iTRAQ reagent vial was added to one sample tube, vortexed for 1 minute and centrifuged at 9,000 x *g* for 1 min. For example, for a duplex-type experiment, the contents of the iTRAQ reagent 114 vial was added to the control sample, and the contents of the iTRAQ reagent 117 vial was added to the test sample. Sample tubes containing the iTRAQ reagents were incubated at room temperature for 1 h prior to combining them together into a 0.5 ml tube and lyophilizing *in vacuo* using an Eppendorf Concentrator plus (Eppendorf).

2.3.9.4 Fractionation of iTRAQ labelled peptides using strong cation exchange chromatography

The combined iTRAQ peptide mixture was separated by strong cation exchange (SCX) chromatography on a Dionex Ultimate nano-LC system using a Polysulfoethyl A column (4.6 mm i.d. x 150 mm, 5 µm, 300Å, Phenomenex, Cheshire, UK) or a ProPac® (1 mm i.d. x 250 mm SCX column, Dionex).

Samples were dissolved in 500 µl of buffer A (Buffer A: 25% (v/v) acetonitrile, 10 mM phosphoric acid) and loaded onto the column using a 500 µl loop. The loaded sample was washed with buffer A for 20 min at 200 µl/min to remove excess reagent. Peptides were eluted with a linear gradient of 0 – 500 mM KCl in 25% (v/v) acetonitrile, 10 mM phosphoric acid, at 200 µl/min with fractions collected at 1 min intervals. Peptide elution was monitored using UV absorbance at 214, 235 and 280 nm. Fractions were lyophilised and re-suspended in 50 µl of MilliQ H_2O prior to analysis by nanoLC – MS/MS as detailed in *sections 2.3.9.5.2 and 2.6.1*)

2.3.9.5 Reversed phase NanoLC-MS/MS

NanoLC separations were performed using a CapLC system (Waters Corporation, Manchester, UK), consisting of a µHPLC pump and low volume autosampler coupled to a Streamselect µ-column switching module (Waters Corporation).

Samples were stored in microtitre plates in a chilled area at 10ºC. Samples were loaded via a low volume autosampler (Waters Corporation) onto a PepMap C_{18} guard column (5 mm x 300 µm i.d., Dionex) for pre-concentration and desalting using 100% solvent C (acetonitrile/water, 3:97; (v/v) 0.1% (v/v) formic acid) at a flow rate of 20 µl/min. The eluent was diverted to waste. After 6 min of washing, the nano-reversed phase C_{18} PepMap analytical column (150 mm x 75 µm i.d., Dionex) was switched into line using the Streamselect µ-column switching module and a separation gradient performed as detailed below. For all experiments performed, a flow rate of 200 nl/min was set. The 75 µm i.d., 360 µm o.d analytical capillary LC column was coupled to 20 µm i.d., fused silica of the nanoLC sprayer (Waters Corporation) using a teflon zero dead volume connector.

2.3.9.5.1 NanoLC of Gel extracts

Separation of gel extracted peptides were performed using a linear RP gradient at a flow rate of ~200 nl/min: 5% Solvent A (acetonitrile/water, 5:95 (v/v); 0.1% (v/v) formic acid) to 60% Solvent B (acetonitrile/water, 97:3 (v/v); 0.1% formic acid (v/v)) formed over 40 min, before a steeper 15 minute gradient to 80% solvent B. The column was maintained at 80% solvent B to remove remaining material and then re-equilibrated over 10 min to the initial starting conditions. MS/MS data dependent acquisitions were performed on a Q-Tof Global Ultima as described in *section 2.6.1*.

2.3.9.5.2 NanoLC of iTRAQ samples

Separation of iTRAQ samples were achieved using a slow linear gradient of 7% solvent A (acetonitrile/water, 3:97 (v/v); 0.1% (v/v) formic acid) to 80% solvent B (acetonitrile/water, 97:3 (v/v); 0.1% (v/v) formic acid) formed over 145 min, maintained at 80% B for a further 10 min and returned to the initial starting conditions over a period of 8 min. MS/MS data dependent acquisitions were performed on a Q-Tof Global Ultima as described in *section 2.6.1*

2.4. AQUA

2.4.1 AQUA analysis of BipA protein from *E. coli*

Samples of *E. coli* K-12 (MG1655) taken at 15, 30, 45, 60, 120, 180 and 240 min intervals during growth were prepared as described in *section 2.2.5*, separated using a 4-12% NuPAGE SDS polyacrylamide gel, and stained using Colloidal CCB as detailed in *section 2.3.2*. Gel bands were excised from the gel, corresponding to the approximate molecular mass of BipA ± 15 kDa (actual molecular mass = 67355 Da) and in-gel digested in the presence of 500 fmol of the BipA reference peptide, corresponding to residues 440 – 446, (Leu *– Asp – Tyr – Val – Ileu – Pro – Ser – Arg, where * indicates a leucine contaning 6 x [^{13}C] atoms) as detailed in *section 2.3.3*. The concentration of the BipA peptide was determined by amino acid analysis (Alta Biosciences). The lyophilised samples were re-suspended in 10 µl of 10% (v/v) acetonitrile containing 1% (v/v) formic acid, injected onto the nanoLC and separated as described in *section 2.3.9.5* for the analysis of gel extracts. Mass spectrometry was performed using a Q-Tof Global Ultima (Waters Corporation). An MS/MS experiment in positive ion mode was performed isolating and fragmenting the doubly charged precursor ions, *m/z* = 481 and 484, corresponding to the native and reference peptide, respectively, using a collision energy of 15 eV. Data was collected from 50 to 1200 *m/z* at 1.0 scan/sec over the 65 min LC gradient. All other MS settings were as detailed in *section 2.6.1*. Ion chromatograms were reconstructed for the $[M+H]^+$ *y*4 ion, *m/z* = 472.28 for both the native and reference peptide. The peak areas of each extracted ion chromatogram were calculated using MassLynx 4.0 (Waters Corporation) and used to calculate an abundance ratio. Using the previously determined concentration of the BipA reference peptide, the concentration of the protein BipA was determined. This process was repeated for each sampled time point.

2.4.2 AQUA analysis of the Major Outer Membrane Protein and a Metalloprotease from *C. trachomatis*

Purified elementary and reticulate bodies were re-suspended in 200 µl of 50 mM NH_4HCO_3 containing 0.1% (v/v) RapiGest, incubated for 10 min on ice and sonicated for 10 min in a sonication bath containing chilled water (~4ºC). Samples were centrifuged at 9,000 x *g* for 15 min to remove cell debris, the supernatant removed and the protein concentration determined using the BCA assay as detailed in *section 2.2.9*. 30 µg of both EB and RB protein lysates were separated using a 4-12% NuPAGE SDS polyacrylamide gel and stained using Colloidal CBB as detailed in *section 2.3.2*. Gel bands were excised from the gel corresponding to the approximate molecular mass of the Major Outer Membrane Protein (CTL0050) and the Metalloprotease (CTL0328) ± 15 kDa (actual molecular masses = 42,400 and 69,200 Da, respectively) and in-gel digested in the presence of 500 fmol of the MOMP and Metalloprotease reference peptides G-Y-V-G-Q-E-F-P-L*-D-L-K (MOMP) and I-S-L*-G-I-P-L-K (Metalloprotease), where * indicates a leucine contaning 6 x [^{13}C] atoms) as detailed in *section 2.3.3*. Peptide concentration was determined by amino acid analysis (Alta Biosciences). The lyophilised samples were re-suspended in 10 µl of 10% (v/v) acetonitrile containing 1% (v/v) formic acid, injected onto the nanoLC and separated as described in *section 2.3.9.5* for

the analysis of gel extracts. Mass spectrometry was performed using a Q-Tof Global Ultima (Waters corporation). MS/MS experiments were performed in positive ion mode isolating and fragmenting the doubly charged precursor ions corresponding to the native and reference peptides of both MOMP (*m/z* = 683.4 and *m/z* = 686.4) and the metalloprotease (*m/z* = 420.7 and *m/z* = 423.7), using collision energies of 22 eV and 17 eV, respectively. Data was collected from 50 to 1500 *m/z* at 1.0 scan/sec over the 65 min LC gradient. All other MS settings were as detailed in *section 2.6.1*. Ion chromatograms were reconstructed for the *y* ions of the native and reference peptides for both MOMP ($y6_{\text{native}}$ *m/z* = 586.4; $y6_{\text{reference}}$ *m/z* = 591.4) and the metalloprotease ($y7_{\text{native}}$ *m/z* = 727.6; $y7_{\text{reference}}$ *m/z* = 733.6). The peak areas of each extracted ion chromatogram were calculated using MassLynx 4.0 (Waters Corporation) and used to calculate abundance ratios. Using the previously determined concentrations of the reference peptides, the concentration of the metalloprotease and MOMP were determined.

2.4.3 Absolute quantitation of horse heart myoglobin using infusion

1 mg of horse heart myoglobin was reduced, alkylated and digested with trypsin in 1 ml of 50 mM NH_4CO_3 containing 0.1% (w/v) Rapigest as described in *section 2.3.5*. Samples were prepared containing a myoglobin digest at a range of concentrations from 500 amol to 15 pmol, each containing 1 pmol of the [^{13}C] labelled myoglobin reference peptide (A-L-E-L*-F-R, where * indicates a leucine containing 6 x [^{13}C] atoms). The final volume was 10 µl and all dilutions were prepared in 50% methanol (v/v) containing 1% acetic acid (v/v). The concentration of the labelled reference peptide was determined by amino acid analysis (Alta Biosciences, Birmingham, UK). Each sample was centrifuged at 9,000 x *g* for 5 min and transferred to a 200 µl 96 well microtitre plate. 10 µl of sample was aspirated using the Nanomate system and an infusion MS/MS experiment performed as detailed in *section 2.6.3*. Samples were sprayed for 30 min at a chip voltage of +1.7 kV and a gas pressure of 0.6 psi. MS/MS data was acquired for 2 min from 50 to 1500 *m/z* at 5 scan/sec, isolating and fragmenting the doubly charged precursor ion for the native peptide and isotopically labelled reference peptide (*m/z* = 375.2 and 378.2, respectively). A collision energy of 14 eV was applied, a value previously determined by infusion of 10 µl of 1 pmol of the isotopically labelled myoglobin reference peptide prepared in 50% (v) methanol containing 1% (v/v) acetic acid as detailed in *section 2.6.3*. Tuning of the collision energy was performed to maximise the signal response for the $[M+H]^+$ *y*4 ion of the reference peptide, *m/z* = 570.2, corresponding to the native peptide $[M+H]^+$ *y*4 ion, *m/z* = 564.2. After data collection, ion chromatograms were reconstructed for the *m/z* transitions $[M+2H]^{+2} \rightarrow$ *y*4 for the native and reference peptide, respectively (378.2 → 570.2 and 375.2 → 564.2), and the intensities used to obtain an abundance ratio. The previously determined concentration of the myoglobin reference peptide was used to calculate the concentration of the native reference peptide and hence the myoglobin concentration. This was repeated for each time point.

2.4.4 Absolute quantitation of myoglobin in human serum by infusion

Human serum (A kind gift from Prof. Swee Lay Thein, Kings College, London) at a concentration of ~75 mg/ml was diluted 1 in 200 with 50 mM ammonium bicarbonate containing 0.1% (v/v) Rapigest and digested as described in *section 2.3.5*. Digested serum was diluted 1:1 with 50% (v/v) methanol containing 1% (v/v) acetic acid. Samples were spiked with 1 pmol of the myoglobin reference peptide and the native

horse heart myoglobin digest at a range of concentrations (1 fmol - 30 pmol) in a final volume of 10 µl. Samples were centrifuged at 9,000 x *g* for 5 min and transferred to a 200 µl 96 well microtitre plate. Infusion and MS/MS were performed as described in *section 2.6.3*. Myoglobin protein concentrations were determined as detailed in *section 2.4.3*.

2.5 Label-free

2.5.1 Preparation of EB and RB samples for 2D-RPLC-MS^E

Protein lysates of EBs (two biological replicates) and RBs (two biological replicates) were prepared as described in *section 2.3.9.1*. Each sample containing 100 µg of EB and RB protein lysates (four samples) were made up to 100 µl using 0.5 M TEAB. 2 µl of the reducing agent, 50 mM (tris-(2-carboxyethyl) phosphine [TCEP] was added and incubated for 1 h at 60ºC. After incubation, 1 µl of the alkylation reagent, 200 mM methyl methane-thiosulfonate in isopropanol was added and incubated for 10 min at room temperature. Samples were proteolytically digested by the addition of 10 µl of trypsin at 1 mg/ml containing 88.8 µg of $CaCl_2$ and incubated overnight at 37ºC. Digested protein lysates were lyophilized *in vacuo* and re-suspended in 50 µl of 100 mM ammonium formate containing 8% (v/v) acetonitrile and a tryptic digest of the protein alcohol dehydrogenase from *Saccharomyces cerevisiae* (Waters Corporation), used as an internal standard at a final concentration of 20 fmol/µl. Each sample was analysed using 2D-RPLC-MS^E as described in *section 2.6.4* and the data processed according to *sections 2.7.3* and *2.8.3*.

2.6 Mass Spectrometry

2.6.1 ESI – Q-ToF MS and MS/MS

Except the infusion studies and label-free quantitative analyses, described in *section 2.6.3* and *2.6.4*, all data were acquired using a Q-Tof Global Ultima fitted with a Z-spray and nanoLockSpray source (Waters). [Glu^1]-fibrinopeptide B ($[M+ 2H]^{+2}$ = 785.8426) was used as the internal lockmass calibrant for the nanoLockSpray source. For automated Data Directed Acquisitions (DDA), a survey scan was acquired from *m/z* 375 to 1800 (1.0 scan/sec, 0.1 sec inter-scan delay) in positive ion mode with the switching criteria for MS to MS/MS including (i) ion intensity (15 counts per second); (ii) charge state ($^+2$, $^+3$, $^+4$) and (iii) exclusion list (see *section 2.6.1*). Product ion spectra were acquired from *m/z* 50 to 1800 at 1 scan/sec until the ion intensity fell below a threshold of 5 counts per second or data had been collected for 12 s, which ever occurred first. Six channels for product ion acquisition were used. The collision energy used to perform MS/MS was automatically varied according to the mass and charge state of the eluting peptide using a collision energy profile. For all experiments performed the general parameters shown in Table 2.2 were used unless otherwise stated. In MS mode, the quadrupole is operated in RF mode, transmitting a wide range of ions. To increase the *m/z* range transmitted, this transmission window can be moved during a scan. This movement is called the MS profile and is controlled by 5 variables: *m/z* 1, *m/z* 2, the dwell time at *m/z* 1, the ramp time between *m/z* 1 and *m/z* 2 and the dwell time at *m/z* 2. These three variables are given values as a percentage of the total scan time. Typically, these values were maintained at *m/z* 1 = 400 and *m/z* 2 = 600

with a dwell time of 30% and a ramp rate of 20% for both. This allows detection of ions at 0.6 of the lower boundary *m/z* (0.6 x 400 = 240 *m/z*) and 4 x the upper *m/z* value (4 x 600 = 2400 *m/z*). However, for iTRAQ experiments it was found necessary to lower *m/z* 1 to 100 *m/z* to allow for transmission of the low mass reporter ions in the range 114 – 117 *m/z*. The RF lens value was also decreased from 1.0 to 0.6 to further provide improved signal response in the low mass region.

Instrument control and data acquisition were provided by the software MassLynx 4.0 (Waters Corporation). The instrument was calibrated daily using the fragment ions of [Glu^1]-fibrinopeptide B ($[M+2H]^{+2}$ (*m/z* = 785.8426) as shown **Table 2.1**. Additionally, 180 fmol of an enolase digest (Waters Corporation) was used routinely to check the nanoLC-MS/MS analysis process.

Table 2.1. List of product ion masses from [Glu^1]-fibrinopeptide B, ($[M+2H]^{+2}$ = 785.8426 *m/z*) used for the calibration of the Q-Tof Global Ultima and Q-Tof Micro.

m/z
72.081
120.081
175.119
187.071
246.156
333.188
480.257
627.325
684.346
813.389
942.432
1056.475
1171.502
1285.544

Table 2.2. General parameters used for mass spectrometry performed using a Q-ToF Global Ultima and Q-ToF Micro.

Parameter	**Q-Tof Global Ultima**	**Q-Tof Micro**
Capillary voltage (V)	+3.00	Applied via Nanomate
Cone voltage (V)	100	20
Extraction cone (V)	n/a	1.5
Source temperature (ºC)	50	40
Desolvation temperature (ºC)	0	0
Cone gas flow (l/h)	20	20
Resting Collision energy (eV)	10	7

2.6.2 MALDI-ToF MS

MALDI-ToF experiments were performed using a M@LDI HT (Waters, Manchester, UK) in positive ion mode.

Samples were spotted onto stainless steel lockmass target plates (Waters) according to the dried droplet method (Karas and Hillenkamp, 1988). Samples were mixed 1:1 with matrix and then 1 µl was spotted on to the target plate and allowed to air dry. The matrix used for peptide analysis comprised of 2 mg/ml α-cyano-4-hydoxycinnamic acid in 50:50 (v/v) acetonitrile:water containing 0.1% (v/v) trifluoroacetic acid. External calibration was performed using monoisotopic masses of a bovine serum albumin protein digest at the beginning of each day. All analyses were performed using a lockmass correction. [Glu^1]-fibrinopeptide B $[M+H2]^{2+}$ (*m/z* = 785.8426) prepared in 2 mg/ml α-cyano-4-hydoxycinnamic acid in 50:50 (v/v) acetonitrile:water containing 0.1% (v/v) trifluoroacetic acid and was used for lockmass correction at a concentration of 200 fmol/µl on target.

All samples were analysed using positive ion reflectron mode with delayed extraction. The pulse and source voltages were set at 2,340 V and 15,000 V, respectively. A low or medium coarse laser setting was used and fine laser energy tuning was provided under software control. Data was collected from *m/z* 800 – 3500. Data was processed using MassLynx 4.0 (Waters Corporation).

2.6.3 Infusion MS and MS/MS

Infusion of samples was achieved using an automated chip infusion system (TriVersa™ Nanomate, Advion Biosciences, NY, USA) coupled to a Q-Tof Micro mass spectrometer fitted with a z-spray source (Waters Corporation). Lyophilised samples were re-suspended in 10 µl 50% (v/v) methanol containing 1% (v/v) acetic acid, otherwise samples were infused as prepared, without further solvent addition. Samples were maintained at 4ºC in a 200 µl 96 well microtitre plate. Initially, to generate the nanoelectrospray, a voltage of +1.7 kV with a gas back pressure of 0.6 psi was applied to the samples in a conductive tip applied to the back of a nozzle on the chip (**Figure 2.3**). Depending upon the MS signal response, the voltage was fine-tuned in the range +1.6 to +2.2 kV and the gas pressure from 0.6 to 1.5 psi to maximise the signal intensity. Flow rates vary between ~50 and 300 nl/min depending upon the voltage and gas pressure used. For each analysis, a new tip and nozzle were used preventing carryover. The Nanomate system was controlled using ChipSoft 7.0 software (Advion Biosciences, NY, USA). The Q-Tof micro was operated in positive ion mode using an MS scan range of 300 to 1700 *m/z* (1.0 scan/sec, 0.1 sec inter-scan delay) and an MS/MS scan range of 50 to 1700 *m/z* (1.0 scan/sec, 0.1 sec inter-scan delay). Other MS parameters were set according to **Table 2.2**. Collision energies were tuned for each precursor ion and were dependant upon the fragmentation profile required. Data acquisition and instrument control was achieved using MassLynx 4.0. The instrument was calibrated daily using the fragment ions of [Glu^1]-fibrinopeptide B $[M+2H]^{+2}$ (*m/z* = 785.8426) as shown in **Table 2.1**.

2.6.4 2D-RPLC-MS^E

Two-dimensional separations were performed using a nanoAcquity 2D UPLC system (Waters Corporation). For the first dimension separation, 4.5 µl (9.0 µg) of the prepared EB and RB protein lysates containing 90 fmol of an ADH digest (see *section 2.5.1*) were injected onto a 5µm, Xbridge BEH130 C18, 300 µm i.d. x 50 mm (Waters Corporation) column equilibrated in 20 mM ammonium formate, pH 10 (buffer A). The first dimension separation was achieved by increasing the concentration of acetonitrile (buffer B) in 11 steps consisting of 8.2%, 11.7%, 13%, 14.5%, 15.9%, 17.4%, 18.9%, 20.8%, 23.6%, 45%, 65%. At each step the programmed percentage composition was held for 1 min at a flow rate of 2 µl/min and the eluant diluted by buffer C (H_2O + 0.1% (v/v) formic acid) from the second dimension pump at a flow rate of 20 µl/min. This effectively dilutes the ammonium formate and acetonitrile, allowing trapping of the eluting peptides onto a Symmetry C18, 180 µm i.d. x 20mm trapping cartridge (Waters Corporation). After 15 min washing of the trap column, peptides were separated using an in-line second dimension analytical separation performed on a 75 µm i.d. x 200 mm, 1.7 µm, BEH130 C18, column (Waters Corporation) using a linear gradient of 5 to 40% B (buffer A = 0.1% (v/v) formic acid in water, buffer B = 0.1% (v/v) formic acid in acetonitrile) over 90 min with a wash to 85% B at a flow rate of 300 nl/min. All separations were automated and performed on-line to the mass spectrometer.

All mass spectrometry was performed using a Synapt Q-Tof mass spectrometer fitted with a nanolockspray source operating in MS^E mode (Waters Corporation). Data was acquired from 50 to 1990 *m/z* using alternate low and elevated collision energy (CE) scans. Low CE was 5V (Trap), 4V (Transfer) and elevated was 12-35V ramp (Trap), 10V (transfer). The lock mass Glu-fibrinopeptide, ($[M+2H]^{+2}$, *m/z* = 785.8426) was infused at a concentration of 100 fmol/µl at 250 nl/min and acquired every 60 seconds. Operation of the Waters Synapt mass spectrometer was performed in collaboration with Dr Chris Hughes from Waters Corporation.

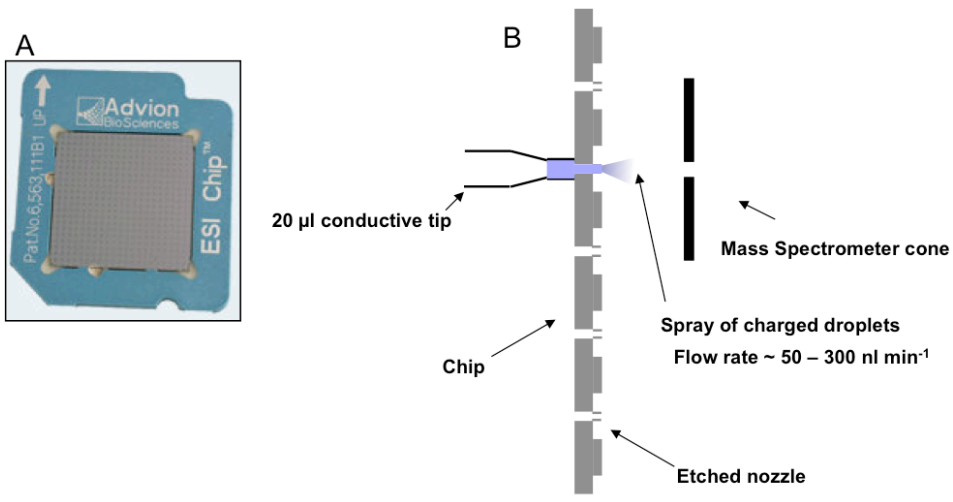


Figure 2.3. The Triversa Nanomate chip system (Advion Biosciences). Panel A shows a photograph of a 400 nozzle chip, while Panel B provides a schematic representation of spray production.

2.7 Data processing

2.7.1 Exclusion list generation

In both the MudPIT (*section 2.3.8*) and GeLC-MS/MS (*section 2.3.7*) analyses performed in this study, the technique of real-time database searching (RTDS) was implemented. This was established in collaboration with the Waters Corporation who provided a developmental version of a real-time database-searching algorithm.

In this process, EB and RB samples were first analysed using a nanoLC-MS/MS experiment without any RTDS as described in *sections 2.3.9.5* and *2.6.1*. In the next stage, the same samples were analysed a second and third time utilising the RTDS algorithm. Exclusion lists were generated 'on the fly' using the RTDS algorithm within the ProteinLynx Global Server 2.05 software suite (Waters Corporation). Once an MS/MS spectrum had been acquired, it was processed to generate a peak list and searched against a protein translation of the *Chlamydia trachomatis*, serovar D genome. This process occurs on a ~30 to 60 sec time scale. During the LC-MS/MS experiment, when three peptides matching a specific protein are assigned, a peptide exclusion list is generated. This exclusion list consists of peptide masses generated 'on the fly' from the *in-silico* trypsin digestion of those proteins meeting the three peptide criteria. The mass spectrometer is automatically updated with the new exclusion lists, excluding the MS/MS analysis of any further peptides from those proteins already assigned by three peptides and so allowing the analysis of further low abundant peptides and hence potentially improving proteome coverage.

2.7.2 Data processing of GeLC-MS/MS, iTRAQ and MudPIT data

MS/MS spectra collected using DDA were processed using either ProteinLynx Global Server (PLGS) version 2.05 for GeLC-MS/MS and MudPIT data, or PLGS version 2.2.5 for iTRAQ data. The following parameters were used for the processing of MS/MS spectra; normal background subtraction with a 25% threshold and medium de-isotoping with a threshold of 1%, no smoothing of the data was performed. Peak lists were generated in a .pkl format ready for database searching.

2.7.3 Data processing of Label-free data

LC-MS^E data were processed using PLGS version 2.3 for submission to the IDENTITY^E search engine (Waters corporation, Milford, MA). The ion detection, clustering and normalization procedures have been described in detail previously (Geromanos *et al*., 2009). Briefly, LC-MS^E spectra were lockmass-corrected, centroided, deisotoped and charge state reduced and intensity measurements adjusted. Each detected component is expressed as an AMRT (accurate mass retention time) pair. These AMRT pairs are used to associate fragments ions with their corresponding precursor ions using the embedded time alignment algorithm. AMRT pairs are then clustered and matched with the same AMRTs across all experiments based upon mass precision and a retention deviation threshold. Once processed, the 11-peak list files corresponding to the LC-MS^E acquisitions performed for each of the 11 fractionation steps were merged. The eight merged

files created, correspond to the two biological and two technical replicates of both EB and RB samples. These were subsequently submitted for database searching to IDENTITY^{E} (*section 2.8.3*).

2.8 Database searches

2.8.1 Database searches of GeLC-MS/MS and MudPIT data

Peak lists in .pkl format were submitted to PLGS version 2.05 and the data searched against a protein translation of a concatenated database consisting of the *C. trachomatis*, serovar D genome (May, 2004) and the NCBI human genome (May, 2004). The data was also subsequently searched against the *C. trachomatis* L2/434/Bu genome (September 2007). Early access to this completed genome sequence was kindly provided by Dr Nicholas Thomson, Wellcome Trust Sanger Institute, Cambridge, UK. The database was publically released in January 2008 (Thomson *et al*., 2008).

A maximum of one missed cleavage for tryptic digestion, a fixed modification for carbamidomethylation of cysteine and a variable modification for the oxidation of methionine were allowed. Precursor and fragment ion tolerances were set at 100 ppm and 0.25 Da, respectively. Protein identifications required the assignment of ≥ 2 different peptides or where identified by only a single peptide, spectra were manually inspected and assigned. Proteins were only assigned if, for each peptide ion, greater or equal to three consecutive experimentally derived *y* or *b* ions could be matched to the predicted spectra. Data was imported into Microsoft Excel 2000. The data was checked for homologous chlamydial and human peptides. No homologous peptides were detected in these datasets.

2.8.2 Protein identification and quantification of iTRAQ data

Peak lists in .pkl format were submitted to the Mascot search engine (Matrix Science, London, UK). Mascot version 2.0 was used to search iTRAQ labeled *S.* Typhimurium samples for the iTRAQ optimisation and iTRAQ parser development studies described in Chapter 3. Mascot version 2.2.1 was used for the chlamydial iTRAQ studies presented in chapter 5. *S.* Typhimurium iTRAQ peak lists were searched against a protein translation of the *Salmonella enterica,* Serovar Typhimurium genome (December 2005). For the chlamydial iTRAQ studies, peak lists were searched against a protein translation of the *C. trachomatis* L2/434/BU genome including L2 plasmid genes (January, 2008) and a concatenated database consisting of the *C. trachomatis* L2/434/BU genome and the NCBI *human* genome (January, 2008).

A maximum of one missed cleavage for trypsin digestion and fixed modifications for methyl methane-thiosulphonation (MMTS) of cysteine and the N-terminus and lysine side chains using the 4-plex iTRAQ label were allowed. Variable modifications for the oxidation of methionine and iTRAQ modification of tyrosine were also allowed. Precursor and fragment ion tolerances were set at 100 ppm and 0.25 Da, respectively. Protein identifications required the assignment of ≥ 2 different peptides with a significance threshold for accepting a match of $p < 0.03$ (≥ 98% confidence). Using Mascot 2.2.1, the chlamydial MS data

was searched against a decoy database of the *C. trachomatis* L2 genome including the L2 plasmid. The FDR of peptides above the identity threshold was estimated to be 3.23%.

For the chlamydial data, .pkl files were merged and protein ratios calculated using Mascot 2.2.1. Peptide ratios were weighted and median normalization performed, automatic outlier removal was chosen and the peptide threshold was set to 'at least homology'. False discovery rates were calculated by searching all spectra against a decoy database using the Mascot software. For the *S.* Typhimurium iTRAQ data, protein assignments were exported as a comma separated values (.csv) file for association to their respective reporter ion ratios generated by the iTRAQ parser. For chlamydial iTRAQ data, merged protein assignments and quantitative data were exported as a .csv file for data curating.

2.8.3 Protein identification and quantification of label-free data

LC-MS^E data was processed as described in *section 2.7.3* and submitted to the PLGS database search algorithm IDENTITY^E version 2.3. In IDENTITY^E, data is further correlated based upon physiochemical properties of peptides in the liquid and gas phase. The search strategy is a three stage iterative process, where each successive iteration incrementally increases the selectivity, specificity and sensitivity of each search. The outline of this process is described briefly in *section 5.3.2* and has been described in detail previously (Geromanos *et al*., 2009; Li *et al*., 2009). Each processed file was searched against a protein translation of the *C. trachomatis* L2/434/Bu genome sequence, including the L2 plasmid sequence and the internal standard alcohol dehydrogenase from *Saccharomyces cerevisiae*. A full 1x reverse database was also appended as a decoy search to determine the peptide false discovery rate. To account for possible host peptides that were homologous to chlamydial peptides, data was also searched against the *C. trachomatis* L2/434/BU database appended with the UniProt *human* database (http://www.uniprot.org, February 2009). Search parameters were as follows: Precursor and product ion tolerance were 10 ppm and 15 ppm respectively. Enzyme specificity was set to tryptic; fixed modifications included MMTS; variable modifications included deamidated asparagine and glutamine, acetylation at the N-terminus and oxidized methionine. A maximum of two missed tryptic cleavages were allowed. A false discovery rate of 4% was applied. Each protein was identified with ≥ 3 different peptides and each peptide was assigned with a minimum of at least 3 fragment ions per peptide.

2.8.3.1 Data normalization and protein quantification of the label-free data

Normalization to obtain absolute concentrations was performed within IDENTITY^E. The average intensity measurement for the three most abundant tryptic peptides for each assigned protein, including the internal protein standard ADH were determined. The average intensity of the 'Top3' most abundant peptides of the internal standard were used to calculate a universal response factor (counts/mol of protein). This response factor was applied to the average intensity of the 'Top3' abundant peptides from all other assigned proteins to determine their absolute concentration, expressed as fmol on column (Silva *et al*., 2006b). The results for each of the 8 samples were exported as a .csv file for further manipulation in Microsoft Excel. Peptides identified as being homologous between *Chlamydia* and human (closest genome with homology to

the African green monkey) were identified and removed using a script written by Dr Richard Edwards. Data was further filtered for protein quantitation by only considering proteins that were identified in at least two replicates of the same chlamydial development form, i.e., EBs or RBs. These could be either technical or biological replicates or both. Proteins assigned in a single replicate present in both EB and RB, were not considered for quantitation, but are reported for qualitative purposes. Protein identifications present in a single replicate are reported for qualitative purposes and reach the criteria of being matched by ≥ 3 different peptides. By using the replication of protein assignments across different injections, the false positive rate is minimized, since chemical noise is random in nature and so does not tend to replicate across injections, so reducing the false positive rate and improving confidence in protein assignment (Vissers *et al*., 2007).

2.8.3.2 Normalization to the number of bacterial cells

Label-free

For the label-free analysis, each sample was normalized on a per bacterium basis. The number of bacteria per µl of EB and RB suspensions used to prepare protein lysates were determined by measuring the number of copies of the *omcB* gene using quantitative PCR, as described in *section 2.2.10*. The chlamydial genome has a single copy of the *omcB* gene. Based upon the assumption that one *omcB* gene is equivalent to one genome, and that one genome is equivalent to a single bacterium, the number of bacterial cells equivalent to the 4.5 ul of digested protein lysate(s) injected on column were determined. Using the Avogadro's number, the number of molecules of each protein were calculated and divided by the number of bacterial cells loaded on column to provide quantitative values for each protein expressed as molecules per cell. Values were calculated for the entire dataset using Microsoft Excel. Protein ratios were calculated and the standard deviation between biological and technical replicates determined for each developmental form. These data are shown in **Appendix 1, Table 5.2**.

iTRAQ

iTRAQ protein ratios for each biological replicate were normalized according to the number of bacterium contained within each sample preparation. The number of bacterial cells in each EB and RB biological replicate were used to calculate a correction factor ratio, which was applied to each protein ratio before calculating the mean ratio and the standard deviation for each protein.

2.8.4 Functional protein categories

Proteins assigned for the chlamydial studies were grouped into functional categories according to the classification originally described by Stephens *et al*, (1998). Proteins were only classified within a single functional group and assigned to each protein via their *loci* using Microsoft Excel.

2.8.5 Protein identifiers

Using the ID mapping tool within UniProt (http://www.uniprot.org/), gi numbers were mapped to UniProt identifiers and exported as an Excel file. CTL (*C. trachomatis* L2 *loci*) and CT (*C. trachomatis*,

serovar D *loci*) numbers were mapped using a *loci* translation table kindly provided by Dr Nicholas Thomson (Wellcome Trust Sanger Institute).

CHAPTER 3

DEVELOPMENT OF QUANTITATIVE PROTEOMIC TECHNOLOGIES

3.0 DEVELOPMENT OF QUANTITATIVE PROTEOMIC TECHNOLOGIES

3. 1 Introduction

3.1.1 Relative quantitation

The ability to compare protein expression levels in cells or tissues in different biological states is an important and fundamental aspect of proteomic research. Traditionally, such measurements have been achieved using 2-D gel electrophoresis coupled with densitometry, providing a relative measure of protein abundance between two different cellular states. However, the approach lacks the required dynamic range, representing only the most abundant proteins, and is not applicable to the analysis of proteins with extreme physio-chemical properties (Gygi *et al*., 2000; Tannu and Hemby, 2006). Further, the relative nature of the comparison means that it is difficult to place a quantitative value on the expression level of a protein when it is not detected in cells in one particular state. It is these limitations that have led to the development of alternative, non-gel based, quantitative technologies.

The majority of these non-gel based methodologies use stable isotope labelling strategies. In these types of proteomic analyses, the difference in mass between pairs of chemically identical analytes of different stable-isotope compositions, are measured in a mass spectrometer. The measured ratio of the signal intensities between the pairs indicates the abundance ratio of the two analytes.

Such methodologies can be divided into two categories. The first called SILAC (Stable Isotope Labelling by Amino acids in Cell culture), is an *'in-vivo'* labelling method (Ong *et al*., 2003; Mann, 2006), utilizing the incorporation of essential amino acids such as [^{13}C]-arginine and [^{13}C]-lysine into proteins within a particular cellular state. By comparing the relative amounts of heavy and light peptides in test and control samples, respectively, it is possible to estimate the protein abundance ratio between the two conditions. Because samples are combined at an early stage, the methodology has the advantage of eliminating errors due to any subsequent downstream processing. However, the requirement for the incorporation of essential amino acids restricts its application; for example, it would be too impractical and expensive to achieve in humans. The second category, the use of external stable-isotope labelling strategies, requires the differential labelling or modification of specific amino acid residues in proteins or peptides. A well-characterised approach is the Isotope-Coded Affinity Tag (ICAT) method. Developed by Gygi and co-workers (1999b), differentially labelled heavy and light tags specific for cysteine residues are used to provide relative quantification of cysteine-containing peptides. Additionally, these tags contain biotin, allowing the selective retrieval of cysteine-containing peptides from non-cysteine containing peptides using affinity purification, greatly reducing the

complexity of the sample and so providing increases in the dynamic range of the mass spectrometric analysis.

All of these methods are based upon an incorporated mass difference to support relative quantification by measurement of the corresponding relative peak areas in either MS or MS/MS (Pan and Aebersold, 2007). Because of this requirement, they are restricted to experiments looking at the difference between two-to-three conditions (3-plex), (Hilger and Mann, 2011). This limits the types of experiments that can be performed; for example, it precludes time-course studies. Although 2-plex studies can be combined post-analysis, not all of the same peptides will be identified in every sample, making comparison and statistical interpretation of the data difficult.

An alternative method, initially allowing the multiplexing of up to four samples (Ross *et al*., 2004) and then subsequently extended to the multiplexing of up to eight samples (Choe *et al*., 2007) has been described. The methodology, termed iTRAQ, is based upon a multiplexed set of isobaric reagents that are specific for primary amines (**Figure 3.1**). Proteins from up to eight different cellular states are first reduced, alkylated and digested with trypsin before labelling with one of eight different tags. The labelled peptides from each cellular state are then combined and nanoLC–tandem mass spectrometry performed. Since the iTRAQ tags are isobaric, in MS mode the masses of the iTRAQ labelled peptides from the eight states are indistinguishable from one another. However, when the labelled peptides are isolated and fragmented in tandem MS mode, each tag generates a unique *m/z* reporter ion in a relatively quiet region of the mass spectrum from 113 – 119 and 121 *m/z*. Protein quantification is achieved by comparing the relative intensities of these reporter ions in the MS/MS spectrum. The corresponding peptides are identified using the fragment ions from the peptide backbone. Because there is potentially a contribution of signal from all eight labelled peptides, improvements in signal to noise ratio are observed leading to more confident peptide identifications. Although the technique produces peptide mixtures that are more complex than those obtained using the cysteine specific ICAT approach. The increased number of peptides improves protein coverage, leading to increased statistical confidence in both protein identification and quantification, outweighing the extra fractionation steps required. Furthermore, as essentially all peptides are labelled, the approach lends itself to the detection of post-translational modifications. An alternative to iTRAQ reagents, are Tandem Mass Tags (TMT), which are based upon the same principle, but have different reporter ion masses (Thompson *et al*., 2003).

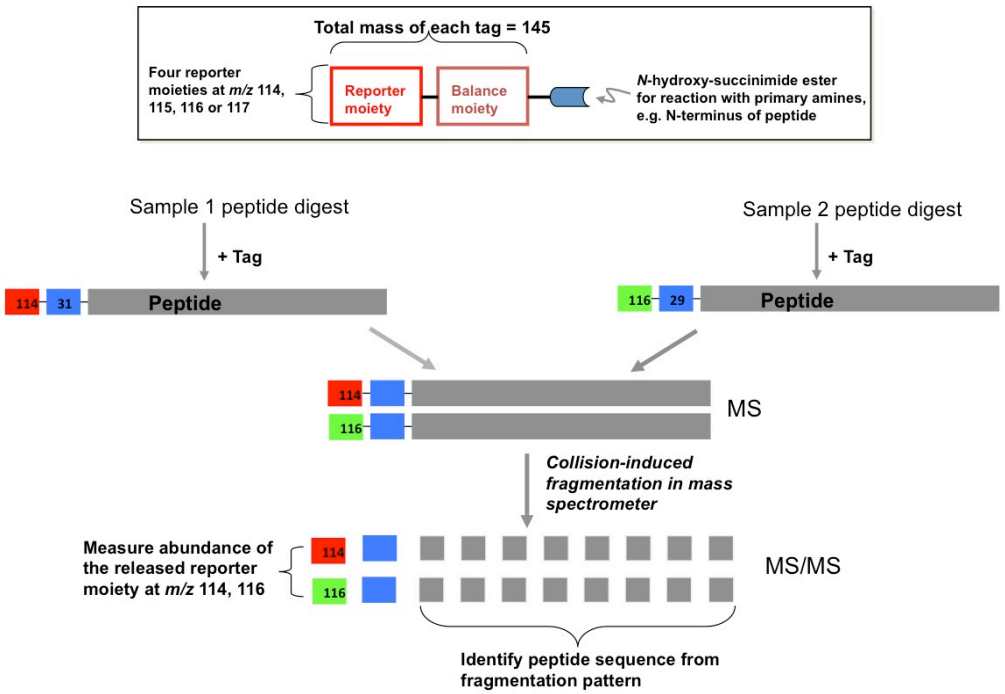


Figure 3.1 iTRAQ tag structure and workflow (after Ross *et al*., 2004).

3.1.2 Absolute quantitation

The majority of routinely used quantitative technologies provide only a relative measure of protein abundance between two cellular states. It is now apparent that this places restrictions on the comparisons that can be made across different experiments. There is therefore a requirement to be able to express the quantity of each protein of a complex system in terms of the number of copies per cell, as opposed to -fold changes obtained in a differential expression type of experiment. Such information represents the 'gold standard' in quantification, allowing quantitative comparisons to be made between different proteins within the same and/or different systems.

Absolute quantification of small organic molecules is considered relatively routine (reviewed by Hoke *et al*., 2001), and is well established in the pharmaceutical industry. The type of experiment often performed is termed multiple reaction monitoring (MRM), and is traditionally performed on a triple quadrupole mass spectrometer, because of the maximum sensitivity achieved when operated in this mode. Here, the first quadrupole (Q1) and the last quadrupole (Q3) are used as a mass filter, isolating a specific precursor *m/z* and monitoring one or more of its corresponding fragment ions. Using prior knowledge about a drugs structure, for example, can be used to predict the precursor *m/z* and a corresponding fragment *m/z* (MRM transition). The fragment ion is then monitored during the chromatographic elution, providing both maximum selectivity and sensitivity. By using an isotopically labelled version of the drug as an internal standard, the ratio of drug to internal standard can be used to calculate the absolute amount of the drug present in the sample. This same approach has in recent years, been applied to the absolute quantification of proteins.

The application of the isotope-dilution method for the quantification of proteins, was first described by Barr (1996), as a solution to replacing immunoassays. This was re- introduced to the proteomics community by Gerber *et al*. (2003). Here they demonstrated that 1-D SDS-PAGE in combination with stable isotope labelled peptide standards and LC-MS/MS, can be used to fractionate and obtain absolute quantitative measurements of proteins and phosphoproteins from complex proteomes such as those of HeLa cells. The procedure, termed AQUA (Absolute QUAntitation of proteins), involved the separation of whole cell lysates of HeLa cells using 1-D SDS-PAGE. Peptides, corresponding to specific tryptic fragments from proteins of interest, were chemically synthesized with an incorporated stable isotope such as [^{13}C]. Samples containing the protein of interest (e.g. slices excised from SDS-polyacrylamide gels) were subjected to limit digestion with trypsin, following the addition of a known amount of the isotopically labelled peptide. NanoLC-MS/MS using Multiple Reaction Monitoring (MRM), was then used to measure the amount of the native tryptic peptide from the protein of interest by comparison with the corresponding internal standard. They demonstrated the approach to be precise (± 5%) and highly specific. For example, they were able to discriminate between phosphorylated and non-phosphorylated protein isoforms. It was also markedly more sensitive than many existing proteomic approaches, demonstrating that <20 μg of HeLa cell lysates was required to quantify the protein separase and its phosphorylation state.

Although conceptually simple, the implementation of such quantitative approaches in a lab with different mass spectrometers is not trivial. Accordingly, this chapter focuses on the establishment and validation of such technologies for quantitative proteomic analysis in Southampton. The implementation of the iTRAQ strategy for quantitative analysis of the proteome from the intracellular bacterium *Chlamydia trachomatis* has required considerable method development. These developments have required the optimisation of sample preparation and fractionation procedures; the optimisation of MS analysis using alternative, as apposed to recommended instrumentation; and the processing of raw iTRAQ data to obtain qualitative and quantitative information.

The second part of this chapter considers the development of absolute quantification methods and their application to the analysis of the *Chlamydia trachomatis* proteome. Initial studies focus on the implementation and validation of the AQUA strategy developed by Gerber *et al*. (2003) using the more accessible and characterised system, *Escherichia coli* (MG1655). Establishing this technology using *E. coli* has allowed the technique to be extended to allow preliminary data for the absolute quantification of proteins expressed in elementary bodies (EBs) and reticulate bodies (RBs), the two developmental forms of *Chlamydia trachomatis*.

As with the completion of a genome sequence, the measurement of absolute quantities of proteins on a proteome-wide scale would represent a major advance. However, to achieve such an ambitious goal requires a substantially increase in throughput. One of the limitations of the AQUA strategy in this respect is the requirement for lengthly chromatography prior to MS analysis. Typically,

the analysis time is ~ 1 h per sample and can take several hours of chromatography method development to establish conditions for each reference or set of reference peptides. The direct analysis of samples within complex mixtures, without recourse to on-line chromatography, would therefore offer several advantages in terms of speed and development time. Accordingly, this chapter also reports a proof of concept strategy developed using automated nanoelectrospray infusion, using a chip-based electrospray system (Nanomate system, Advion Biosciences) in combination with high-resolution mass spectrometry, for the absolute quantification of proteins within simple and potentially complex mixtures, eliminating the requirement for on-line RP chromatography. To demonstrate the potential of this approach, the biological fluid human serum and the model protein myoglobin are used.

3.2 Results

3.2.1 Development and implementation of iTRAQ technology

As a prerequisite to the quantitative analysis of *Chlamydia* using iTRAQ, a suitable experimental workflow for the quantitative analysis of protein lysates using the iTRAQ reagents from Applied Biosystems was implemented using protein lysates from *Salmonella* Typhimurium. *Salmonella* was selected as it offered an easily obtainable source of protein lysates from cells that are readily cultured, compared to the more difficult chlamydial cells, which require a host for growth.

3.2.2 Fractionation of iTRAQ labelled peptide mixtures

Whole cell lysates of *Salmonella* Typhimurium SL1344 were prepared as described in Chapter 2. In brief, 4 x 100 µg samples of protein lysates were reduced, alkylated, proteolytically digested using trypsin and labelled using the 4-plex iTRAQ reagents as described in *section 2.3.9.3*. Lysates from two batches of unstressed cells were labelled using the 114 and 116 iTRAQ reagents, respectively. Similarly, lysates from two batches of cells osmotically stressed using 0.5 M NaCl were labelled with the 115 and 117 reagents. Preliminary mass spectrometric analysis (data not shown) indicated that the labelled peptide extract was highly complex, requiring fractionation to increase proteome coverage. Therefore, as part of the iTRAQ workflow, two fractionation columns were evaluated.

Strong-cation-exchange (SCX) chromatography is a commonly used fractionation technique for proteomics and was chosen for fractionation of the iTRAQ labelled peptide mixture. Initial separations were performed using the ProPac® 250 mm x 1 mm ID SCX column (Dionex), in conjunction with a sodium phosphate buffer pH 3.0 and NaCl gradient (0 – 500 mM) as described in *section 2.3.9.4*. **Figure 3.2** shows a UV absorbance chromatogram of 400 µg (4 x 100 µg) of iTRAQ labelled peptides loaded on column. With the exception of the large UV absorbance signal obtained between 0 and 12 min, no peptide elution was observed during the 50 min separation gradient. This suggested either no effective binding of the labelled peptides to the column and hence the peptides were eluted along with the injection artefact, or that peptides were bound with high affinity to the column, with up to 500 mM NaCl not being of sufficient ionic strength to elute peptides from the column. Subsequent MS analysis revealed that the iTRAQ labelled peptides were contained within the injection artefact. Maintaining the pH of the mobile phase buffers at pH 3.0 is essential to ensure that the peptides are positively charged, allowing effective binding to the sulphonic acid groups of the SCX column packing material. To be certain that the pH was maintained at 3.0 during the separation gradient, the experiment was repeated using 10 mM phosphoric acid in combination with a 0 – 500 mM NaCl gradient. Similarly, no peptide separation was observed.

In light of the results obtained using the Dionex ProPac column, an alternative SCX column from the manufacturer Phenomenex was evaluated. iTRAQ labelled peptides were separated using a

Luna, 5 µm, SCX, 100 Å, 250 x 4.6 mm ID column (Phenomenex), equilibrated in 10 mM phosphoric acid. Peptide elution was achieved using a NaCl gradient (0-500 mM) as detailed in *section 2.3.9.4*. Using this column and buffer system, iTRAQ labelled peptides were separated to provide sufficient fractionation of the sample for further downstream analysis. A UV absorbance chromatogram of the separated iTRAQ peptides is shown in **Figure 3.3**. Since this column provided sufficient fractionation of the iTRAQ labelled peptides, all subsequent iTRAQ experiments, as presented in Chapter 5 were performed using the Phenomenex Luna SCX column.

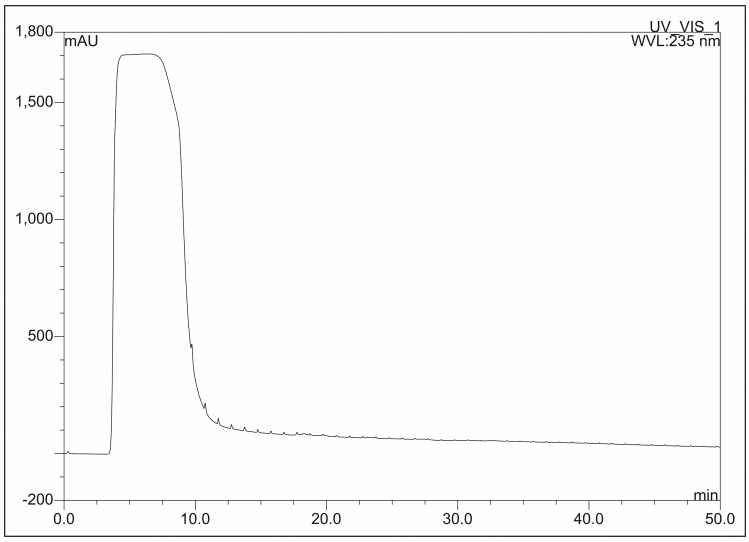


Figure 3.2. A UV chromatogram of a SCX separation of an iTRAQ labelled peptide extract from *Salmonella* Typhimurium using a ProPac® 250 mm x 1mm ID SCX-10 column (Dionex). Separation was achieved using sodium phosphate pH 3 and a NaCl gradient (0-500 mM) at 0.2ml/min over 50 min. UV absorbance was monitored at 235 nm.

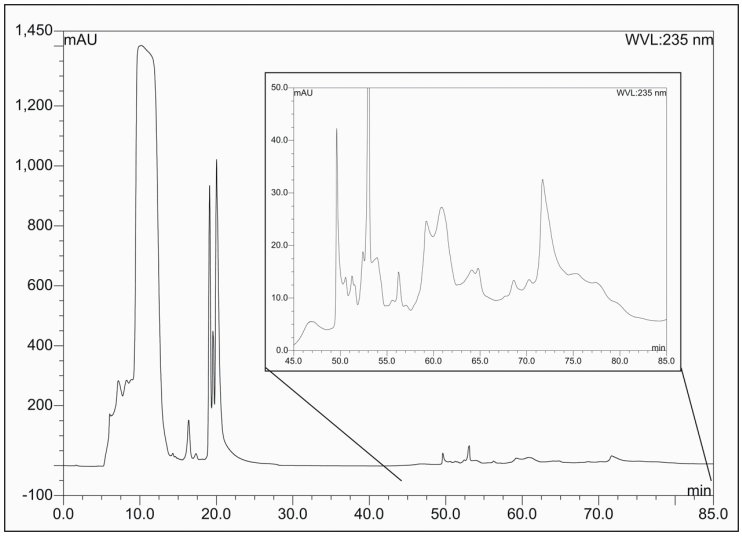


Figure 3.3. A UV absorbance chromatogram of a SCX separation of an iTRAQ labelled peptide extract from *Salmonella* Typhimurium using a 250 mm x 4.6 mm ID Luna SCX column (Phenomenex). Separation was achieved using 10 mM phosphoric acid and a NaCl gradient (0-500 mM) at 0.2 ml/min. UV absorbance was monitored at 235 nm. The inset provides an expansion of the region where peptide elution occurred. Fractions were collected at 1 min intervals.

3.2.3 Extraction of quantitative information from iTRAQ data

At the initiation of these studies, the iTRAQ™ approach had been developed primarily for use with mass spectrometry instrumentation and software developed by Applied Biosystems (ABI) (Ross *et al*., 2004). The ABI software package ProQuant (ABI), provided both quantitation and identification of iTRAQ labelled samples using MS data submitted in proprietary ABI formats. As such, the analysis of iTRAQ labelled samples using MS instrumentation and software from the manufacturer Waters (formerly, Micromass), presented some challenges. These included the optimisation of instrument parameters, peptide assignment and in particular, the extraction of iTRAQ reporter ion information.

The assignment of proteins from iTRAQ labelled peptides using alternative protein identification tools such as, ProteinLynx Global Server and Mascot, can easily be achieved. However, obtaining quantitative information using iTRAQ, at the time, was restricted to ProQuant (ABI). In collaboration with the School of Electronics and Computer Science (James Rodger, Dr Simon Miles, University of Southampton), a parser, written in Java™ was developed to extract the reporter ion intensities from the MS/MS spectrum of each fragmented iTRAQ labelled peptide ion.

Before raw MS/MS data can be submitted to PLGS or Mascot for searching, they require pre-processing to generate a peak list. As part of the PLGS software suite, peak lists can be generated in the .pkl file format. This .pkl file contains on one line, the *m/z* value of a peptide ion, along with its corresponding ion intensity and charge state. This is followed by a list of fragment ion *m/z*'s and their corresponding ion intensity values, including the reporter ion *m/z* and intensities.

The input of the iTRAQ parser is a .pkl file, which parses the following information into a comma-separated-variable (CSV) file: .pkl filename, charge state, peptide ion *m/z*, the reporter ion *m/z* of 114.1±0.1, 115.1±0.1, 116.1±0.1, 117.1±0.1 and their corresponding ion intensities. The mass window of ±0.1 was established to ensure that the reporter ion peak was captured. This was achieved using an iterative process, by manual inspection of the reporter ions in the spectra and output file. The Graphical User Interface (GUI) of the iTRAQ parser is shown in **Figure 3.4**.

Figure 3.4. iTRAQ Parser - Screenshot of the Graphical User Interface.

Peptides are assigned by submission of the same set of .pkl files to the Mascot search engine and the results exported as a .CSV file. This reporter ion intensity information can then be integrated with the peptide assignment data using the .pkl filename and corresponding peptide ion mass (to 4 decimal places) by manipulation in Microsoft Excel. The data is then filtered and processed as follows:

Thresholding

Ratios with intensities <20 counts were removed since low ion counts close to background can cause errors. A value of 20 was chosen, since this is about 10 x the ion counts obtained from background in this reporter ion region.

Corrections for isotope overlap

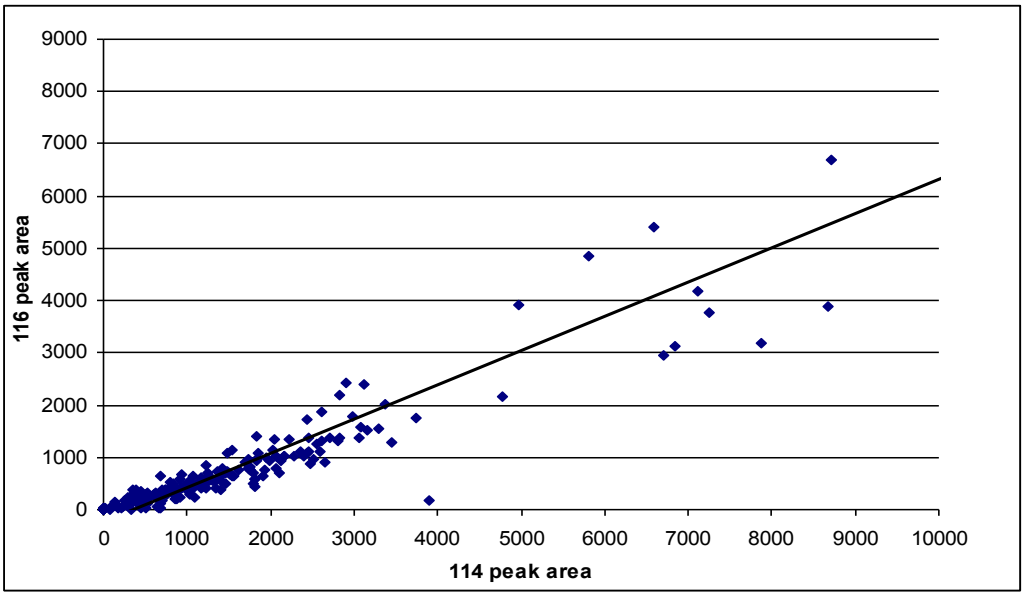
Although small (< 3% for 4-plex; <1% for 8-plex), correction factors to account for isotope overlap from the natural isotope abundance (+1, +2 Da) and from incomplete enrichment at any carbon or nitrogen (-1, -2) were applied. The information for these purity correction values is provided with the iTRAQ reagent kit.

Intensity weighting

An intensity weighting was applied by summing the reporter ion intensities in any given channel (114, 115, 116, 117, etc.) from all peptides arising from a given protein. This provides a natural way of the most intense spectra contributing more to the protein ratio, whereas contributions from weaker ions closer to background are minimized.

To test the ability of the parser to successfully extract iTRAQ reporter ion data, *Salmonella* protein lysates were labelled with the 114 and 116 iTRAQ reagents. 100 µg of a *Salmonella* protein lysate was labelled with the 114 iTRAQ reagent and 50 µg of the same sample labelled with the 116 iTRAQ reagent (*section 2.3.9.3*) and LC-MS/MS analysis performed without SCX fractionation (*section 2.3.9.5*). Using the iTRAQ parser, the reporter ion intensities were extracted and plotted as shown in graphs A and B of **Figure 3.5.** Correlation of the data with the expected 2:1 differential expression ratio was reasonable with the exception for a few outlying values (**Figure 3.5, A**). Upon further analysis these outliers were identified as peptide autolysis products from the protease bovine trypsin, used for digestion of the protein lysates. Exclusion of these data points from the dataset showed that there was a good correlation with the expected 2:1 labelling ratio (**Figure 3.5, B**).

A



B

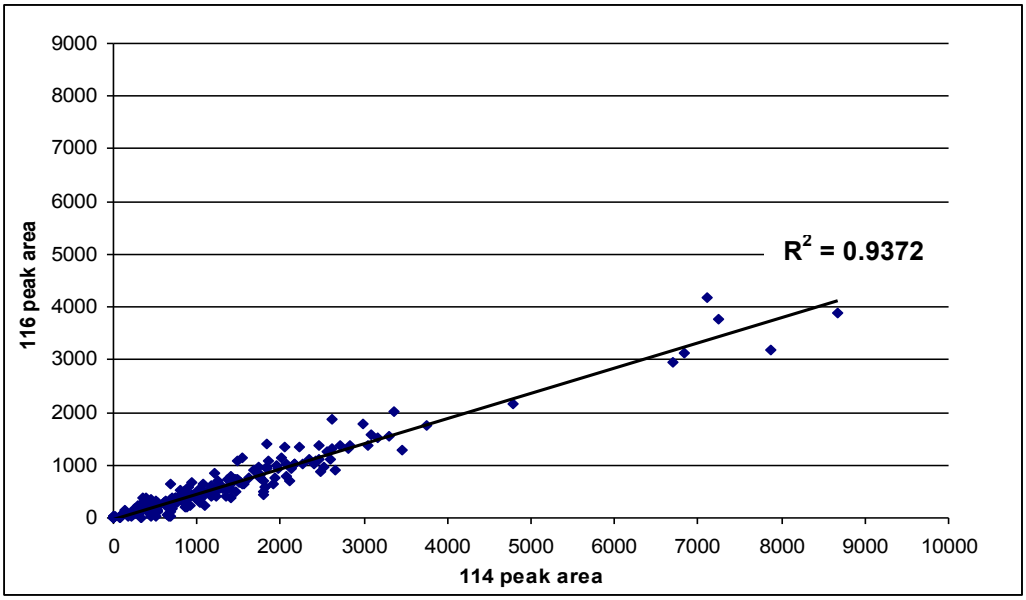


Figure 3.5. iTRAQ reporter ion intensities of peptides from *Salmonella* lysates labelled with the 114 and 116 iTRAQ reagents in a 2:1 ratio. 114 reagent = 100 µg of lysate, 116 reagent = 50 µg of lysate. Graph A represents the labelled *Salmonella* lysates including the data points for the autolysis products of trypsin; Graph B represents the labelled *Salmonella* lysates excluding data points from the trypsin autolysis products. This data was in good agreement with the expected 2:1 ratio with a correlation coefficient of 0.9372. Data was extracted using the iTRAQ parser written at Southampton University.

3.3 Implementation and validation of the AQUA strategy

To establish and validate a methodology for the absolute quantitation of proteins, the AQUA strategy (Gerber *et al*., 2003) was used to determine the expression levels of the regulatory protein

BipA during the growth of the bacterium *Escherichia coli*. The AQUA strategy has two stages. The first of these is the design and synthesis of an internal peptide standard and its validation; the second is development of the assay and implementation.

3.3.1 Internal peptide standard selection and validation

Peptide selection is based upon a number of criteria which includes: the length of the peptide, normally selected to be between approximately 5 and 15 amino acids for the ease of synthesis; the presence of an amino acid that can easily and cost effectively be replaced with a stable isotopic equivalent such as leucine, isoleucine or phenylalanine; and the lack of cysteine or methionine residues, since these are easily modified by chemical groups or oxidised within the mixture, leading to uncharacterised mass shifts. There is also a preference for the amino acid arginine at the C-terminus, as these peptides tend to have improved ionisation efficiencies over peptides with a C-terminal lysine. Importantly, peptides should be unique to the protein under study in their respective organism. However, synthesising peptides targeted at conserved regions of a protein can be valuable when studying the same protein in several different organisms, thereby reducing the number of peptides required to be synthesised.

The following peptide, L*-D-Y-V-I-P-S-R (where * indicates a Leucine containing 6 x [^{13}C] atoms) targeted for the measurement of the protein BipA (residues 440 to 446), was selected according to the above criteria and chemically synthesised (Peptide Research Products). This provided a peptide that was chemically identical to its native counterpart, but was six actual mass units heavier. Amino acid analysis was performed to determine the absolute concentration of the synthetic peptide (Alta Biosciences).

An MS/MS infusion experiment of the internal peptide was performed as described in *section 2.5.3* to identify a suitable fragment ion that could be used for monitoring the abundance of the BipA peptide. The signal intensity of the *y*4 fragment ion from the doubly charged precursor ion, *m/z* = 484 (**Figure 3.6**), was maximised by optimisation of the collision energy. Using this optimised collision energy, an LC-MS/MS experiment was performed using 500 fmol of the peptide to determine its retention time (rt = 16.30 min) and MS/MS profile.

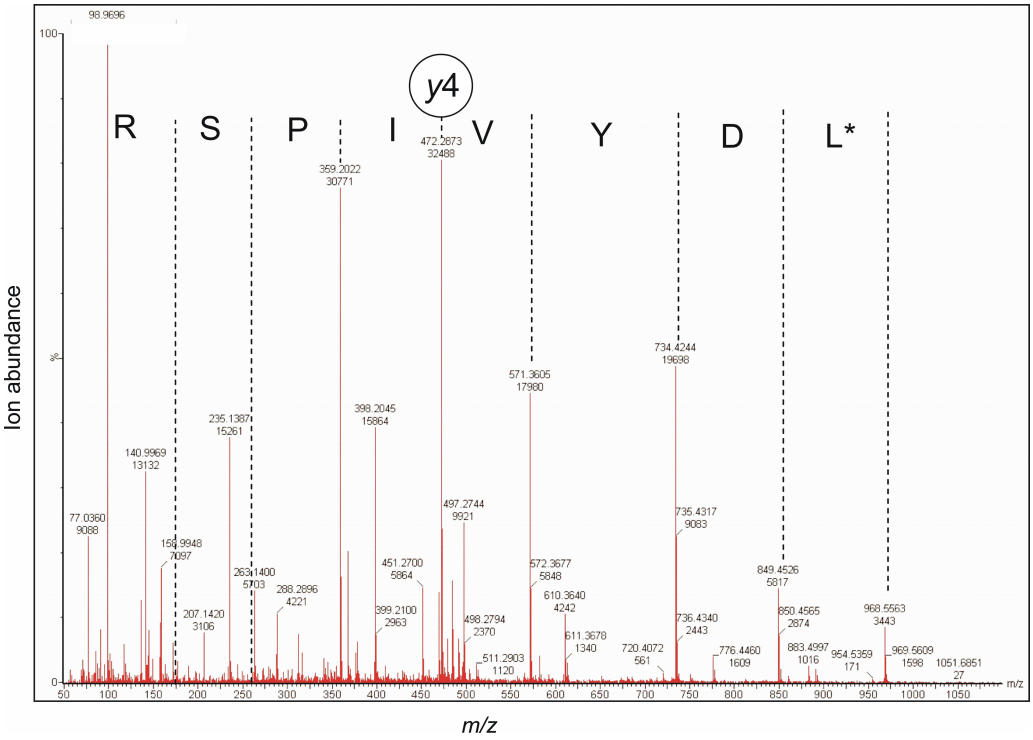


Figure 3.6. MS/MS fragmentation spectrum of the AQUA peptide standard selected for the absolute quantitation of the protein BipA from *E. coli.* Collision energy was tuned to 15 eV to maximise the signal intensity for the *y*4 ion, *m/z* = 472.28. Residues 440 to 446. L* indicates a Leucine containing 6 x [^{13}C].

3.3.2 Absolute quantification of the protein BipA during growth of *E. coli* K-12

Samples of culture were taken at set time intervals during the growth of *E. coli* K-12 (MG1655*)* and the number of cells estimated using a haemocytometer and A_{600} as detailed in *section 2.2.5*. The harvested cells were separated by 1D SDS-PAGE, gel bands excised corresponding to the molecular mass of the BipA protein ± 15 kDa (molecular mass = 67355 Da) (**Figure 3.7**) and in-gel digested in the presence of 500 fmol of the BipA isotopically labelled reference peptide, as detailed in *section 2.4.1*. An LC-MS/MS experiment was performed isolating the precursors *m/z* 484 and 481 ions, corresponding to the doubly charged ion of the reference peptide and native peptide, respectively (**Figure 3.8**). Using reconstructed ion chromatograms of the *y*4 fragment ion (*m/z* = 472.28), the peak areas were used to calculate the concentration of the native peptide and hence the BipA protein (**Figure 3.9**). Using the calculated number of cells loaded onto the SDS-PAGE gel, the number of molecules of BipA per cell, at each time point during the growth of *E. coli* K-12 was determined.

BipA expression was maximal in cells in the early exponential phase of growth (~18,000 molecules/cell), but decreased over 25-fold (to ~700 molecules/cell) as cells progressed into the stationary phase. These changes correlated well with immunoblotting experiments conducted in parallel, where BipA was also shown to decrease with progressive cell growth (see inset **Figure 3.8**). These immunoblotting experiments were performed in collaboration with Mrs Karen Platt.

It was concluded that the AQUA strategy accurately reflected the growth profile of the BipA protein. Further, in this particular case, the limit of detection was 700 molecules per cell and was comparable to the detection limits of the semi-quantitative immunoblotting procedure used for comparison.

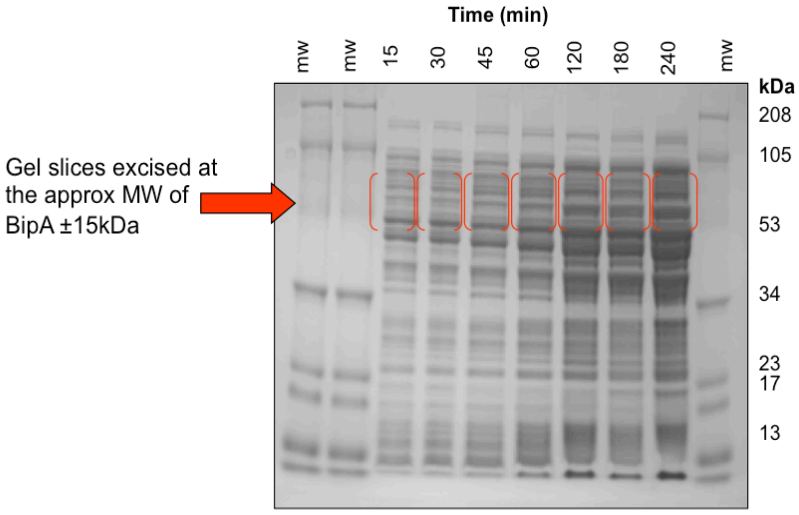


Figure 3.7. SDS-PAGE image of crude cell extracts from *E. coli* (MG1655) extracted at different time points during growth and separated for subsequent AQUA analysis of BipA. Colloidal Coomassie blue staining; molecular weight markers are indicated on the right hand side.

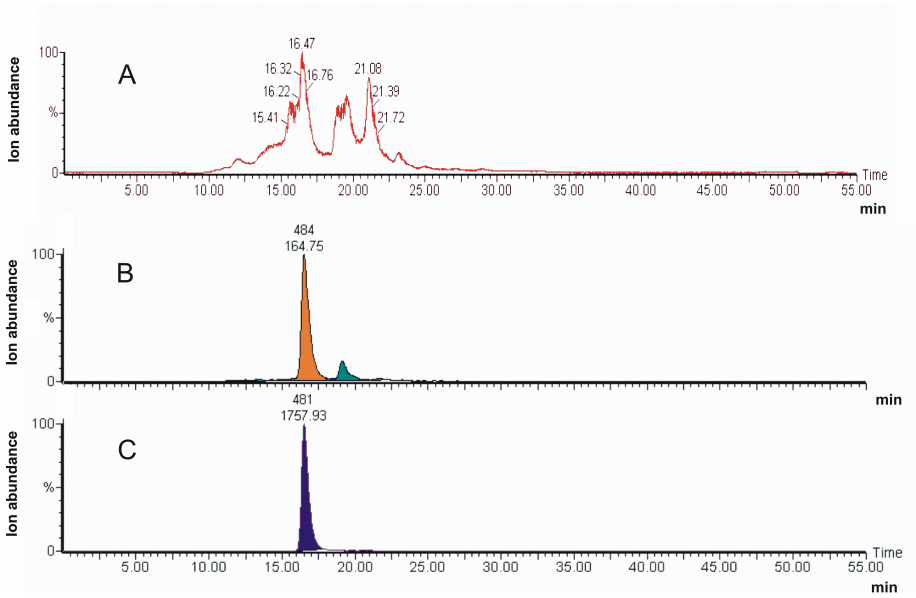


Figure 3.8. AQUA analysis of the protein BipA from *E. Coli* (MG1655) after 15 min of growth. A = LC-MS/MS ion chromatogram of the native BipA peptide ion, *m/z* = 481. B = Reconstructed ion chromatogram of the *y*4 ion (*m/z* = 472.2) from MS/MS of the $[M+2H]^{2+}$,6 x $[^{13}C]$ labelled reference peptide ion (*m/z* = 484) . C = Reconstructed ion chromatogram of the *y*4 ion (*m/z* = 472.2) from MS/MS of the $[M+2H]^{2+}$, native peptide ion (*m/z* = 481).

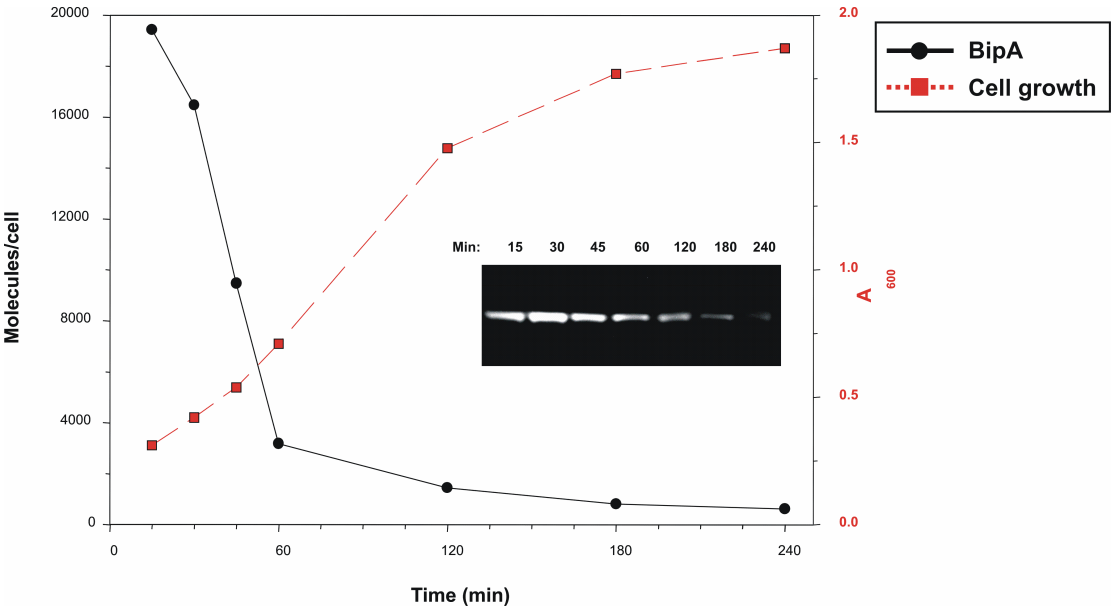


Figure 3.9. BipA synthesis is growth regulated. The number of molecules of BipA per cell measured using the AQUA approach during growth of *E. coli* (MG1655). Inset: Immunoblot of samples from the same time points probed with BipA-specific antibodies.

3.3.3 Absolute quantification of the Major Outer Membrane protein and a putative metalloprotease from *C. trachomatis.*

As a further test of the AQUA strategy, it was used to quantify two proteins from *C. trachomatis,* serovar L2, identified from qualitative proteomic analyses performed and reported in Chapter 4 of this thesis. These two proteins were found to be present in both EBs and RBs, the two developmental forms of *C. trachomatis*. The first of these was the Major Outer Membrane Protein (MOMP) (CTL0050/CT681), a protein that has been reported to be in relatively high abundance in both EBs and RBs (Hatch *et al*., 1981; Salari and Ward, 1981; Caldwell *et al*., 1981). The second protein selected (CTL0328/CT072) was originally annotated as a putative protease (Stephens *et al*., 1998), and has subsequently been re-annotated as a metalloprotease (YaeL) in the *C. trachomatis* L2 genome (Thomson *et al*., 2008).

In brief, 30 µg of protein lysates from both EBs and RBs were separated using SDS-PAGE and gel slices corresponding to the molecular mass of MOMP (42,550 kDa) and the metalloprotease (69,271 kDa) excised ± 15 kDa. AQUA analysis was performed for both proteins in EBs and RBs as detailed in *section 2.4.2* and spiked with 500 fmol of the relevant isotopically labelled reference peptides (**Figure 3.10**). The sequences of the reference peptides were G-Y-V-G-Q-E-F-P-L*-D-L-K (MOMP) and I-S-L*-G-I-P-L-K (metalloprotease) (where * indicates a Leucine containing 6 x [^{13}C]), corresponding to amino acid residues 242 to 251 and 479 to 486, respectively. These peptides were selected based upon the proteotypic data generated in the qualitative studies presented in Chapter 4.

Figures 3.11 and **3.12** show LC-MS/MS ion chromatograms of the $[M+2H]^{2+}$ precursor ion from the native metalloprotease peptide (*m/z* = 423.7) along with reconstructed ion chromatograms for the selected *y* fragment ions of both native and reference peptides (y7, *m/z* = 727.6 and *y*7*, *m/z* = 733.6, respectively) (see **Figure 3.10**) for both EBs and RBs. Peak areas from the reconstructed ion chromatograms for both the native and reference peptide were used to calculate the protein concentrations. The concentration of the metalloprotease in EB samples was 210 fmol per 30 µg of protein loaded onto an SDS-PAGE gel and more than 3.3 -fold greater in RBs, with a concentration of 693 fmol per 30 µg of protein confirming the expression of this protein in both EBs and RBs.

In contrast, the concentration of the highly abundant MOMP was unable to be determined in either EBs or RBs due to effects caused by column overloading and possible saturation of the MS detector. Therefore, a repeat analysis of the sample at a diluted concentration would be required. Since these analyses consume significant amounts of the isolated EB and RB preparations, the analysis was not repeated and the samples were retained for the studies presented in Chapter 5.

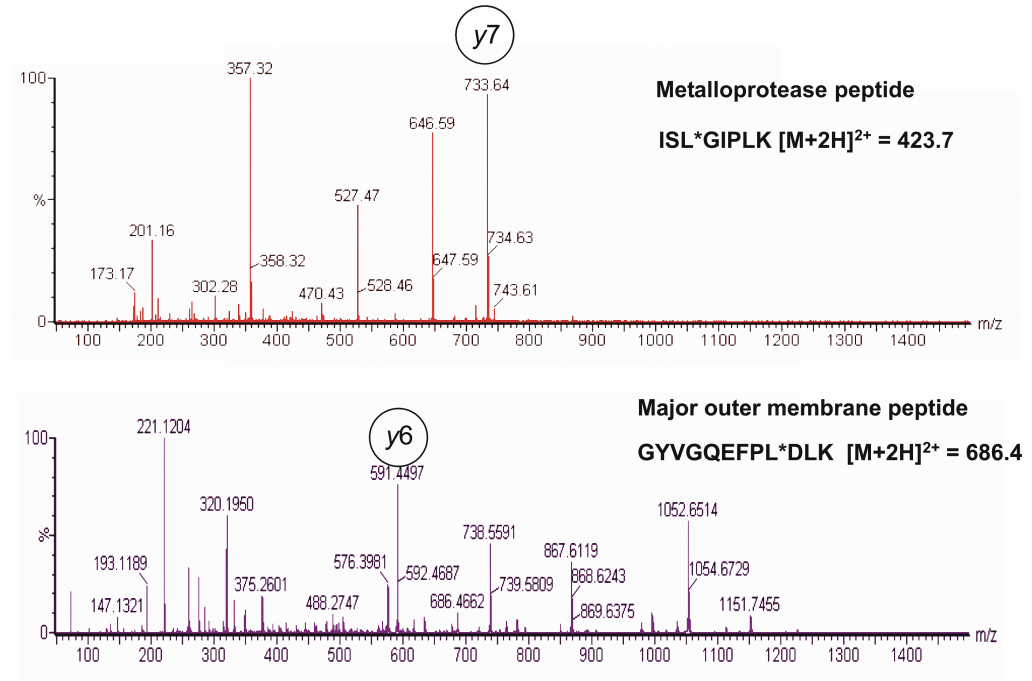


Figure 3.10. MS/MS fragmentation spectra of the AQUA reference peptides selected for absolute quantitation of the metalloprotease (CTL0328/CT072) and the Major Outer Membrane Protein (CTL0050/CT681) from *C.trachomatis* L2. A collision energy of 17 eV was used to maximise the intensity of the y7 ion (*m/z* = 733.6) for the Metalloprotease peptide, and 22 eV for maximising the intensity of the *y*6 ion (*m/z* = 591.4) of the Major Outer Membrane Protein. L* indicates a Leucine containing six x [^{13}C].

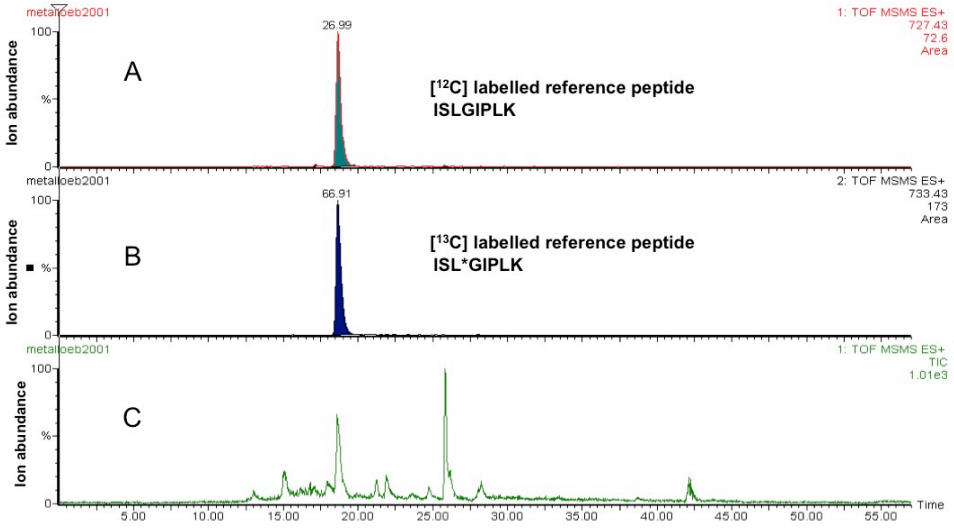


Figure 3.11. AQUA analysis of the metalloprotease protein in EBs from *C. trachomatis,* serovar L2.
A = Reconstructed ion chromatogram of the *y*7 ion (*m/z* = 727.4) from MS/MS of the [M+2H]$^{2+}$, native peptide ion (*m/z* = 420.4). B = Reconstructed ion chromatogram of the *y*7 ion (*m/z* = 733.4) from MS/MS of the [M+2H]$^{2+}$ [^{13}C] labelled reference peptide ion (*m/z* = 423.4). C = LC-MS/MS total ion chromatogram of the native metalloprotease peptide ion, *m/z* = 420.4. Using the peak areas of the native and reference peptide, a calculated expression level of 210 fmol per 30 µg of protein was obtanined. L* indicates a leucine containing 6 x [^{13}C].

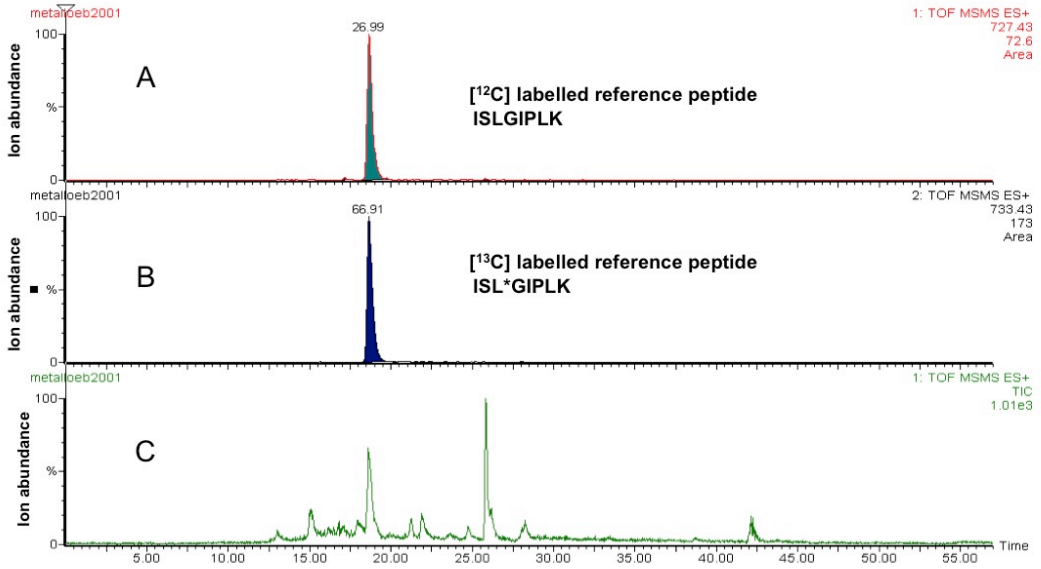


Figure 3.12. AQUA analysis of the metalloprotease protein in RBs from *C. trachomatis,* serovar L2. A = Reconstructed ion chromatogram of the *y*7 ion (*m/z* = 727.43) from MS/MS of the $[M+2H]^{2+}$ native peptide ion (*m/z* = 420.4). B = Reconstructed ion chromatogram of the *y*7 ion (*m/z* = 733.4) from MS/MS of the $[M+2H]^{2+}$ [^{13}C] labelled reference peptide ion (*m/z* = 423.4) . C = LC-MS/MS total ion chromatogram of the metalloprotease reference peptide ion, *m/z* = 420.4. Using the peak areas of the reference and native reference peptide, a calculated expression level of 693 fmol per 30 µg of protein was obtained. L* indicates a leucine containing 6 x [^{13}C].

3.3.4 Development of an infusion based approach for absolute quantification

To increase the throughput of the absolute quantitation studies, it was of interest to explore the possibility of using an infusion based approach that circumvents the requirement for lengthy LC separations. This possibility has become more realistic with the introduction of the Nanomate chip technology (Advion Biosciences), which allows efficient infusion of small volumes of samples via nanoelectrospray into a mass spectrometer in an automated fashion. The signal averaging that accrues as a result of infusing sample over time, results in improved S/N ratios and hence increases the sensitivity and precision with which peptides can be measured. As a pre-requisite to the analysis of more complex mixtures (*section 3.3.6*), the approach was first evaluated using a relatively simple mixture, across a wide concentration range.

3.3.5 Absolute quantitation of a protein in a simple mixture

An internal reference peptide corresponding to a tryptic peptide of horse heart myoglobin, A-L-E-L*-F-R (residues 135 to 140) was chemically synthesised, replacing the amino acid leucine (indicated by *) with a stable isotope of leucine containing 6 x [^{13}C] (Peptide Research Products). The concentration of this peptide was determined by amino acid analysis (Alta Biosciences). Digests of the protein myoglobin from horse heart were prepared in the concentration range from 500 amol to 15 pmol with the myoglobin internal

reference standard added to a final concentration of 1 pmol in a volume of 10 µl as detailed in Chapter 2. The sample was infused using the Triversa Nanomate (Advion Biosciences) into a Q-TOF Micro fitted with a z-spray source (Waters) at a flow rate of ~ 200 nl/min. The samples were each infused over 30 minutes and tandem mass spectrometry performed as described in *section 2.4.3*. **Figure 3.13** shows an MS spectrum of the myoglobin digest at a concentration of 500 fmol and an expanded region of the native peptide *m/z* under measurement. A typical MS/MS spectrum for the native (**A**) and internal reference peptide (**B**) is shown in **Figure 3.14**. The peak intensities of the *y*4 fragment ions *m/z* = 564.3 and 570.2, corresponding to the native and reference peptides, respectively, were used to calculate ratios for quantitation. The x and y values used for the linear regression analysis in **Figure 3.15** represent the expected concentration of myoglobin (x-axis) and the observed concentration of myoglobin (y-axis). The statistics associated with the linear regression are indicated on the calibration curve. For an ideal system, the slope of the line should be equal to unity and intersect through the origin. The observed slope of the line was 1.0807 with an intercept of +0.1058. The r^2 value calculated for the linear regression was 0.9977 indicating a linear response over the concentration range evaluated. The inset graph represents the lower concentrations of the response curve (500 amol to 250 fmol) and had an observed slope of 1.2104 and an intercept of -0.0008, r^2= 0.9989, demonstrating linearity at low concentrations. The lower limit of detection (LOD) was 500 amol for the quantification of the purified myoglobin.

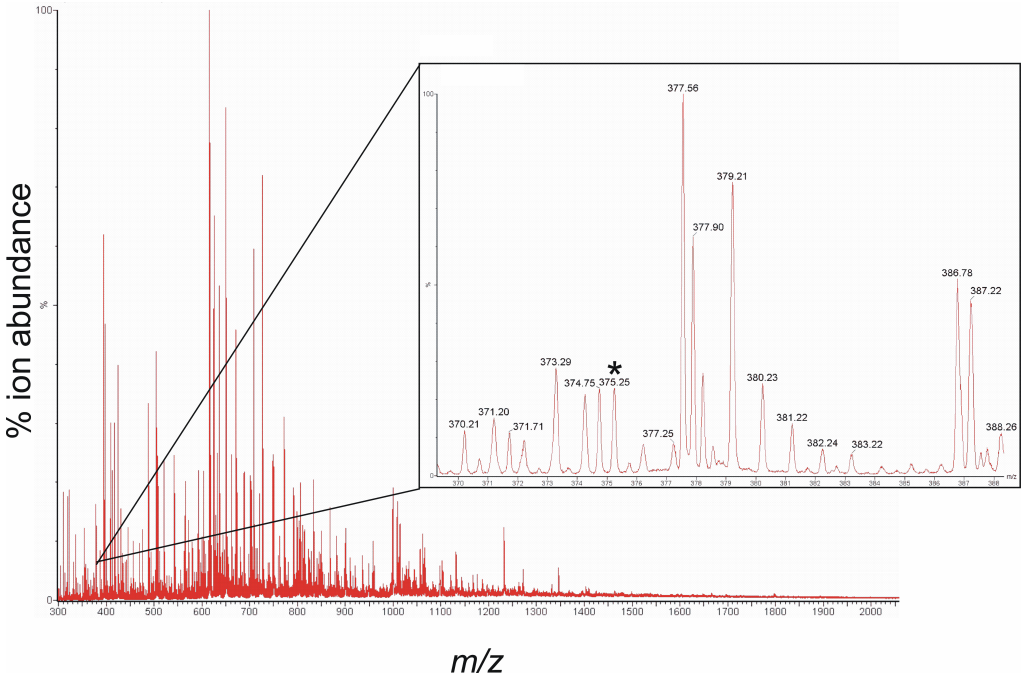


Figure 3.13 An MS spectrum from summed scans over a 2 min infusion of 500 fmol of a horse heart myoglobin digest using the Nanomate system (Advion Biosciences). The inset shows an expanded region of the *m/z* spectrum indicating (*) the native peptide ion used for quantitation by MS/MS ($[M+H]^{2+}$ *m/z* = 375.3).

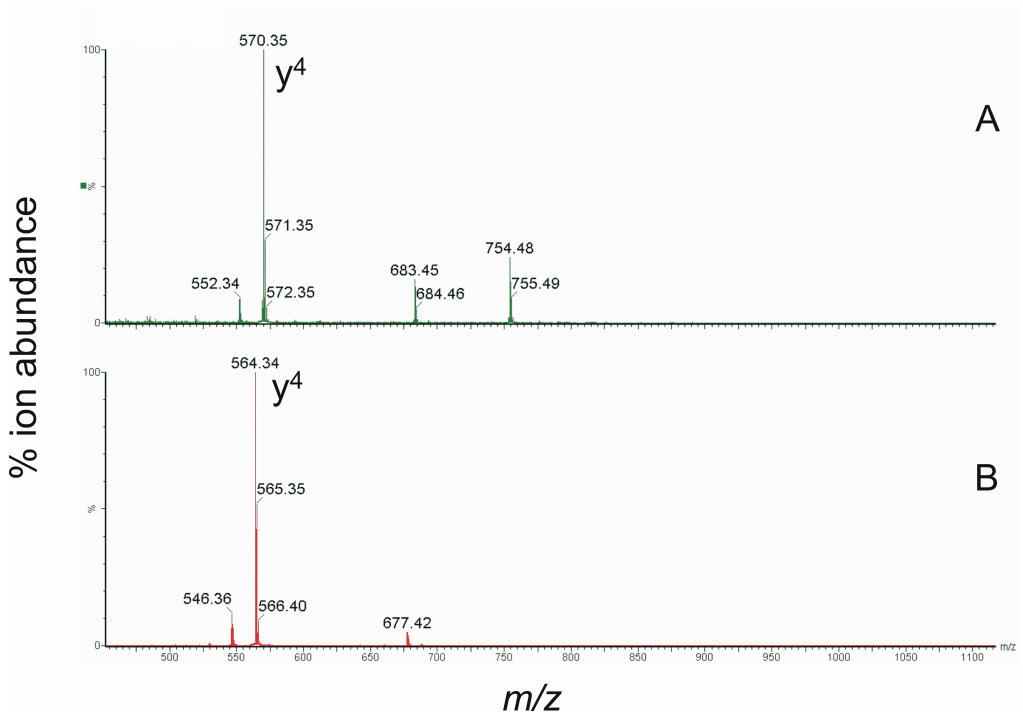


Figure 3.14. MS/MS spectra of (A) the AQUA reference peptide for horse heart Myoglobin and (B) the corresponding native peptide from myoglobin. A collision energy of 14 eV was used to obtain maximum signal intensity for the y4 fragment ion (A) [M+2H]2+ 378.2 m/z to y4 570.2 m/z; (B) [M+2H]2+ 375 m/z to y4 564.2 m/z.

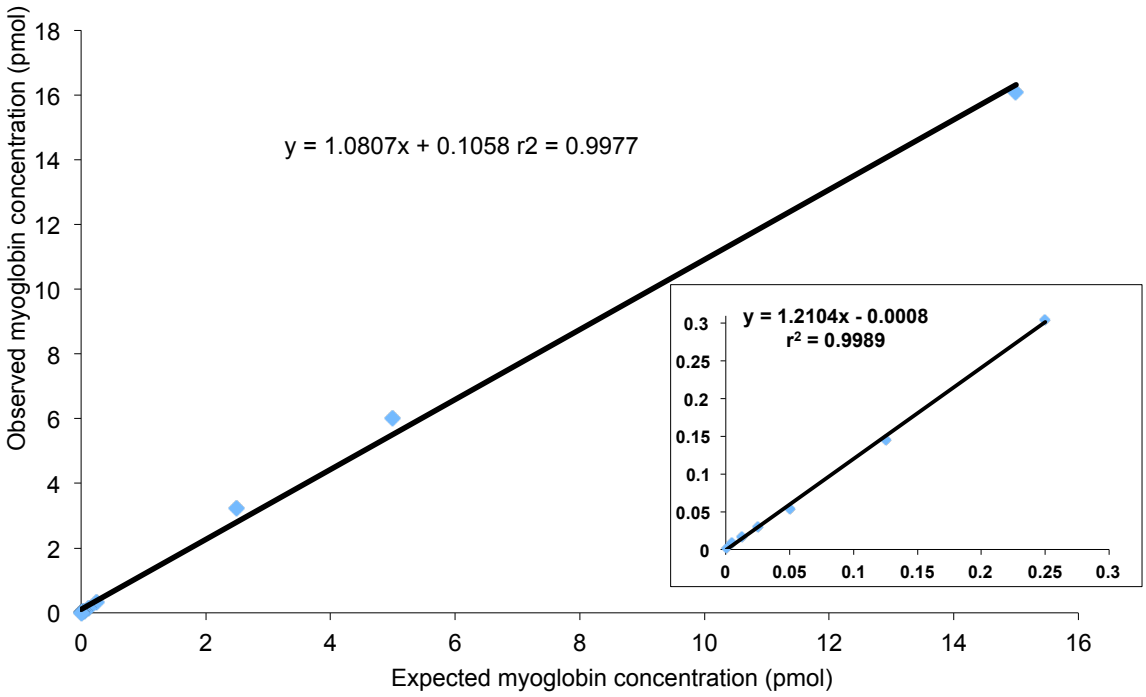


Figure 3.15. Observed vs. expected response curve for the quantification of myoglobin from 500 amol to 15 pmol by nanoESI-infusion using the Triversa Nanomate. Inset = the low concentration range (500 amol–250 fmol).

3.3.6 Absolute quantitation of a protein in a complex mixture

To test the approach with a significantly more complex mixture, human serum was spiked with 1 pmol of myoglobin reference peptide and a digest of horse heart myoglobin at a range of concentrations (1 fmol - 15 pmol) in a final volume of 10 µl. Samples were infused over 30 min and tandem mass spectrometry performed as described in *section 2.4.4*. An MS spectrum of digested human serum in 50% methanol + 1% acetic acid is shown in **Figure 3.16**. The observed vs. expected response curve for the quantitation of myoglobin in a background of human serum is shown in **Figure 3.17**. The r^2 value = 0.9972 across the full concentration range evaluated with a value of 0.995 in the low concentration range from 10 fmol to 500 fmol. The LOD for myoglobin in the complex mixture human serum was 10 fmol.

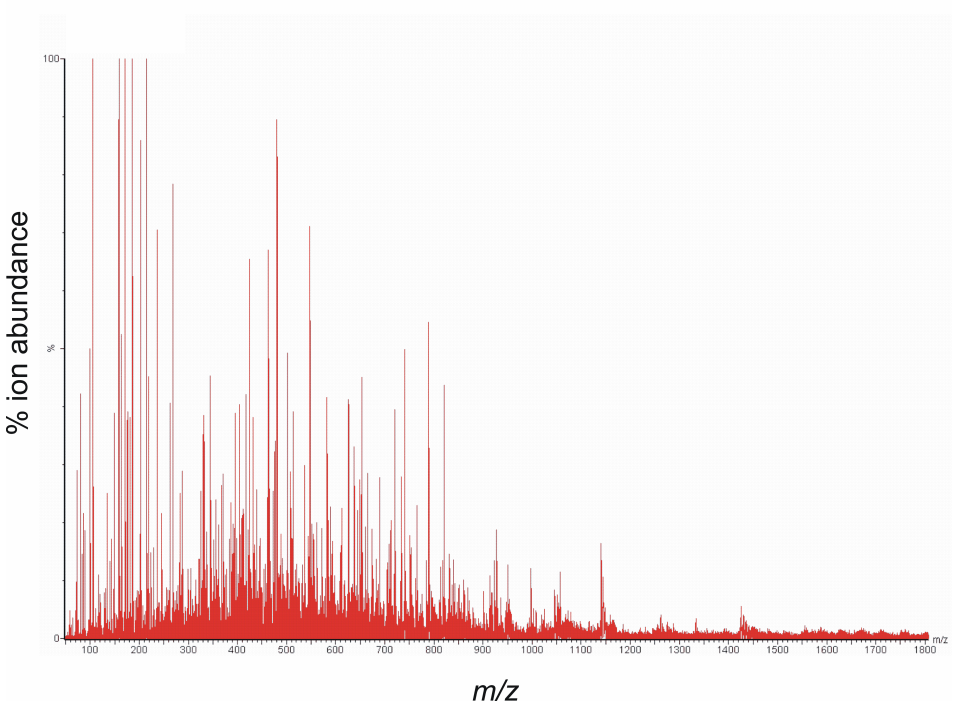


Figure 3.16. An MS spectrum from summed scans over a 2 min infusion of a human serum digest using the Nanomate system (Advion Biosciences).

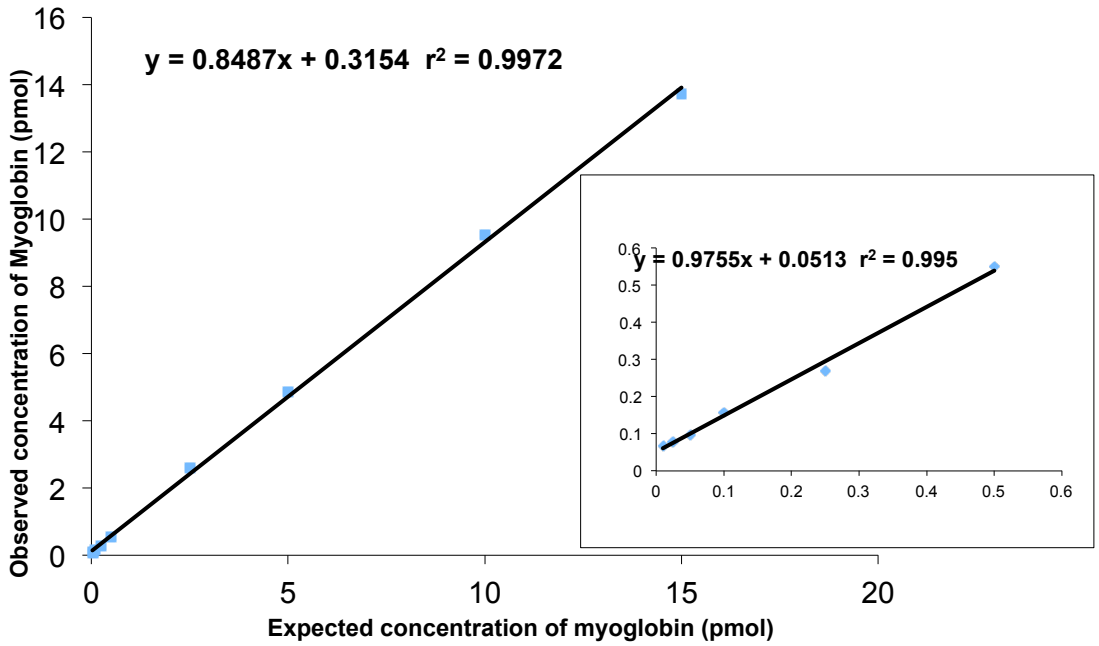


Figure 3.17. Observed vs. expected response curve for the quantification of myoglobin in a background of human serum by infusion using the Triversa Nanomate for the concentration range 10 fmol to 15 pmol. Inset = the low concentration range (10 fmol–250 fmol).

3.4 Discussion

A major aspect of proteome research is the concurrent identification and quantitation of proteins within a complex mixture. At the beginning of the studies presented in this thesis, 2-DGE and the ICAT approach, were the most widely used gel-based and LC-based proteomic techniques used for relative quantitation, respectively. However, the commercialisation of iTRAQ reagents from Applied Biosystems, allowed the Southampton proteomics laboratory to become early adopters of the iTRAQ methodology (Ross *et al*., 2004). To achieve this required overcoming a number of hurdles, since ABI instrumentation and the associated iTRAQ identification and quantitation software, were unavailable within the Southampton laboratory. As such, the first part of this Chapter describes the implementation and optimisation of methodologies for peptide fractionation and quantitation of iTRAQ labelled peptides. **Table 3.1** provides a comparison of the advantages and disadvantages of the three technologies 2-DGE, ICAT and iTRAQ.

The second part focuses on the development and application of technologies for the absolute quantification of proteins. Here, preliminary data using the AQUA approach is obtained using the model system *E. Coli*, before measuring proteins from the *C. trachomatis* proteome. The Chapter also reports a proof-of-concept application using nanoelectrospray infusion (nanoESI-infusion) MS/MS for the absolute quantification of proteins in complex mixtures. Even when using two orthogonal approaches such as SCX and RP-LC, the complexity of a proteome such as *Chlamydia*, although modest in size when compared to the human proteome, still presents a major challenge in achieving complete proteome coverage. It is therefore clear that alternative more effective methods of fractionation maybe required if complete proteome coverage of an organism such as *Chlamydia* is to be realised. There are a number of methodologies that have been successfully used for the efficient fractionation of complex mixtures. These include, COFRADIC (Gevaert et al., 2002), gas phase fractionation (*section 1.11.6*), SCX using pH gradient elution, SDS-PAGE gel electrophoresis, RP-HPLC, high-pH RP-HPLC and OFFGel. Nonetheless, studies directly comparing these technologies are limited. Dowell *et al* (2008), compared the 2-D methodologies, SCX using salt gradients, SCX using pH gradients, RP-HPLC and high-pH RP-HPLC (pH 10). When applied to the analysis of an *E.Coli* peptide lysate, they reported that RP-HPLC and high-pH RP-HPLC were more effective than SDS-PAGE and SCX using pH gradients, identifying 281 and 261 proteins, respectively. Separation using SCX with pH gradients and SDS-PAGE performed better than SCX with salt gradients, identifying 178 and 139 proteins, respectively. Separations using SCX with salt gradients were poor by comparison, with only 81 proteins identified. Further, High-pH RP has also been shown to be more effective at identifying peptides and proteins in membrane preparations, than both SCX with pH gradients and the isoelectric focusing technique OFFGel (Manadas *et al*., 2009). Nevertheless, OFFGel has been effectively used in many studies, including the separation of iTRAQ labelled peptides (de Godoy *et al*., 2008; Chenau *et al*., 2008). One particular study used OFFGel in combination with a label-free quantitative MS approach to measure the absolute cellular concentrations of 51% of the proteins encoded by the human pathogen, *Leptospira interrogans* (Malmstrom *et al*., 2009). Studies can also employ multiple orthogonal stages of separation. Garbis *et al*. (2011) used a combination of high-pressure size exclusion chromatography (SEC), Zwitterion –Ion Hydrophillic

Interaction Chromatography (HILIC), and RP-nanoLC to assign 1955 proteins spanning a dynamic range of 12 orders of magnitude in blood serum.

Table 3.1. Pros and Cons of 2-DGE, ICAT and iTRAQ

	PROS	***CONS***
2-DGE	Allows easy visualisation of post-translational modifications	Co-migration of proteins can cause errors in quantitation and protein assignment
	Allows Peptide Mass Fingerprinting to be used for identification purposes	Poor representation of proteins with extreme physio-chemical properties, e.g. hydrophobic, very basic or acidic and very large or small
	Molecular mass and pI information retained	Sampling is biased towards highly abundant proteins
		Difficult to automate
ICAT	An LC-based method offering flexibility of choosing a wide range of stationary and mobile phases available to resolve complex biological samples at the protein and peptide level	Reliance on the presence of cysteine residues reduces protein coverage, potentially misses certain proteins and post-translational modifications
	Reduces complexity of sample by affinity selecting peptides only containing cysteine residues	Labelling at the protein level, potentially reduces the efficiency of labelling as a result of steric hindrance.
	Extra information provided for database searching, that is, the knowledge that each peptide must contain a cysteine	Poor recovery of biotin tagged peptides reduces sensitivity
		Increased complexity of MS spectra
		Interpretation of fragmentation spectra can be complicated by the attached biotin label
iTRAQ	Tagging of potentially all peptides increases protein coverage improving statistical confidence in identification and quantitation	Increases complexity of sample requiring further fractionation procedures
	An LC-based method offering flexibility of choosing a wide range of stationary and mobile phases available to resolve complex biological samples at the protein and peptide level	Labelling at the peptide level can cause potential sources of error in earlier sample handling or variable degrees of trypsin digestion.
	Multiplexing of up to 8 samples and good technical reproducibility	Problems identifying the same peptides in replicate experiments
	Allows the detection of post-translational modifications	Poor precursor ion selection can cause errors in quantitation.

3.4.1. Extraction of quantitative reporter ion information

The instrumentation and software for high-throughput proteomic technologies have been improving and developing at a rapid rate. The large-scale proteomic datasets generated from these high-throughput technologies, require effective methodologies to process extract and report, both qualitative and quantitative information. However, it can be difficult to exploit such innovative methodologies, when their access is restricted to a single commercial supplier and often requires considerable financial resource to implement. At the time of these studies, although iTRAQ reagents were becoming freely available, software for performing quantitative analysis of iTRAQ data was limited to the commercial software ProQuant from Applied Biosystems.

In the absence of the ProQuant software, a Java parser was written to extract quantitative reporter ion information from iTRAQ peptide MS/MS spectra into a format, where the reporter ion ratios could easily be calculated and integrated with protein identifications generated from PLGS or Mascot. Although quite basic in function, the iTRAQ parser provided a useful tool to extract this otherwise inaccessible information. The results demonstrated that the parser was able to successfully extract reporter ion intensities from centroided peak lists in the .pkl format. Relative quantitation data obtained using iTRAQ labelled peptide extracts from *Salmonella,* showed good correlation with the expected 2:1 labelling ratio. The reporter ion intensities were extracted into a .CSV file format, which provided a familiar format for post-processing of the data to apply a reporter ion count threshold, isotope purity correction and the calculation of reporter ion ratios. Intensity information was subsequently related to their corresponding peptides using the precursor *m/z* of the peptide. **Figure 3.18** shows a data workflow of the iTRAQ parser.

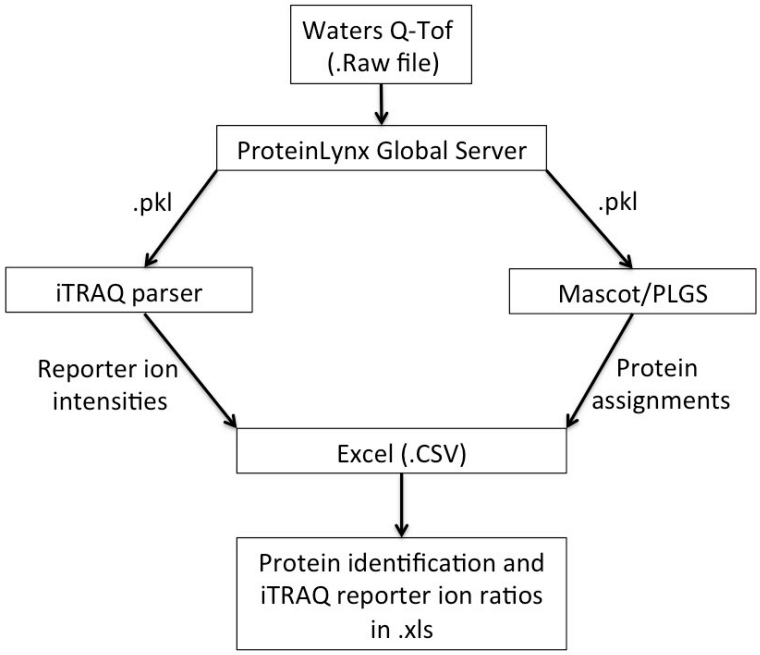


Figure 3.18. An overview of the data processing workflow for iTRAQ data acquired using a Waters Q-Tof.

The iTRAQ parser provides a basic tool for the extraction of reporter ion information. However, a limitation of this software comes from the .pkl peak list format. The only available discriminator in the .pkl file to associate reporter ion intensities with their corresponding peptide assignment in the Mascot or PLGS output file, is the *m/z* of the precursor peptide ion, calculated to 4 decimal places. Therefore, if different proteins contain a homologous peptide precursor ion with an identical *m/z* to within 4 decimal places, these must be removed, since there is no further discriminator such as spectrum number, retention time, to ensure the correct set of reporter ion intensities are assigned to the correct peptide. However, for precise quantitation, typically only peptides that are unique to a protein are used for calculating the reporter ion ratio and so by default, this weakness is beneficial. Although practically useful, it was clear that the iTRAQ parser would require further development. Such improvements needed to include using alternative file formats, making use of discriminators such as a spectrum title, or query number for each MS/MS spectrum, so that reporter ion intensities can be assigned to their corresponding peptides. For example, using generic formats such as *mz*XML, *mz*Data, or pepXML would allow data to be processed without platform specific constraints. Other improvements, such as automated isotopic purity correction, low ion count thresholding and reporter ion calculation would all be beneficial.

During the development of the iTRAQ parser, Shadforth *et al*. (2005) reported an open source software tool for the extraction of iTRAQ reporter ion ratios using the Mascot .mgf and Sequest .dta formats. In addition to improving the integration of peptide assignments with their corresponding reporter ion intensities, the software also performs automated isotope purity correction, applies a user defined low ion

count threshold filter based upon quantisation errors, and calculates a reporter ion ratio for each protein. Shortly after, Mascot released an updated version of their software, providing simultaneous identification of proteins along with their calculated reporter ion ratios within a single integrated platform. This platform offered in addition to the above features, flexibility for different types of peptide and protein normalisation, peptide outlier removal and protein ratio calculations. The software also processed iTRAQ data in a range of file formats (including the .pkl format) from the most common MS vendors. Since these software developments were moving so rapidly, continued development of the iTRAQ parser was halted and the Mascot platform was adopted for iTRAQ analysis of *C. trachomatis* L2 presented in Chapter 5. There are now a number of alternative freely distributed packages for iTRAQ quantitation available and these are shown in **Table 3.2.**

Table 3.2. Freely available iTRAQ quantitation software

Software	Format required	Reference
iTracker	Peak list (*.dta, *.mgf)	Shadforth *et al*., 2005
jTraqX	MascotDatFile library	Muth *et al*., 2010
Census	mzXML, pepXML, DTAselect	Park *et al*., 2008
Libra	mzXML, pepXML	http://tools.proteomecentre.org/Libra.php

3.4.2 Absolute quantitation

The advent of systems biology and biomarker proteomics has created a pressing need for approaches that provide absolute measures of protein quantity within cells and biofluids, offering many advantages over relative quantitation techniques and representing the gold standard of proteomic analysis (**Table 3.3**). Based upon the well-established technique of stable isotope dilution, the Gygi laboratory demonstrated absolute quantitative measurements of horse heart myoglobin and both the phosphorylated and un-phosphorylated isoforms of human separase using the AQUA strategy (Gerber *et al*., 2003). Typically, these AQUA experiments are performed using LC-MS/MS using selected-reaction monitoring (SRM) or multiple-reaction monitoring (MRM) on a triple quadrupole mass spectrometer. Since a triple quadrupole instrument was unavailable in our laboratory, the AQUA strategy was developed and implemented using a Q-ToF mass spectrometer (Waters) in combination with a CapLC nanoLC system (Waters). An instrument traditionally used for untargeted discovery based analyses.

To implement and validate the AQUA strategy, an AQUA experiment was performed measuring the expression levels of the regulatory protein BipA during different stages of growth. Using an isotopically labelled internal peptide standard corresponding to a tryptic peptide of BipA, changes in the abundance of BipA following nutrient upshift of *E.Coli* cells were monitored using AQUA. The expression of BipA was

maximal in cells in the early exponential phase of growth (>18,000 molecules per cell), but decreased over 25-fold (<700 molecules per cell) as *E.Coli* cells progressed into stationary phase. Independent immunoblot analysis of the BipA protein showed a good correlation with the AQUA data, showing a similar decrease in BipA expression from 15 min to 240 min. These data confirm that the AQUA methodology can be used to provide precise quantitative measurements of proteins within complex mixtures and can even provide greater sensitivity than traditional antibody based assays, measuring to <700 molecules per cell. The study also highlights the usefulness of the AQUA approach to measuring dynamic changes of protein expression over multiple time points and without the requirement for antibodies. These types of time-course experiments are likely to provide important clues about the regulation and control of cellular systems, such as the developmental cycle of *Chlamydia*.

Using this established AQUA strategy, the endogenous levels of two *C. trachomatis* L2 proteins, the 42.4 kDa, Major Outer Membrane Protein (MOMP) and a 69.2 kDa, metalloprotease (CTL0328) were determined (*section 2.4.2*). Both these proteins were detected in RBs and EBs in the qualitative studies presented in Chapter 4 and later in the quantitative studies presented in Chapter 5. The determined concentration for the metalloprotease was 693 fmol per 30µg (23 fmol/µg) of protein lysate in RBs and 210 fmol/30µg (7 fmol/µg) in EBs, representing a 3.3 –fold down-regulation in EBs, a decrease in expression that is consistent with the data for this metalloprotease presented in Chapter 5. The measurement of the MOMP was unsuccessful in both EB and RB samples and was attributed to poor chromatography. An explanation for this was the overloading of the RP column, caused by high concentrations of MOMP. This could be resolved by reducing the amount of sample loaded on column, or dilution of the sample prior to loading. However, since each AQUA analysis consumes 30 µg of protein lysate from both EB and RB samples, and in light of the limited amount of sample available, the remaining sample was retained for the iTRAQ studies presented in Chapter 5. Unlike qualitative or relative quantification proteomic studies, obtaining consistent quantitative measurements using the AQUA strategy is dependent upon the efficient digestion and recovery of target protein peptides from the gel matrix. However, the yield of individual peptides from such in-gel digests have been shown to be protein dependent and vary greatly, resulting in error margins of up to 50%, often requiring the use of several reference peptides, if accurate quantitation is to be achieved (Havlis and Shevchenko, 2004). Additionally, and as highlighted by the AQUA studies presented in this Chapter, the approach as it stands, where one protein is measured at a time, does not lend itself to large-scale quantitative proteomic studies. As such, AQUA methodologies using in-solution digestion followed by LC-MS, without any pre-fractionation, are now more commonplace (Wolf-Yadlin *et al*., 2007; Kitteringham *et al*., 2009; Unwin *et al*., 2009).

Table 3.3. Pros and Cons of relative and absolute quantitation

	PROS	CONS
Relative Quantitationon	High-throughput - providing a relative measure of multiple proteins per experiment.	Only provides a relative measure of quantity between individual proteins, restricting analysis to comparing 2 to 3 samples (Triplex) or in the case of iTRAQ and TMT, the comparison between 6 to 8 samples.
	Relative quantitation based on multiplexed stable-isotope labelling strategies provides an internal standard, improving reproducibility between experiments.	Protein amounts are not easily compared across experiments.
Absolute Quantitation	Allows the comparison of protein concentrations across experiments.	AQUA/MRM studies require development time for each protein being measured.
		AQUA/MRM based experiments are relatively low throughput. However, new high-throughput technologies are emerging (see *Chapter 5*).

Since the amount of time to perform an AQUA experiment can be significant, multiplexing strategies measuring 10's to 100's of proteins per analysis are starting to emerge. For example, Kuzyk *et al*. (2009) multiplexed 45 peptide reference standards to measure the endogenous levels of 45 plasma proteins using a single LC-MRM assay. Using these multiplexing approaches opens up possibilities for measuring sets of proteins precisely, under different conditions. For example, measuring the expression levels of all the glycolytic enzymes during the developmental cycle of *Chlamydia*. Ideally, each protein to be measured should be quantified using 2-3 reference peptides, since there may be different forms of the protein, i.e., degradation products, isoforms. However, the high cost of synthesising, purifying and quantifying AQUA peptides has in many studies been restricted to using a single protein per protein target (Cheng *et al*., 2006; Bondar *et al*., 2007). One cost effective and novel methodology for generating large numbers of reference peptides is Q-conCAT (Pratt *et al*., 2006; Rivers *et al*., 2007). Here, up to 100 different proteotypic peptides from different proteins are concatenated to form a synthetic gene construct. The artificial proteins are expressed in *Escherichia coli* and cultured in medium containing stable isotope precursors. The resulting synthetic protein is then purified and the concentration determined. The QconCAT proteins are added to complex mixtures in know amounts, proteolytically digested to generate reference peptides, which are used to quantify proteins using LC-MRM. Since the QconCATs are introduced at the stage of digestion, systematic errors are minimised, providing precise quantitative measurement. By reducing the limitation of generating stable-isotope reference peptides, the possibility of using the AQUA strategy more ambitious studies can be realised.

.

3.4.3 Quantitation using chip-based nanoelectrospray infusion

The length of sample run time and the usage of expensive machine time required to perform an AQUA/LC-MRM experiment is significant. A major component of this time is the chromatographic separation required prior to MS/MS. Additionally, for each new reference peptide or set of reference peptides, chromatographic optimisation must also be performed, often requiring several attempts before an optimal separation is achieved. This is further compounded by chromatographic cleaning regimes required to minimise chromatographic carry-over effects. These time and cost constraints can limit the application of AQUA for larger scale absolute quantification studies in the majority of proteomic laboratories.

Several mass spectrometric assays for the quantitation of small molecule pharmaceutical compounds have been developed using automated chip-based nanoelectrospray infusion (Nanomate, Advion BioSciences), without the use of chromatography prior to MS/MS (Leuthold *et al*., 2004; Dethy *et al*., 2003; Corkery *et al*., 2005). These types of analyses are made possible because of the high selectivity afforded by MS/MS, where qualitative and quantitative measurements can be performed without the need to baseline resolve components and often achieved without LC separation (Leuthold *et al*., 2004).

This Chapter reports a proof-of-concept application using nanoESI-infusion (Nanomate, Advion Biosciences) for the direct quantitation of proteins in solution, removing the requirement for LC separation. This has the potential to reduce run times (typically 60 min for LC-MRM), the cost per sample (columns and LC solvents) and eliminate the risk of sample carry-over.

3.4.3.1 Chip-based infusion of simple mixtures

Using ten different concentrations (from 500 amol to 15 pmol) of a myoglobin digest spiked with a myoglobin reference peptide, the sensitivity and linearity of response of the nanoESI infusion approach was evaluated. The results showed high linearity across the concentration range measured, from 15 pmol to the lower limit of quantitation (LOQ) 500 amol. Even in these simple mixtures, these data suggest that this approach could have possible application for the routine quantification of semi-purified recombinant proteins, where measures of absolute concentration are required. Although acquisition times per sample in these studies were 30 minutes, acquisition times of 2-10 min, could easily be achieved depending upon the abundance of the protein(s) being measured, providing a rapid precise assay. Alternatively, the ability to perform extended periods of infusion could also be used to improve the S/N ratio for low intensity peptides, or allow time to perform further analyses such as MS^3 (see later).

3.4.3.2 Chip-based infusion of complex mixtures

Increased sample complexity presents several issues for quantitative analysis, the first of which is selectivity. Chromatographic separation brings an added selectivity to MS analysis, reducing the possibility of crosstalk between analytes. Therefore, methodologies such as ESI-infusion, where chromatography is not employed, are dependant upon tandem mass spectrometry (MS/MS) to eliminate this crosstalk. Using MRM-based experiments and measuring multiple fragment ions of a selected precursor ion, can dramatically increase the selectivity, eliminating interferences and improving both linearity and limits of quantitation.

Another important aspect of chromatography is that it provides concentration of the analyte and minimises matrix suppression effects. In comparison to the quantification of purified myoglobin, the lower limits of quantification (LOQ) for myoglobin in a background of digested human serum was found to be 20-fold less [LOQ = 10 fmol] and was not detectable below 10 fmol. Matrix effects causing suppression of the analyte signal are the most likely explanation for this reduced limit of quantitation. This is not so surprising when considering that ~2000 proteins have been confirmed in human serum (Garbis *et al*., 2011) and when the complexity is increased further by proteolytically digestion. By increasing the selectivity of the analysis, these matrix effects can be reduced or eliminated. However, in the absence of chromatographic separation, increased selectivity is reliant on the MS methodologies employed. One possibility is to introduce a third stage of MS. This has been shown to greatly increase selectivity and thereby reducing matrix effects. Olsen and Mann (2004), used MS^3 of C- and N-terminal peptide ions generated from MS/MS (MS^2) to increase identification specificity, even at the sub-femtomole level. Similarly, MS^3 has been employed in combination with nanoESI-infusion to increase quantitative precision of a parent drug and its metabolites in human plasma, without the use of LC (Leuthold *et al*., 2004).

Flow rates in nanoESI display a strong influence on ion signals and the lower flow rates of nanoESI have been shown to be more tolerant of salts compared to conventional ESI (Wilm *et al*., 1996). In 'true nanoESI', where flow rates are maintained below 50 nL/min, in addition to a higher tolerance towards salts, less analyte suppression is observed, with claims that at flow rates of a few nL/min, the signal suppression effects totally disappear (Schmidt *et al*., 2003). Nonetheless, flow rates through the ~10 μM nozzles of the Nanomate platform used in this study are between 100-200 nL/min and ion suppression is still observed in highly complex matrices, albeit, at lower levels than convention ESI (Chen *et al*., 2011). Recently, nanoESI-infusion has been used in combination with FT-ICR-MS to identify proteolytic digests using four narrow overlapping mass ranges. Here, ion suppression effects normally observed using a full range scan were minimized, improving the number of moderate to low abundance peptides identified. Each analysis was performed in 3 min and the results were comparable to LC-MS/MS (Chen *et al*., 2011). By combining different MS strategies, such as MS^3, over-lapping mass ranges or using alternative separation principles such as ion mobility (Giles *et al*., 2011), increases in selectivity can be realised, offering the possibility of achieving absolute quantitation for even low abundance proteins in highly complex mixtures, without the requirement for LC.

During the final preparation of this thesis, Xiang and Koomen, (2012), evaluated the concept of nanoESI-infusion (Nanomate) in combination with MRMs for the absolute quantification of standard proteins and the expression of heat shock proteins in digests of whole cell lysates. They concluded that nanoESI-infusion was comparable to LC-MRM when peak intensities where sufficient or when corrected to eliminate interference. Applied to the measurement of abundant proteins or enriched fractions, nanoESI-infusion provided a fast, high-throughput method for the precise absolute quantitation of proteins. However, they also report that measurements for the quantification of low intensity peptides in complex mixtures, were different from those obtained using LC-MRM and required further investigation. Extended periods of infusion also

open up the possibility for high-throughput screening, especially when multiplexed. Xiang and Koomen, (2012), calculated that up to 6000 transitions could be measured using only 5 µl of sample.

3.4.4 Relative and Absolute quantification in *Chlamydia*

There are significant barriers to investigating Chlamydial genes and their regulation, For instance, the obligate intracellular developmental cycle means there is no host – free system to culture *Chlamydia* and genetic manipulation has historically been difficult. As a consequence, little molecular detail about the biological function of its gene products is known. The precise quantification of proteins under specific conditions would therefore answer some key questions about chlamydial biology. Such questions are likely to include: What are the most abundant proteins during the different stages of the developmental cycle? This will be vital if better diagnostic reagents are to be developed. How many hypothetical proteins are actually expressed, and at what levels, during the developmental cycle? How does the transcriptomic profile (Belland *et al*., 2003; Nicholson *et al*., 2003) relate to the proteomic profiles? After measurement of all these proteins during different stages of the developmental cycle, can they be clustered, providing clues about their biological roles in chlamydial biology?

Although relatively modest in size, the qualitative and quantitative analysis of the *C. trachomatis* proteome still presents a challenge. However, with less than 900 proteins, the proposition of relative and absolute quantitation of a complete chlamydial proteome becomes tangible, offering the prospect of deepening our understanding of how this pathogen mediates the diverse range of pathological outcomes. Development and implementation of the iTRAQ approach in this Chapter, has paved the way for comparisons of protein expression between RBs and EBs using relative quantitation (*Chapter 5*). Nevertheless, the approaches for the absolute quantitation of proteins established in this Chapter, are still in the majority of proteomic laboratories relatively low throughput (10-20 proteins per day for LC-MRM based assays) and require a considerable investment of resource. Moreover, the amount of sample consumed, realistically restricts quantitative measurement to protein subsets when investigating intracellular pathogens such as *Chlamydia*, where protein yields are low. During the course of these studies, a high-throughput label-free MS approach (LC-MS^E), allowing the identification and quantification of up to thousands of proteins measured in absolute amounts was described (Silva *et al*., 2006a; Silva *et al*., 2006b). Collaborative studies with Waters (LifeSciences development laboratory, Manchester, UK) opened up the opportunity to apply this MS technology to the measurement of cellular concentrations of chlamydial proteins on a proteome-wide scale (*Chapter 5*). As such, attention was re-focused on using this approach for the comparison of chlamydial proteins in RBs and EBs. Whilst, the AQUA and nanoESI-infusion approaches are not easily extended to the measurement of an entire proteome, they will be highly useful in future studies. For example, the precise measurement of pathway components early in the developmental cycle, before replication begins will require highly sensitive assays owing to the limited number of EBs present at this stage. The approach will also be invaluable for validation of the data presented in Chapter 5, replacing traditional antibody-based assays.

CHAPTER 4

SHOTGUN PROTEOMIC ANALYSIS OF *C. TRACHOMATIS*

4.0 Shotgun proteomic analysis of *C. trachomatis**

4.1 Introduction

The availability of the complete genome sequence of *Chlamydia trachomatis* serovar D and subsequently *C. trachomatis*, serovar L2, have facilitated studies focused on the protein components expressed by chlamydial genomes. Given the intrinsic imprecision of genome annotation software, the validation of genome annotations has become increasingly important. Proteomics provides confirmation that predicted genes encode *bona fide* proteins distinguishing between authentic and pseudo-genes. The vast majority of genome annotation efforts now rely on computational methods for gene prediction or comparison to homologous proteins in other organsims, but are currently rarely validated experimentally (Jaffe *et al*., 2004; Ansong *et al*., 2008). For example, the error rate in annotation of the 340 genes of *Mycoplasma genitalium* genome has been estimated to be ~8% (Brenner, 1999). Even for extremely well studied species such as *E. coli* strain K-12, substrain JM109, investigations a decade later revealed that many ORFs were incorrectly annotated (Maillet *et al*., 2007). Additionally, studies in the proteomic domain provide biological insights, complement quantitative experiments and help to define the limits of the proteome under study.

* This Chapter has been modified from a paper published in *Proteomics* entitled "Shotgun proteomic analysis of *Chlamydia trachomatis"* (Skipp *et al*., 2005)

4.1.1 Proteomic analysis of *Chlamydia*

2-D gel electrophoresis in combination with MS technology has been extensively used for large-scale protein separation and identification experiments and has been applied to the study of chlamydiae. Prior to the availability of genome sequences, there have been a number of chlamydial proteomic studies, which are reviewed by Vandahl *et al*. (2002) and which were discussed in *section 1.8.1* of this thesis.

Bini *et al*. (1996) using the improved resolution of Immobilised pH Gradient (IPG) strips, produced the first high-resolution 2-D gel electrophoresis map of proteins from the extracellular EB of *C. trachomatis* L2. This map showed up to 600 silver-stained spots, although the vast majority were not identified. Pulse-labelling and pulse-chase experiments have also been employed to address the protein expression, producing definitive maps of EBs for both *C. trachomatis* (Shaw *et al*., 2002a) and *C. pneumoniae* (Vandahl *et al*., 2001). However, although many of the protein spots from these maps have been identified, they only provide a maximum of 16% coverage of the predicted proteome and do not give data on protein expression in RBs. Belland *et al*. (2003) demonstrated that at the mid point in the infectious cycle of *C. trachomatis* (15-24 h PI), virtually every chromosomal and plasmid gene is transcribed. The remaining 28 genes, corresponding to ~3% of the genome were shown to be transcribed during the later stages of the developmental cycle (40 h PI). Although several studies report a disparity between the relative expression of mRNAs and the corresponding expressed proteins (Anderson and Seilhamer, 1997; Gygi *et al* 1999a), it is highly likely that a large percentage of these transcripts are expressed at the protein level and exceed the 16% coverage currently

detected. Additionally, the small size of the *C. trachomatis* genome (serovar D, 894 ORFs; serovar L2, 897) makes it an ideal candidate for efforts aimed at achieving comprehensive proteome coverage and understanding the relationship between the transcriptome and proteome.

The limited proteome coverage of chalmydiae achieved so far, can, in part, be attributed to the necessity to culture *Chlamydia* in a host cell and difficulties in obtaining good yields of highly pure preparations of EBs and RBs. Additionally, the poor representation of some classes of proteins when using 2-D gel electrophoresis (Gygi *et al*., 2000; Rehm, 2006) has also been a contributing factor. Nonetheless, the value of the 2-D gel electrophoresis approach should not be under estimated. It has proved useful in the identification of novel proteins, finding differences between preparations, investigating the effects of treatments on *Chlamydia*-infected cells and comparing *Chlamydia* at different times during the developmental cycle (*section 1.8.1*). However, the majority of the predicted chlamydial components are yet to be detected at the protein level, either in EBs or RBs. This suggests that alternative approaches are needed to provide increased coverage of the chlamydial proteome

A particular method that has generated a notable amount of interest is multidimensional liquid chromatography in combination with MS/MS for the direct analysis of complex peptide mixtures derived from an organism of interest (reviewed by Fournier *et al*., 2007). One commonly used configuration, termed MudPIT (Multidimensional Identification Technology) has proved to be very effective, combining strong cation exchange with RP nanocapillary columns arranged in series, to provide high-resolution separation and concentration of peptides prior to their identification by MS/MS. (Link *et al*.,1999; Washburn *et al*., 2001). As a peptide–centric approach, complex mixtures of proteins are first proteolytically digested to produce peptides, before loading onto a SCX column. Fractions of these peptides are subsequently eluted from the SCX column using successive steps of increasing salt concentration. Each eluted peptide fraction is either loaded off-line or directly onto a RPLC column for separation and analysis using tandem mass spectrometry. Recently, a modification of MudPIT utilizing on-line SCX in combination with long RP columns and higher column temperatures, achieved the identification of 53% and 46% of the predicted proteomes for the prokaryotes, *Corynebacterium glutamicum* and *E. coli* strain MG1655, respectively. (Fränzel and Wolters, 2011).

An alternative protein-centric approach, termed GeLC-MS/MS, uses conventional 1D SDS-PAGE for protein separation in conjunction with LC-MS/MS analysis (GeLC-MS/MS) of peptides generated by *in-gel* digestion using a site-specific endo-protease such as trypsin (Schirle *et al*., 2003). Similar to MudPIT, this technique boosts peak and load capacity thereby potentially increasing the number of peptides that can be identified. It has been applied in a wide range of applications, ranging from extensive proteome coverage of different organisms (Jones *et al*., 2011; Wang *et al*., 2010), or their sub-celluar compartments (Vaughan *et al*., 2006; Williams *et al*., 2007), to biofluids and tissues (Nicholas *et al*., 2006; Albrethsen *et al*., 2010).

In this study, the techniques of MudPIT, GeLC-MS/MS and 2-D gel electrophoresis have been used to characterise the expressed proteome of *C. trachomatis* strain L2 in both the EB and RB forms. At the same time, the study provides a comparison of the technologies used to characterise this proteome. At the initiation of these studies the genome sequence of *C. trachomatis* L2 was still being completed at the Wellcome Trust Sanger Institute. It was therefore necessary to initially search data against the completed genome sequence of *C. trachomatis* serovar D (Stephens *et al*., 1998) for protein identification and classification. The data was subsequently re-searched against the *C. trachomatis* L2 genome sequence at a later date (Thomson *et al*., 2008).

4.2 Results

4.2.1 Purification of EBs and RBs

EBs contain several proteins in the outer membrane that are synthesised during the late stages of the developmental cycle that are not observed in RBs (Hatch *et al*., 1984; Hatch *et al*., 1986; Newhall *et al*., 1987; Moroni *et al*., 1996). Additionally, the EB membrane is cross-linked by disulphide bonds and the association of the major outer membrane protein (MOMP) and LPS in the membrane is different in EBs and RBs (Hackstadt *et al*., 1985; Birkelund *et al*., 1988). These characteristics, in combination with the increased transcriptional acitivity observed during the mid cycle when RBs are undergoing replication, suggests that there will be major differences in the proteomes of RBs and EBs and underscores the importance of defining the former. While the methods for purification of infectious EBs are well established, RBs are osmotically fragile and must be purified from inclusions, presenting difficulties and requiring the development of a purification strategy.

C. trachomatis L2 grows well in a range of mammalian cells with a relatively short developmental cycle and gives high yields compared to isolates from the trachoma biovar. Additionally, infection of host cells is highly efficient, occurring without the need for centrifugation of *Chlamydia* onto the host cell. These features made this strain particularly suitable for this study. In collaboration with Professor Ian Clarke, Dr Joanne Spencer and Mrs Leslie Cutliffe (Southampton General Hospital, UK), a purification strategy for EBs and RBs was developed.

C. trachomatis L2 were cultured in BGMK cells according to the methods described in Chapter 2. Purification of RBs from infected cells was achieved at 15 h PI after replication of RBs by binary fission, but before the second stage differentiation of RB to EB within the inclusions. RBs were purified by two cycles of density gradient centrifugation in discontinuous Urografin gradients. Yields were typically low, with 1 – 2 mg of RB cells being obtained for 14 x 175 cm^2 flasks where >95% of host cells were infected. EBs were similarly purified and assessment of the purity of both chlamydial forms was achieved using thin section EM. Micrographs demonstrated the absence of RB in EB preparations and vice versa (**Figure 4.1**). Additionally, individual preparations of RBs were verified for translational activity using [^{35}S] methionine in host free protein synthesis reactions as described in *section 2.2.8*. RBs purified at 15 h were translationally active; by contrast, purified EBs were metabolically inert.

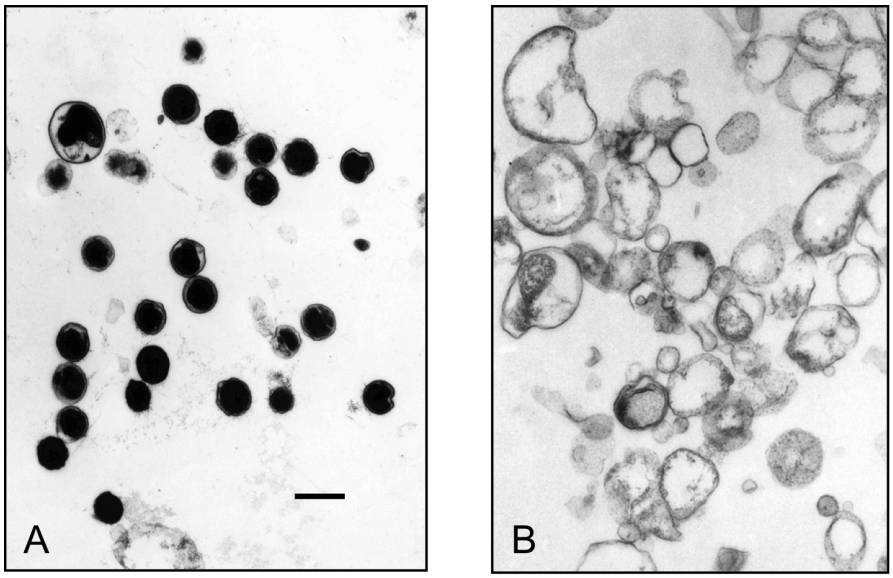


Figure 4.1. Thin section EM of gradient-purified EBs (A) and RBs (B) used for subsequent analyses. RBs were purified from infected cells at 15 h post-infection and EBs at 48 h post-infection The scale bar represents 0.5 μM. Micrographs were produced by the Southampton Biomedical Imaging Unit (Southampton General Hospital, UK).

4.2.2 Two-dimensional gel analyses of EB and RB whole cell lysates from *C. trachomatis*

Having isolated pure preparations of EBs and RBs, proteomic analysis of these preparations was performed using the different techniques outlined above. Accordingly, the first technique used was two-dimensional gel electrophoresis (2-DGE). Protein profiles of EBs and RBs were obtained by solubilisation of protein extracts in 2D-GE sample lysis buffer and fractionated by 2-DGE as described in Chapter 2. **Figure 4.2** illustrates the protein profile obtained with whole cell lysates of EBs from *C. trachomatis* L2. Three independent biological preparations were analysed of both EBs and RBs. Despite using a 2-DGE clean-up procedure, gels often showed extensive streaking, indicating the presence of interfering chemicals, such as salts, lipids and nucleic acids. Reproducibility of RB 2-D gels was extremely poor and alignment with other RB 2-D gels and EB 2-D gels for comparison was not possible except for a limited number of highly abundant spots. The separation of EB and RB samples using 2-DGE was performed in collaboration with Dr. Joanne Spencer (Southampton General Hospital, UK).

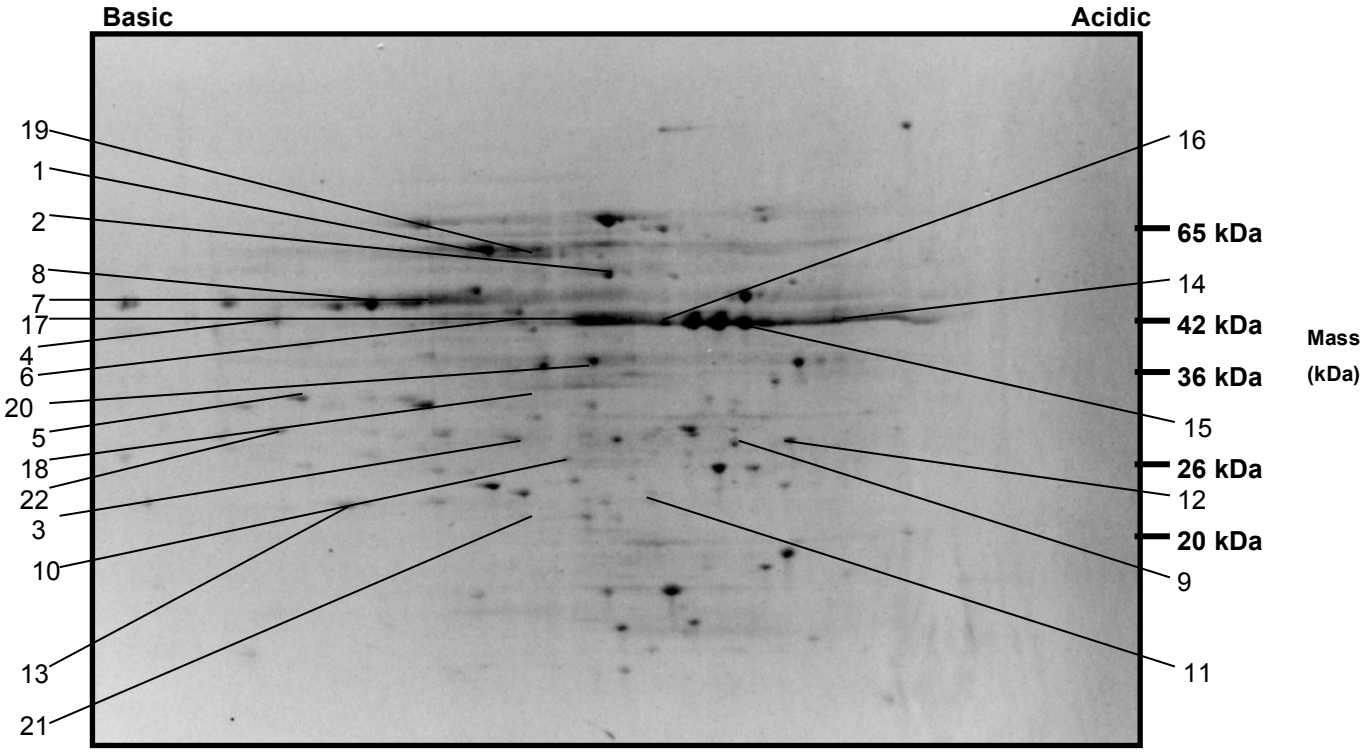


Figure 4.2. 2-D gel image of protein extracts separated from purified EBs of *C. trachomatis* L2. The first dimension range was pH 4.0 to 7.0. The gel was stained with SYPRO Ruby and the image analysed as described in Chapter 2. Identified proteins are indicated by the relevant spot numbers and listed in **Table 4.1** and **Table 4.2**.

Spots that were reproducibly present in at least three biological replicate gels for each type of sample were excised and digested *in situ* with trypsin. Two approaches were used for the identification of the resulting extracted peptides. Initially, all samples were screened using MALDI-TOF MS to determine the presence of peptides. In samples where peptides were observed, nanoLC-MS/MS was used to identify the peptides and hence assign the proteins. **Table 4.1** provides a list of those proteins identified in EBs and RBs using 2-DGE. All proteins identified in RBs were also identified in EBs. Of the 19 proteins identified, the majority corresponded to house-keeping components and other previously reported high abundance proteins (Shaw *et al*., 2002a). Only two hitherto hypothetical proteins were found and other classes of protein (e.g., integral membrane proteins) were poorly represented (**Table 4.1**).

Table 4.1. A list of identified protein spots from 2D gels of EBs from *C. trachomatis* L2

Gel spot number[a]	UniProt accession	Protein Name	Predicted[b] MW/pI
1*	B0B7W6	HSP 70	70/5.0
2[$]	B0B815	60 kDa Cysteine Rich OMP	59/7.4
2[$]	B0B9L8	HSP 60	58/5.3
3	B0B7V0	Arginine Binding Protein	28/5.8
4	B0B8L2	Aromatic AA Aminotransferase	44/5.7
5	B0B8Q5	Elongation Factor TS	30/5.4
6*	B0B7N8	Elongation Factor Tu	43/5.5
7*	B0B7N8	Elongation Factor Tu	43/5.5
8*	B0B7N8	Elongation Factor Tu	43/5.5
9*	B0B8B5	FKBP type peptidyl prolyl cis trans isomerase	26/5.0
10	B0B9T6	Glucose 6 P Dehydrogenase DevB family	28/5.4
11*	B0B8F8	hypothetical protein	21/5.9
12	B0B8B2	hypothetical protein	27/5.2
13	B0B8F8	hypothetical protein	21/5.9
14*	B0B8Q7	Major Outer Membrane Protein	42/5.1
15	B0B8Q7	Major Outer Membrane Protein	42/5.1
16	B0B8Q7	Major Outer Membrane Protein	42/5.1
17*	B0B881	RNA Polymerase Alpha	41/5.4
18*	B0B7N0	RNA Polymerase Beta	140/5.7
19	B0B9K6	S1 Ribosomal Protein	63/5.22
20*	B0B7M9	Transaldolase	36/4.98
21	B0B7N6	Transcriptional termination factor	20/5.31
22	B0B7P5	Triosephosphate Isomerase	29/5.62
23*	B0B8D3	Yop proteins translocation lipoprotein J	35/5.66

[a] Proteins also found in 2-D gels of RBs are indicated by *

[b] pI and molecular mass (kDa) were calculated using ProteinLynx Global Server 2.05 (Waters)

[C] In the case of exported proteins and integral membrane proteins, the coordinates include the signal sequence region where this is known.

[$] The proteins HSP 60 and the 60 kDa cysteine rich OMP were identified within the same gel spot.

4.2.3 Analysis of the chlamydial proteome by MudPIT

In light of the results obtained from the 2-D gel electrophoresis experiments, MudPIT was employed to increase proteome coverage. ~100 μg of both EB and RB samples were digested in solution with trypsin containing Rapigest and the resulting peptides were trapped onto an Optipak SCX column before being eluted using seven steps of increasing salt concentration (*section 2.3.8*). At each step, the peptide fraction was

eluted onto a C18 reverse phase nanoLC column, separated and analysed using tandem mass spectrometry. MS/MS data were searched against a protein translation of the *C. trachomatis* and human genome. The human genome was selected since this offered the closest available genome sequence with significant homology to the African green monkey (*Cercopithecus aethiops*) host cells used for culturing *Chlamydia* (Almeida *et al*., 2011).

MudPIT analysis of EB samples generated a total of 9048 MS/MS spectra, resulting in the identification of 222 unique peptides and 107 proteins. Despite extensive purification of EBs and RBs, host cell protein contamination was still observed. Although a similar number of MS/MS spectra were obtained (8762 spectra), the total number of chlamydial proteins assigned in RBs using MudPIT was 70. This reduced sampling of the predicted RB proteome is likely to reflect greater levels of host cell contamination in RB preparations. Despite the contamination of preparations with host cell proteins, no homolgous peptides between *Chlamydia* and the human genome were identified.

Protein identification using the MudPIT approach, in contrast to the poor proteome coverage obtained using 2-DGE, was relatively efficient and unbiased, identifying proteins across a range of functional categories and encompassing proteins with isoelectric pH values ranging from 4.5 to 11.4; and molecular weights from 8.4 to 182.8 kDa. Proteins identified via the MudPIT approach are listed in **Table 4.2**.

Table 4.2. Proteins identified from *C. trachomatis* L2 by 2-DGE, MudPIT and GeLC-MS/MS.

Category/Protein name	Gene name	UniProt accession	Primary *locus*	Serovar D *locus*	MW (kDa) [a]	pI [a]	No. of peptides	Tech [b]	EB/RB
Amino acid Biosynthesis									
Leucine Dehydrogenase	*ldh*	B0B8Z9	CTL0142	CT773	37	5.3	2	G	EB
Serine Hydroxymethyltransferase	*glyA*	B0B804	CTL0691	CT432	54	6.6	1	G	EB
Tryptophan Synthase Alpha Chain	*trpA*	B0B9S2	CTL0424	CT171	28	4.8	1	G	EB
Biosynthesis of Cofactors									
Dihydropteroate Synthase	*folP*	B0B8I8	CTL0877	CT613	50	5.6	1	M	EB RB
Geranylgeranyl pyrophosphate synthase	*ispA*	B0B8K3	CTL0892	CT628	33	4.9	3	G,M	EB
Ribityllumazine Synthase	*ribH*	B0B8V8	CTL0101	CT732	16	6.3	2	G	EB
Thioredoxin Reductase	*trxB*	B0B9K7	CTL0354	CT099	38	8.0	1	G	EB
Cell Envelope									
Putative Outer Membrane Protein A	*pmpA*	B0B7Y2	CTL0669	CT412	106	8.6	2	G	RB
Putative outer membrane protein B	*pmpB*	B0B7Y3	CTL0670	CT413	183	5.7	11	G,M	EB
Putative outer membrane protein C	*pmpC*	B0B7Y4	CTL0671	CT414	187	4.5	6	G	EB RB
Putative Outer Membrane Protein D	*pmpD*	B0B940	CTL0183	CT812	161	4.8	29	G,M	EB RB
Putative Outer Membrane Protein E	*pmpE*	B0B9A1	CTL0248	CT869	105	7.2	3	G	EB
Putative Outer Membrane Protein F	*pmpF*	B0B9A2	CTL0249	CT870	112	9.0	5	G	EB
Putative Outer Membrane Protein G	*pmpG*	B0B9A3	CTL0250	CT871	107	5.5	7	G,M	EB RB
Putative Outer Membrane Protein H	*pmpH*	B0B9A4	CTL0251	CT872	108	6.4	11	G	EB
Putative Outer Membrane Protein I	*pmpI*	B0B9A6	CTL0254	CT874	96	6.5	6	G,M	EB RB
60kDa Cysteine Rich OMP	*omcB*	B0B815	CTL0702	CT443	59	7.4	14	G,M,2D(2)	EB RB
60kDa Inner Membrane Protein	*oxaA*	B0B7G7	CTL0503	CT251	88	9.1	6	G,M	EB RB
OmpH Like Outer Membrane Protein	*ompH*	B0B7F8	CTL0494	CT242	19	4.8	2	G	EB
Major Outer Membrane Protein	*ompA*	B0B8Q7	CTL0050	CT681	42	5.1	20	G,M,2D(14, 15,16)	EB RB
Outer Membrane Protein analog	*ompB*	B0B8T9	CTL0082	CT713	37	5.2	3	G	EB
Peptidoglycan-associated lipoprotein	*pal*	B0B8H4	CTL0863	CT600	22	9.0	3	G	EB
Omp85 Analog		B0B7F7	CTL0493	CT241	89	9.3	3	G,M	EB RB
UDP N acetylmuramoylalanylglutamyl DAP Ligase	*murE*	B0B715	CTL0521	CT269	53	5.6	1	M	RB
UDP 3 O 3hydroxymyristoyl glucosamine N acyltransferase	*lpxD*	B0B7F9	CTL0495	CT243	38	7.6	1	G	EB
KDO Synthetase	*kdsA*	B0B8N1	CTL0024	CT655	30	6.0	2	G,M	EB
Central Intermediary Metabolism									
Glycogen Phosphorylase	*glgP*	B0B7G4	CTL0500	CT248	93	5.9	5	G	EB
Glycogen Synthase	*glgA*	B0B925	CTL0167	CT798	53	5.7	3	G	EB
Inorganic Pyrophosphatase	*ppa*	B0B8Z8	CTL0141	CT772	23	4.8	1	G	EB
Cellular Processes									
chromosome partitioning ATPase CHLTR plasmid protein homolog GP5D	*minD*	B0B8F6	CTL0845	CT582	28	7.1	3	G,M	EB RB
ATP dependent zinc protease	*ftsH*	B0B970	CTL0213	CT841	102	6.1	10	G,M	EB RB
HSP 60	*groEL*	B0B9L8	CTL0365	CT110	58	5.3	15	G,M,2D(2)	EB RB
HSP 70	*dnaK*	B0B7W6	CTL0652	CT396	71	5.1	15	G,M,2D(1)	EB RB
10KDa Chaperonin	*groES*	B0B9L9	CTL0366	CT111	11	5.0	2	G	EB
possible Disulfide Bond Chaperone	*dsbG*	B0B9S7	CTL0429	CT177	27	8.1	1	G	EB
Heat Shock Protein J	*dnaJ*	B0B7R0	CTL0595	CT341	42	7.9	3	G,M	EB RB
HSP 70 Cofactor	*grpE*	B0B7W5	CTL0651	CT395	22	4.6	1	G	EB
Leader 60 peptide periplasmic		B0B914	CTL0156	CT788	19	4.8	1	G	EB

Thio specific Antioxidant TSA Peroxidase	*ahpC*	B0B8H7	CTL0866	CT603	22	4.8	2	G,M	EB RB
Superoxide Dismutase Mn	*sodM*	B0B7L0	CTL0546	CT294	23	6.2	5	G,M	EB RB
Secretion Chaperone	*scc1*	B0B9J6	CTL0343	CT088	16	6.7	1	G	EB
Translocase	*secY*	B0B884	CTL0772	CT510	50	10.4	3	G	EB RB
Trigger Factor peptidyl prolyl isomerase	*tig*	B0B8T3	CTL0076	CT707	50	5.0	6	G,M	EB RB
Signal Peptidase I	*lepB*	B0B9C7	CTL0275	CT020	72	8.6	1	G	EB RB
GTP Binding Protein	*ychF*	B0B9K0	CTL0347	CT092	40	5.2	1	M	EB
Yop proteins translocation lipoprotein J	*sctJ*	B0B8D3	CTL0822	CT559	35	5.7	8	G,M,2D(22)	EB RB
Yop proteins translocation protein L	*sctL*	B0B8D5	CTL0824	CT561	25	5.9	4	G,M	EB RB
Yops secretion ATPase	*sctN*	B0B8P5	CTL0038	CT669	48	5.8	3	G	EB
probable Yop proteins translocation protein C general secretion pathway protein	*sctC*	B0B8Q0	CTL0043	CT674	101	5.7	6	G,M	EB RB
Low Calcium Response D	*lcrD*	B0B9J8	CTL0345	CT090	78	8.3	10	G,M	EB
Low Calcium Response H	*scc2*	B0B8F0	CTL0839	CT576	26	9.2	2	G	EB
Yop proteins translocation protein R	*sctR*	B0B8D6	CTL0825	CT562	34	8.4	1	G	EB
Protein Export	*secF*	B0B821	CTL0708	CT448	156	7.7	5	G	RB
Flagellar Motor Switch Domain YscQ family	*sctQ*	B0B8P8	CTL0041	CT672	41	4.7	4	G	EB
Energy Metabolism									
ATP Synthase Subunit I	*atpI*	B0B7M1	CTL0557	CT305	73	6.5	1	G	EB
ATP Synthase Subunit A	*atpA*	B0B7M4	CTL0560	CT308	65	5.1	1	M	EB
ATP Synthase Subunit E	*atpE*	B0B7M6	CTL0562	CT310	23	5.4	1	G	EB
ATP Synthase Subunit B	*atpB*	B0B7M3	CTL0559	CT307	49	5.9	1	G	EB
ATP Synthase Subunit D	*atpD*	B0B7M2	CTL0558	CT306	23	9.3	1	G	EB
NADH Ubiquinone Dehydrogenase	*nqrB*	B0B7J4	CTL0530	CT278	55	8.7	1	G	EB
NADH Ubiquinone Oxidoreductase Gamma	*nqrC*	B0B7J5	CTL0531	CT279	34	6.4	2	G,M	EB
NADH ubiquinone oxidoreductase alpha chain	*nqrA*	B0B8K9	CTL0002	CT634	52	9.1	2	G	EB
Phenolhydrolase NAD ubiquinone oxidoreductase	*dmpP*	B0B8W6	CTL0109	CT740	48	5.3	2	G	EB
ADP ATP Translocase	*npt1*	B0B9H3	CTL0321	CT065	58	8.7	5	G,M	EB RB
ADP ATP Translocase	*npt2*	B0B868	CTL0756	CT495	60	9.4	1	G	RB
Phosphoenolpyruvate Carboxykinase	*pckA*	B0B8T6	CTL0079	CT710	66	5.7	2	G,M	EB
Phosphoglycerate Kinase	*pgk*	B0B8R9	CTL0062	CT693	43	5.9	2	G	EB
Phosphoglycerate Mutase	*pgmA*	B0B8U8	CTL0091	CT722	26	7.2	3	G	EB
Enolase	*eno*	B0B8G1	CTL0850	CT587	45	4.6	5	G,M	EB RB
Fructose 6 P Phosphotransferase	*pfkA_2*	B0B9V7	CTL0459	CT207	61	6.7	1	G	EB
Fructose 6 P Phosphotransferase	*pfkA*	B0B9V5	CTL0457	CT205	62	6.5	1	G	EB
Glyceraldehyde 3 P Dehydrogenase	*gapA*	B0B879	CTL0767	CT505	36	5.8	2	G,M	EB RB
Pyruvate Kinase	*pykF*	B0B7Q0	CTL0586	CT332	54	6.3	1	G	EB
Triosephosphate Isomerase	*tpiS*	B0B7P5	CTL0582	CT328	30	5.6	2	G,2D(22)	EB
Malate Dehydrogenase	*mdhC*	B0B7U5	CTL0630	CT376	36	6.7	6	G	EB
Phosphomannomutase	*mrsA*	B0B7L1	CTL0547	CT295	67	5.1	3	G,M	EB
Predicted 1 6 Fructose biphosphate aldolase dehydrin family	*dhnA*	B0B9W5	CTL0467	CT215	38	/6.71	3	G	EB
Cytochrome Oxidase Subunit I	*cydA*	B0B9C0	CTL0268	CT013	50	9.5	5	G	EB
Cytochrome Oxidase Subunit II	*cydB*	B0B9C1	CTL0269	CT014	40	8.9	1	G	EB
Glucose 6 P Dehydrogenase	*zwf*	B0B9T5	CTL0437	CT185	51	5.4	2	G	EB
Glucose 6 P Dehydrogenase DevB family	*devB*	B0B9T6	CTL0438	CT186	29	5.4	2	G,2D(10)	EB
Glucose 6 P Isomerase	*pgi*	B0B7U7	CTL0633	CT378	58	5.8	5	G	EB
Oxoglutarate Dehydrogenase	*sucA*	B0B9G2	CTL0310	CT054	103	5.4	1	G	EB
Transaldolase	*tal*	B0B7M9	CTL0565	CT313	36	5.0	8	G,M,2D(20)	EB RB
Transketolase	*tktB*	B0B8X6	CTL0119	CT750	73	5.4	1	G	EB
6 Phosphogluconate Dehydrogenase	*gnd*	B0B9H1	CTL0319	CT063	53	5.4	4	G	EB
Ribose 5 P Isomerase A	*rpiA*	B0B9W3	CTL0465	CT213	26	5.5	2	G	EB

Pyruvate Dehydrogenase	*pdhB*	B0B7G2	CTL0498	CT246	36	5.8	1	G	EB
Dihydrolipoamide Acetyltransferase	*pdhC*	B0B7G3	CTL0499	CT247	46	5.9	2	G	EB
Dihydrolipoamide Succinyltransferase	*sucB*	B0B9G3	CTL0311	CT055	40	5.1	2	G	EB
Succinyl CoA Synthetase Alpha	*sucD*	B0B951	CTL0194	CT822	30	5.4	1	G	EB
Succinyl CoA Synthetase Beta	*sucC*	B0B950	CTL0193	CT821	42	5.6	5	G	EB
Fatty Acid and Phospholipid Metabolism									
Biotin Carboxyl Carrier Protein	*accB*	B0B9N1	CTL0378	CT123	18	5.1	1	G	EB
Acyl Carrier Protein Synthase	*fabF*	B0B8Z6	CTL0139	CT770	45	5.5	5	G,M	EB RB
Acyl Carrier Protein Synthase	*acpS*	B0B9K8	CTL0355	CT100	13	9.4	4	G,M	EB
Glycerol 3 P Acyltransferase	*plsC*	B0B826	CTL0713	CT453	24	10.0	1	G	EB
Lipoamide Dehydrogenase	*lpdA*	B0B8D1	CTL0820	CT557	50	6.7	1	G	EB
Malonyl CoA Acyl Carrier Transacylase	*fabD*	B0B7F4	CTL0490	CT238	34	4.9	3	G	EB
predicted acyltransferase family		B0B9V6	CTL0458	CT206	32	6.0	1	M	RB
FA Phospholipid Synthesis Protein	*plsX*	B0B939	CTL0182	CT811	34	6.6	1	G	EB
Enoyl Acyl Carrier Protein Reductase	*fabI*	B0B9L2	CTL0359	CT104	32	5.6	7	G,M	EB RB
Oxoacyl Carrier Protein Reductase	*fabG*	B0B7F3	CTL0489	CT237	26	8.3	5	G	EB
Oxoacyl Carrier Protein Synthase III	*fabH*	B0B7F5	CTL0491	CT239	35	7.6	2	G	EB
CDP diacylglycerol glycerol 3 phospahte 3 phoasphatidyltransferase	*pgsA*	B0B924	CTL0166	CT797	23	8.9	1	G	EB
predicted Lysophospholipase esterase		B0B9P4	CTL0391	CT136	27	5.5	2	G	EB
Phosphatidate Cytidylytransferase	cdsA	B0B824	CTL0711	CT451	34	8.9	2	G	EB
Acyl Carrier UDP GlcNAc O Acyltransferase	*ipxA*	B0B8A5	CTL0793	CT531	31	6.1	2	G	EB
AcCoA Carboxylase Transferase Alpha	*accA*	B0B7I1	CTL0517	CT265	36	5.9	1	G	EB
Acyl CoA Thioester Hydrolase	*vidD*	B0B8A9	CTL0797	CT535	19	18.5 8.75	1	G	EB
Acylglycerophosphoethanolamine Acyltransferase	*aas*	B0B902	CTL0145	CT776	59	7.6	3	G	EB
Purines, Pyrimidines, Nucleosides and Nucleotides									
AMP Nucleosidase	*amn*	B0B8X7	CTL0120	CT751	32	6.6	1	G	EB
CTP Synthetase	*pyrG*	B0B9T3	CTL0435	CT183	60	6.2	3	G	EB
Nucleoside 2 P Kinase	*ndk*	B0B874	CTL0762	CT500	15	5.3	1	G	EB
Thymidylate Kinase	*tdk*	B0B9T8	CTL0440	CT188	23	7.6	1	G	EB
UMP Kinase	*pyrH*	B0B8Q4	CTL0047	CT678	26	5.4	3	G	EB
Ribonucleoside Reductase Large Chain	*nrdA*	B0B956	CTL0199	CT827	119	6.1	1	G,M	RB
dUTP Nucleotidohydrolase	*dut*	B0B7K8	CTL0544	CT292	15	5.3	1	G	EB
Deoxycytidine triphosphate deaminase family protein	*dcd*	B0B9E6	CTL0294	CT039	21	4.9	1	G	EB
adenylate kinase	*adk*	B0B9N6	CTL0383	CT128	28	4.8	1	G	EB
Regulatory Functions									
General Stress Protein	*rplY*	B0B926	CTL0168	CT799	20	9.2	2	G	EB RB
GTPase	*lepA*	B0B9H2	CTL0320	CT064	67	6.3	2	G	EB
HTH Transcriptional Regulatory Protein and Receiver Domain	*tctD*	B0B8K5	CTL0894	CT630	26	8.9	2	G	EB
Replication									
DNA Gyrase Subunit A	*gyrA*	B0B9T9	CTL0441	CT189	94	6.6	4	G	EB RB
DNA Gyrase Subunit B	*gyrB2*	B0B9U0	CTL0442	CT190	90	5.5	9	G,M	EB RB
DNA Pol III Epsilon chain	*dnaQ*	B0B7H7	CTL0513	CT261	27	5.8	1	G	EB
DNA Polymerase I	*polA*	B0B866	CTL0754	CT493	97	5.4	1	M	EB
DNA Topoisomerase I Fused to SWI Domain	*topA*	B0B8L8	CTL0011	CT643	97	9.1	2	G	EB
Endonuclease IV	*nfo*	B0B8K0	CTL0889	CT625	32	6.0	1	G	EB
DNA Helicase	*uvrD*	B0B8I3	CTL0872	CT608	73	6.4	2	G,M	EB

ssDNA Exonuclease	*recJ*	B0B820	CTL0707	CT447	65	9.5	2	G	EB
Transcription									
Polyribonucleotide Nucleotidyltransferase	*pnp*	B0B971	CTL0214	CT842	76	5.7	6	G	EB RB
Transcription antitermination factor	*nusA*	B0B9K5	CTL0352	CT097	49	5.3	2	M	EB RB
RNA Polymerase Alpha	*rpoA*	B0B881	CTL0769	CT507	42	5.4	6	G,M,2D(17)	EB RB
RNA Polymerase Beta	*rpoB*	B0B7N1	CTL0567	CT315	140	5.8	14	G,M	EB RB
RNA Polymerase Beta	*rpoC*	B0B7N0	CTL0566	CT314	155	7.5	21	G,M,2D(18)	EB RB
RNA Polymerase Sigma 66	*rpoD*	B0B8J0	CTL0879	CT615	66	8.4	3	G	EB RB
Transcription Elongation Factor G	*greA*	B0B8L1	CTL0004	CT636	81	5.3	4	G,M	EB
Transcription Termination Factor	*rho*	B0B864	CTL0752	CT491	52	7.3	9	G,M	EB RB
Transcriptional termination protein	*nusG*	B0B7N6	CTL0572	CT320	21	5.3	2	G,2D(21)	EB
Translation									
CLP Protease	*clpP*	B0B803	CTL0690	CT431	21	5.5	3	G,M	EB RB
Clp Protease ATPase	*clpB*	B0B9M1	CTL0368	CT113	96	5.4	11	G,M	EB RB
ClpC Protease ATPase	*clpC*	B0B7K2	CTL0538	CT286	95	6.4	12	G,M	EB
General Stress Protein	*ipiY*	B0B926	CTL0168	CT799	20	9.0	3	G	EB
Elongation Factor P	*efp*	B0B9N0	CTL0377	CT122	21	5.0	2	G	EB
Elongation Factor P	*efp*	B0B8X8	CTL0121	CT752	21	5.0	1	G	EB
Elongation Factor TS	*tsf*	B0B8Q5	CTL0048	CT679	31	5.4	5	G,M,2D(5)	EB RB
Elongation Factor Tu	*tufA*	B0B7N8	CTL0574	CT322	43	5.5	17	G,M,2D(6,7,8)	EB RB
Elongation Factor G	*fusA*	B0B809	CTL0696	CT437	76	5.3	12	G,M	EB RB
Arginyl tRNA Transferase	*argS*	B0B827	CTL0714	CT454	63	6.4	2	G,M	EB
Aspartyl tRNA Synthetase	*aspS*	B0B8B6	CTL0804	CT542	66	5.2	3	G,M	EB
Alanyl tRNA Synthetase	*alaS*	B0B8X5	CTL0118	CT749	98	5.5	3	G,M	EB RB
DO Serine Protease	*htrA*	B0B952	CTL0195	CT823	53	6.8	5	G	EB
Tryptophanyl tRNA Synthetase	*trpS*	B0B8F9	CTL0848	CT585	40	6.9	6	G,M	EB RB
tyrosyl tRNA Synthetase	*tyrS*	B0B9H0	CTL0318	CT062	45	7.1	2	G	EB
Initiation Factor 3	*infC*	B0B962	CTL0205	CT833	20	9.9	1	G	EB
Initiation Factor IF 1	*infA2*	B0B7N9	CTL0575	CT323	8	9.4	1	M	EB
Leucyl Aminopeptidase A	*pepA*	B0B9F3	CTL0301	CT045	54	5.9	11	G	EB
Protease	*sohB*	B0B867	CTL0755	CT494	36	8.4	3	G	EB RB
Glu tRNA Gln Amidotransferase A subunit	*gatA*	B0B9B0	CTL0258	CT003	54	6.1	5	G,M	EB RB
Pet1 12 Glu tRNA Gln Amidotransferase B Subunit	*gatB*	B0B9B1	CTL0259	CT004	55	6.2	4	G,M	EB RB
Lon ATP dependent protease	*lon*	B0B7R3	CTL0598	CT344	92	6.9	2	G,M	EB
Insulinase family Protease III	*ptr*	B0B933	CTL0175	CT806	108	5.1	14	G,M	EB RB
Metalloprotease		B0B9I0	CTL0328	CT072	69	6.6	3	G	EB RB
Metalloprotease	*ispH*	B0B991	CTL0234	CT859	34	6.1	1	G,M	EB,RB
Threonyl tRNA Synthetase	*thrS*	B0B8F5	CTL0844	CT581	73	6.1	1	G,M	EB
Seryl tRNA Synthetase	*serS*	B0B8V5	CTL0098	CT729	48	5.9	1	G	EB
Leucyl tRNA Synthetase	*leuS*	B0B9V9	CTL0461	CT209	93	5.8	1	M	EB
Glutamyl tRNA Synthetase	*gltX*	B0B818	CTL0705	CT445	59	6.5	2	G	EB
Histidyl tRNA Synthetase	*hisS*	B0B8B7	CTL0805	CT543	49	7.1	1	G	EB
Axial Filament Protein	*cafE*	B0B935	CTL0177	CT808	59	7.1	1	G	EB RB
Oligoendopeptidase	*pepF*	B0B9M0	CTL0367	CT112	69	5.6	5	G	EB RB
Peptidyl tRNA Hydrolase	*pth*	B0B927	CTL0169	CT800	20	8.3	1	G	EB
Ribosome Releasing Factor	*rrf*	B0B8Q3	CTL0046	CT677	20	8.9	1	G	EB
Thiol: disulfide Interchange Protein	*dsdD*	B0B8G9	CTL0859	CT595	76	6.7	3	G	EB
Glycyl tRNA Synthetase	*glyQ*	B0B923	CTL0165	CT796	113	5.9	1	G	EB
Lysyl tRNA Synthetase	*lysS*	B0B907	CTL0150	CT781	60	5.4	1	G	EB
CLP Protease	*clpP*	B0B8T2	CTL0075	CT706	22	5.2	2	G	EB
rRNA methylase	*troB*	B0B9H6	CTL0324	CT068	29	7.2	2	G,M	EB RB
Polypeptide Deformylase	*def*	B0B7S2	CTL0607	CT353	21	5.9	1	G	EB
L1 Ribosomal Protein	*rplA*	B0B7N4	CTL0570	CT318	25	9.3	2	G,M	EB RB

L10 Ribosomal Protein	*rplJ*	B0B7N3	CTL0569	CT317	19	6.7	2	G,M	EB RB
L13 Ribosomal Protein	*rplM*	B0B9N3	CTL0380	CT125	17	10.5	2	G	EB
L14 Ribosomal Protein	*rplN*	B0B892	CTL0780	CT518	13	10.2	2	G	EB
L15 Ribosomal Protein	*rplO*	B0B885	CTL0773	CT511	16	10.5	3	G,M	EB RB
L16 Ribosomal Protein	*rplP*	B0B895	CTL0783	CT521	16	11.4	2	G	EB
L17 Ribosomal Protein	*rplQ*	B0B880	CTL0768	CT506	16	11.4	2	G	EB
L18 Ribosomal Protein	*rplR*	B0B887	CTL0775	CT513	13	10.7	2	G,M	EB RB
L19 Ribosomal Protein	*rplS*	B0B9D5	CTL0283	CT028	13	10.3	2	G	EB
L2 Ribosomal Protein	*rplB*	B0B899	CTL0787	CT525	31	10.8	4	G,M	EB RB
L21 Ribosomal Protein	*rplU*	B0B7Z0	CTL0677	CT420	12	9,53	2	G	EB
L22 Ribosomal Protein	*rplV*	B0B897	CTL0785	CT523	12	11.4	1	M	EB
L23 Ribosomal Protein	*rplW*	B0B8A0	CTL0788	CT526	12	10.3	2	G	EB
L24 Ribosomal Protein	*rplX*	B0B891	CTL0779	CT517	13	10.9	2	G,M	EB
L28 Ribosomal Protein	*rpmB*	B0B9J4	CTL0341	CT086	10	11.7	2	G	EB
L29 Ribosomal Protein	*rpmC*	B0B894	CTL0782	CT520	8	10.3	2	G	EB
L33 Ribosomal Protein	*rpmG*	B0B9Q8	CTL0405	CT150	6	10.7	1	G	EB
L4 Ribosomal Protein	*rplD*	B0B8A1	CTL0789	CT527	25	10.1	6	G,M	EB RB
L5 Ribosomal Protein	*rplE*	B0B890	CTL0778	CT516	21	9.7	3	G	EB
L6 Ribosomal Protein	*rplF*	B0B888	CTL0776	CT514	20	10.3	3	G,M	EB RB
L7 L12 Ribosomal Protein	*rplL*	B0B7N2	CTL0568	CT316	14	4.9	3	M	EB RB
L9 Ribosomal Protein	*rplI*	B0B930	CTL0172	CT803	18	6.5	1	G	EB
S1 Ribosomal Protein	*rpsA*	B0B9K6	CTL0353	CT098	64	5.3	8	G,M,2D(19)	EB RB
S10 Ribosomal Protein	*rpsJ*	B0B808	CTL0695	CT436	12	10.8	6	G	EB
S11 Ribosomal Protein	*rpsK*	B0B882	CTL0770	CT508	14	11.3	2	G,M	EB RB
S12 Ribosomal	*rpsL*	B0B811	CTL0698	CT439	15	11.3	1	G	EB
S13 Ribosomal Protein	*rpsM*	B0B883	CTL0771	CT509	14	11.2	1	G	EB
S14 Ribosomal Protein	*rpsN*	B0B913	CTL0155	CT787	12	11.5	2	G	EB
S16 Ribosomal Protein	*rpsP*	B0B9D3	CTL0281	CT026	13	10.6	1	G	EB
S17 Ribosomal Protein	*rpsQ*	B0B893	CTL0781	CT519	10	10.7	1	G	EB
S18 Ribosomal Protein	*rpsR*	B0B929	CTL0171	CT802	9	11.5	1	G	EB
S2 Ribosomal Protein	*rpsB*	B0B8Q6	CTL0049	CT680	31	6.6	3	G,M	EB RB
S3 Ribosomal Protein	*rpsC*	B0B896	CTL0784	CT522	24	10.3	4	G,M	EB RB
S4 Ribosomal Protein	*rpsD*	B0B8K1	CTL0890	CT626	24	10.3	2	G,M	EB RB
S5 Ribosomal Protein	*rpsE*	B0B886	CTL0774	CT512	18	10.2	1	G	EB
S6 Ribosomal Protein	*rpsF*	B0B928	CTL0170	CT801	13	9.0	2	G,M	EB,RB
S7 Ribosomal Protein	*rpsG*	B0B810	CTL0697	CT438	17	10.1	3	G,M	EB RB
S8 Ribosomal Protein	*rpsH*	B0B889	CTL0777	CT515	15	10.5	1	G	EB
S9 Ribosomal Protein	*rpsI*	B0B9N4	CTL0381	CT126	15	11.2	3	G,M	EB RB
Ribosome Binding Factor A	*rbfA*	B0B9K3	CTL0350	CT095	14	9.3	1	G	EB
Transport and Binding Proteins									
ABC Transport ATPase	*dppD*	B0B8R6	CTL0059	CT690			1	G	EB
ABC Transporter ATPase	*ycfV*	B0B9R0	CTL0407	CT152	25	6.9	1	G	EB
ABC Transporter Protein	yjjK	B0B7R7	CTL0602	CT348	59	5.7	1	G	EB
PTS IIA Protein HTH DNA Binding Domain	ptsN	B0B7K6	CTL0542	CT290	26	5.3	2	M	RB
Arginine Binding Protein	*artJ*	B0B7V0	CTL0636	CT381	29	5.8	8	G,M,2D(3)	EB
Solute Protein Binding Family	troA	B0B9H5	CTL0323	CT067	37	5.7	3	G,M	EB RB
Sodium dependent amino acid transporter		B0B9Y2	CTL0483	CT231	55	9.3	1	M	RB
oligo Binding Lipoprotein	*oppA4*	B0B853	CTL0741	CT480	80	5.1	1	G	RB
Protein Translocase	*secA*	B0B8S7	CTL0070	CT701	110	6.0	7	G,M	EB RB
Glutamine Binding protein	*fliY*	B0B859	CTL0747	CT486	29	7.5	6	G	EB RB
Tyrosine Transport	*tyrP*	B0B947	CTL0190	CT818	44	9.3	2	G	EB
Hexosphosphate Transport	*uhpC*	B0B8B8	CTL0806	CT544	52	8.6	3	G,M	EB RB
Dicarboxylate Translocator	*ybhI*	B0B9V4	CTL0456	CT204	51	9.2	1	G	EB
Amino Acid Transporter	*xasA*	B0B9W6	CTL0468	CT216	51	9.1	1	G	EB
Mg Transporter CBS Domain	*mgtE*	B0B9U4	CTL0446	CT194	51	5.0	1	G	EB
ABC Amino Acid Transporter ATPase	*glnQ*	B0B9N8	CTL0385	CT130	26	6.3	1	M	EB

Metal Transport P type ATPase	*zntA*	B0B8V3	CTL0096	CT727	71	7.2	1	G	EB
Hypothetical Proteins									
hypothetical protein		B0B7M7	CTL0563	CT311			1	M	EB
hypothetical protein		B0B9A8	CTL0256	CT001	10	9.0	1	G	EB
hypothetical protein		B0B9B8	CTL0266	CT011	48	7.3	1	G	EB
hypothetical protein		B0B9B9	CTL0267	CT012	30	10.0	1	G	EB
hypothetical protein		B0B9D8	CTL0286	CT031	12	10.0	1	G	EB
hypothetical protein		B0B9F1	CTL0299	CT043	18	5.1	6	G	EB
hypothetical protein		B0B9G4	CTL0312	CT056	27	9.5	1	G	EB
hypothetical protein		B0B9H4	CTL0322	CT066	18	10.3	1	G	EB
hypothetical protein		B0B9L0	CTL0357	CT102	17	9.9	2	G,M	EB
hypothetical protein		B0B9R1	CTL0408	CT153	91	6.4	4	G	EB
hypothetical protein		B0B9T1	CTL0433	CT181	27	5.5	1	G	EB
hypothetical protein		B0B9W2	CTL0464	CT212	17	5.0	1	G	EB
hypothetical protein		B0B9X1	CTL0473	CT221	33	7.3	4	G	EB
hypothetical protein		B0B9X4	CTL0476	CT223	30	8.3	3	G,M	RB
hypothetical protein		B0B7F0	CTL0486	CT234	106	8.8	1	G	EB
hypothetical protein		B0B7G9	CTL0505	CT253	24	9.2	5	G,M	EB RB
hypothetical protein		B0B7H6	CTL0512	CT260	19	4.8	3	M	EB
hypothetical protein		B0B7K4	CTL0540	CT288	63	8.1	2	G	EB
hypothetical protein		B0B7M7	CTL0563	CT311	26	9.1	1	M	EB
hypothetical protein	*aaxA*	B0B7U1	CTL0626	CT372	49	9.2	2	G	EB
hypothetical protein		B0B7V7	CTL0643	CT387	77	6.2	1	G	EB
hypothetical protein		B0B7V8	CTL0644	CT388	13	10.2	1	G	EB
hypothetical protein		B0B7V9	CTL0645	CT389	47	5.8	1	G	EB
hypothetical protein		B0B7W8	CTL0655	CT398	30	7.7	6	G,M	EB
hypothetical protein	*nrdR*	B0B7X6	CTL0663	CT406	18	9.1	1	G	EB
hypothetical protein		B0B7Z1	CTL0678	CT421	26	9.4	1	G	EB
hypothetical protein		B0B7Z7	CTL0684	CT425	70	5.0	2	G,M	EB
hypothetical protein		B0B801	CTL0688	CT429	39	5.1	1	G	EB
hypothetical protein		B0B845	CTL0733	CT472	30	5.3	1	G	EB RB
hypothetical protein		B0B849	CTL0737	CT476	36	8.2	1	M	RB
hypothetical protein		B0B877	CTL0765	CT503	22	5.7	1	G	EB
hypothetical protein		B0B8B2	CTL0800	CT538	27	5.2	7	G,M,2D(12)	EB
hypothetical protein		B0B8D4	CTL0823	CT560	32	7.0	2	G	EB
hypothetical protein		B0B8F1	CTL0840	CT577	13	7.0	3	G	EB
hypothetical protein	*copB*	B0B8F2	CTL0841	CT578	50	9.4	2	G	EB
hypothetical protein	*copD*	B0B8F3	CTL0842	CT579	44	9.6	3	G	EB
hypothetical protein		B0B8F8	CTL0847	CT584	21	5.8	5	G,M,2D(11)	EB RB
hypothetical protein		B0B8G4	CTL0853	CT590	109	5.8	3	G	EB
hypothetical protein		B0B8I5	CTL0874	CT610	27	4.9	3	G,M	EB RB
hypothetical protein		B0B8I6	CTL0875	CT611	27	5.6	2	G	EB
hypothetical protein		B0B8J3	CTL0882	CT618	28	9.1	2	G,M	RB
hypothetical protein		B0B8J6	CTL0885	CT621	93	5.0	1	G	EB
hypothetical protein		B0B8K6	CTL0895	CT631	9	5.3	1	G	EB
hypothetical protein		B0B8K7	CTL0897	CT632	61	6.0	6	G	EB
hypothetical protein		B0B8L0	CTL0003	CT635	17	6.5	1	G	EB
hypothetical protein		B0B8L7	CTL0010	CT642	32	9.4	1	G	EB
hypothetical protein	*recA*	B0B8M5	CTL0018	CT650	38	7.6	5	G,M	EB RB
hypothetical protein		B0B8N5	CTL0028	CT659	88	8.7	4	G	EB
hypothetical protein		B0B8P2	CTL0035	CT666	92	4.7	2	G	EB
hypothetical protein		B0B8P7	CTL0040	CT671	31	4.8	2	M	EB
hypothetical protein		B0B8Q2	CTL0045	CT676	20	5.8	2	M	EB
hypothetical protein		B0B8R7	CTL0060	CT691	25	5.0	1	G	EB
hypothetical protein		B0B8S0	CTL0063	CT694	35	5.4	2	G	EB
hypothetical protein		B0B8S1	CTL0064	CT695	44	5.2	1	G	EB

hypothetical protein		B0B8V4	CTL0097	CT728	28	6.2	1	G	EB
hypothetical protein		B0B8W7	CTL0110	CT741	13	9.5	3	G,M	EB RB
hypothetical protein		B0B8Y9	CTL0132	CT763	15	4.7	1	G,M	EB RB
hypothetical protein		B0B8Z4	CTL0137	CT768	64	5.4	1	G	EB
hypothetical protein		B0B916	CTL0158	CT790	18	4.6	1	G,M	EB RB
hypothetical protein		B0B932	CTL0174	CT805	52	9.6	1	G	EB
hypothetical protein		B0B942	CTL0185	CT814	16	11.1	1	G	EB
hypothetical protein		B0B966	CTL0209	CT837	76	6.1	2	G	EB
hypothetical protein		B0B975	CTL0218	CT846	27	9.8	1	G	EB
hypothetical protein		B0B978	CTL0221	CT849	18	5.1	1	G	EB
hypothetical protein		B0B983	CTL0226	CT853	22	9.0	1	G	EB
hypothetical protein		B0B9A7	CTL0255	CT875	66	6.2	12	G	EB
hypothetical protein		B0B9E9	CTL0297	CT041	7	4.0	2	G	EB RB
hypothetical protein		B0B9Q5	CTL0402	CT147	162	8.6	3	G	RB
hypothetical protein		B0B9E9	CTL0297	CT041	30	5.0	2	G	EB RB
Other									
FHA domain to adenylate cyclase		B0B8P0	CTL0033	CT664	90	4.6	12	G,M	EB RB
FKBP type peptidyl prolyl cis trans isomerase	*mip*	B0B8B5	CTL0803	CT541	26	5.0	9	G,M,2D(9)	EB RB
Predicted metal dependent hydrolase		B0B7V6	CTL0642	CT386	33	5.2	2	G	EB
predicted phosphatase kinase		B0B865	CTL0753	CT492	23	5.7	1	G	EB
Serine threonine Protein Kinase	*pkn1*	B0B9Q3	CTL0400	CT145	70	5.1	1	M	EB RB
SurE like Acid Phosphatase	*surE*	B0B9W8	CTL0470	CT218	32	5.0	1	G	EB
Intergration Host Factor Alpha	*ihfA*	B0B7I3	CTL0519	CT267	11	11.2	4	G	EB
Histone Like Development Protein	*hctA*	B0B8W9	CTL0112	CT743	14	11.0	5	G,M	EB
SWF SNF family helicase	*snf*	B0B8T4	CTL0077	CT708	133	5.4	2	G,M	EB
SWIB YM74 complex		B0B833	CTL0720	CT460	10	9.4	1	G	EB
Yeb C family		B0B830	CTL0717	CT457	27	5.8	3	G	EB
SuA5 Superfamily related Protein		B0B9P5	CTL0392	CT137	31	6.5	1	G	EB
CHLPN 76kDa Homolog		B0B8J7	CTL0886	CT622	68	4.8	1	G	EB
CHLPN 76kDa Homolog		B0B8J8	CTL0887	CT623	48	8.5	16	G,M	EB RB
hydrolase phosphatase homolog		B0B8Z7	CTL0140	CT771	17	5.1	2	G,M	EB,RB
Hit Family Hydrolase	hitA	B0B7V5	CTL0641	CT385	12	5.4	1	M	EB
ACR family	ybgI	B0B9L6	CTL0363	CT108	27	6.0	2	G	EB
phosphohydrolase	icc	B0B8Y0	CTL0123	CT754	33	7.8	1	G	EB
HAD superfamily hydrolase phosphatase		B0B9L1	CTL0358	CT103	34	5.1	2	G	EB
Phosphoglycolate Phosphatase		B0B837	CTL0724	CT464	26	5.2	1	G	EB
Phosphohydrolase	yaeI	B0B834	CTL0721	CT461	37	9.3	1	G	EB

[a] pI and molecular mass were calculated using ProteinLynx Global Server 2.05 (Waters)

[b] The abbreviations for the techniques used are:

G = GeLC-MS/MS

M = MudPIT

2D = 2-DGE

4.2.4 Analysis of the chlamydial proteome by GeLC-MS/MS

To further extend proteome coverage of EBs and RBs, the technique GeLC-MS/MS was employed. EB and RB samples were solubilised in final SDS-PAGE sample buffer and fractionated by conventional SDS-PAGE. The protein separation profile of EBs and RBs obtained by 1-D SDS-PAGE is shown in **Figure 4.3**. Each track of the gel containing the relevant samples was excised into 25 gel bands of equal size and digested *in situ* using trypsin. Peptides extracted from each gel band were identified using nanoLC-MS/MS whereupon all spectra were processed and searched in a similar manner to those obtained via the MudPIT approach.

A total of 30,042 spectra were obtained for EB samples, resulting in 824 peptide assignments and the identification of 298 unique proteins when searched against the serovar D genome. As found in the MudPIT experiments, fewer proteins were identified in RBs, identifying 65 proteins.

The completion of the *C. trachomatis* L2/434/Bu genome during these studies has offered the opportunity to re-search all the MS/MS data obtained from the L2 strain, against the cognate genome data. Re-searching of the GeLC-MS/MS data revealed an additional 5 proteins identified with ≥ 2 peptides, increasing the total proteome coverage of *C. trachomatis* to 36%. Although 2-DGE and MudPIT re-searches did not assign any further proteins, protein sequence coverage was significantly improved across the dataset, for example, the sequence coverage for MOMP (CTL0050), increased from 36% (serovar D) to 83.5% (L2/434/Bu). All protein assignments listed in **Table 4.2** are from searches against the cognate strain, *C. trachomatis* L2/434/Bu.

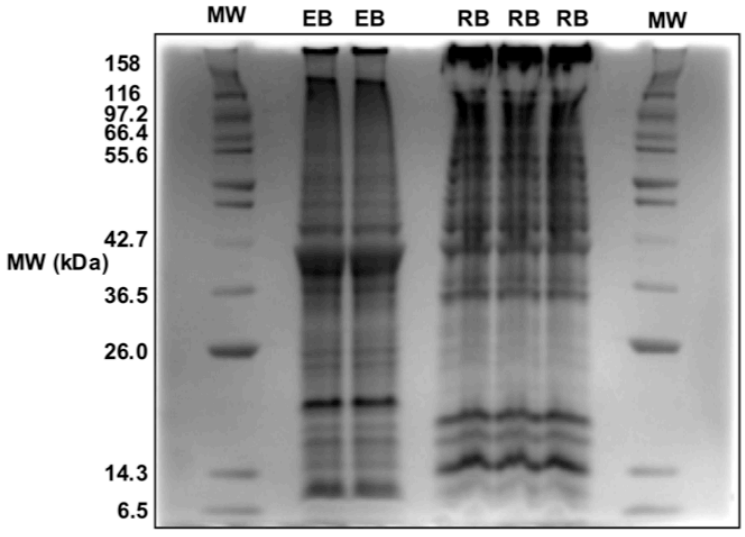


Figure 4.3. Fractionation of EB and RB protein lysates using a NuPAGE 4 – 12% SDS – polyacrylamide gel (Invitrogen, Paisley, UK). Proteins were visualised with Colloidal Coomassie Brilliant Blue. MW = Broad range protein markers (New England Biolabs, UK); EB = 130 µg of EB lysate; RB = 130 µg of RB lysate.

4.3 Discussion

This chapter has described the qualitative proteomic analysis of purified EBs and RBs from *C. trachomatis* L2 using the proteomic technologies 2-D gel electrophoresis, MudPIT and GeLC-MS/MS (O'Farrell, 1975; Link *et al*., 1999; Opiteck *et al*., 1997; Schirle *et al*., 2003). A comparison of the proteomic technologies used in this study, and the roles of some of the identified proteins or groups of proteins in the context of *Chlamydia* biology are now discussed.

4.3.1 Comparison of the sampling characteristics of 2-D gel electrophoresis, MudPIT and GeLC-MS/MS

The bias of 2-DGE towards high abundance proteins resulted in poor proteome coverage of both EB and RB samples. In contrast, protein identification using the MudPIT approach was relatively unbiased and provided increases in proteome coverage. For example, it detected over three times as many predicted membrane proteins and seven-fold more proteins in the transcription and translation categories when compared to 2-D gel electrophoresis (**Table 4.2**).

Analysis by GeLC-MS/MS provided over a two-fold increase in proteome coverage compared to analysis by MudPIT. GeLC-MS/MS identified the entire set of proteins identified by 2D-GE. Similarly, all but 26 out of 107 proteins identified by MudPIT were identified in the GeLC-MS/MS study. A conclusion drawn from this study is that GeLC-MS/MS is significantly more efficient than the two other approaches in detecting components of the chlamydial proteome.

Like the MudPIT approach, GeLC-MS/MS sampled in a relatively unbiased manner representing membrane proteins, low abundance proteins, high molecular weight proteins, and proteins with extreme p*I*s (**Table 4.2**, **Figure 4.4** and **Figure 4.5**). For example, it detected a 186 kDa outer membrane protein (encoded by the *CTL0671* gene) as well as a range of inner membrane proteins (**Table 4.2**). The procedure has allowed the identification of proteins with atypical codon biases, indicative of extremes of expression (**Figure 4.4**). Similarly, the most acidic and basic proteins identified by GeLC-MS/MS have p*I*'s of 4.01 and 11.7 (**Figure 4.5**) respectively, which is in reasonable agreement with the predicted p*I* range of 3.83 to 12.65 for the proteome of serovar L2.

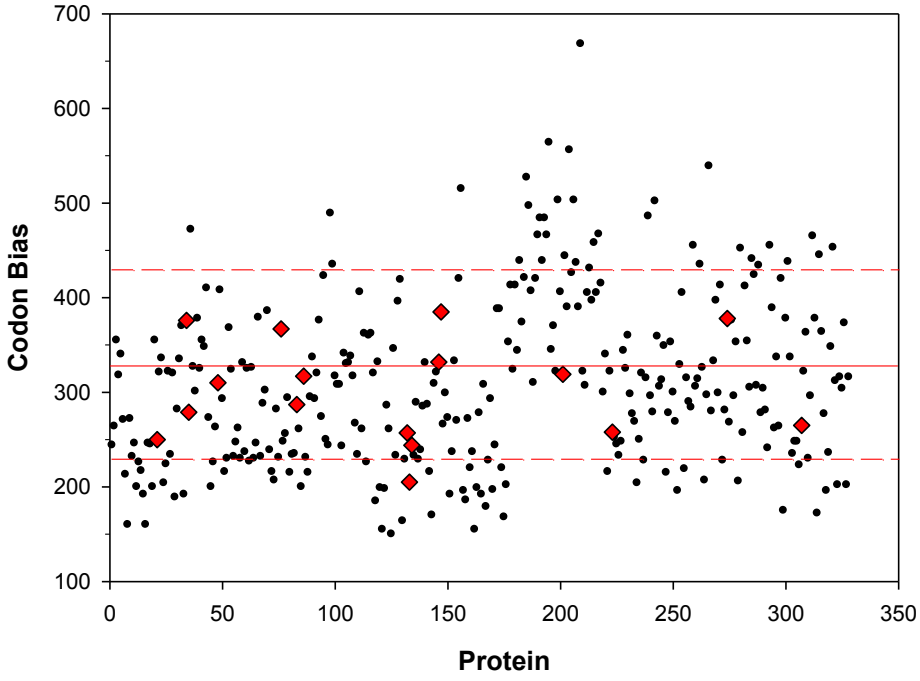


Figure 4.4. Experimentally detected chlamydial proteins with atypical codon biases.
A plot of the codon bias for each gene identified relative to the mean average codon bias for the entire genome (solid horizontal Line on graph). Codon bias values that exceed one SD of the average codon bias (dashed horizontal lines) are indicative of proteins with atypical levels of expression. Codon bias values were calculated by the method of Karlin *et al* (2001). ●, proteins identified by MudPIT or GeLC-MS/MS; ♦, proteins identified by 2-D gel electrophoresis.

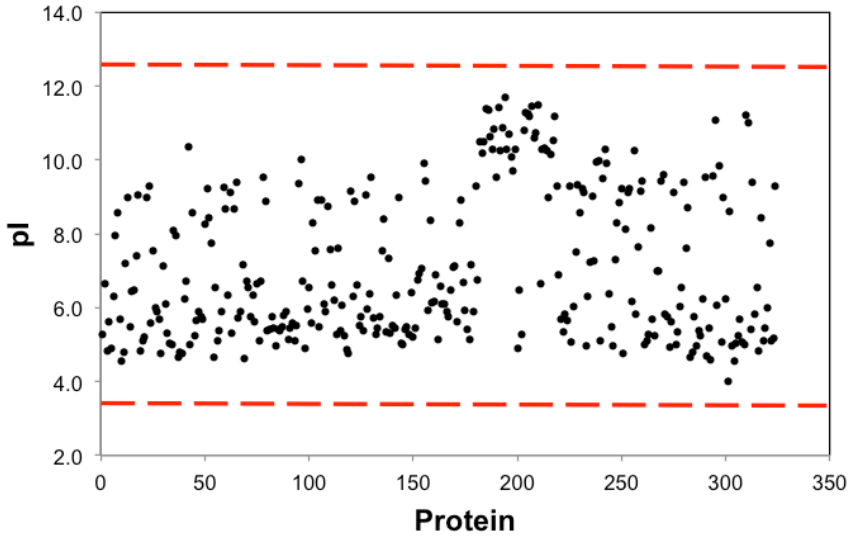


Figure 4.5. p*I* Distribution plot of the proteins identified from *C. trachomatis* serovar L2. p*I* values were determined using the database searching package PLGS 2.2 (Waters, Manchester, UK). The dashed line indicates the predicted p*I* range for *C. trachomatis* L2.

4.3.2 Shotgun identification of proteins from *C. trachomatis* and their context in *Chlamydia* biology

Generally, only tentative conjectures about an organism's cellular processes can be made using genome sequencing data. Genomic transcriptional analyses have become an important tool in studying gene expression especially as prokaryotic gene expression is normally regulated at the transcriptional level. In *Chlamydia*, the developmental cycle appears to be regulated by subsets of temporally expressed genes (Belland *et al*., 2003; Nicholson *et al* 2003). However, verification of the presence of a component at the protein level significantly strengthens the inferences that can be made. This study has identified protein products from all the functional categories of protein-encoding genes and thus provides direct evidence for these processes and pathways (**Table 4.2**). The functional significance of the identified components is discussed in the following sections.

4.3.2.1 Energy metabolism

Traditionally chlamydiae have been considered 'energy parasites, unable to synthesise their own ATP (Moulder, 1962). Evidence for this was provided by the observation that isolated RBs can transport ATP and ADP by an ATP-ADP exchange mechanism (Hatch *et al*., 1982). In this respect, this study has identified in both EBs and RBs, two chlamydial translocases, predicted from the genome sequence and subsequently characterised by cloning and expression in *E. coli* (Tjaden *et al*., 1999). The gene *CTL0024* encodes a protein responsible for exchanging ATP and ADP while the protein encoded by *CTL0756* catalyses the net import of ribonucleoside tri-phosphates. However, the present study also provides evidence to suggest that *Chlamydia* are capable of producing their own energy (Iliffe-Lee and McClarty, 1999) because many of the energy-generating enzymes of central metabolism are present in both EBs and RBs. The entire complement of glycolytic enzymes required to convert glucose-6-phosphate to pyruvate have been detected. Included among these enzymes was the recently predicted fructose-bisphosphate aldolase, the product of *gene CTL0467*, where recent controversy over its presence has led to the proposal of an alternative route via the pentose-phosphate pathway to circumvent this enzyme (Stephens *et al*., 1998). Additionally, all but one of the enzymes associated with the pentose phosphate pathway are present in this proteomic dataset. The fact that all of the glycolytic enzymes were readily detectable in the metabolically inert EBs suggest that this pathway is preassembled, rather than synthesised *de novo* in host cells, and suggests that metabolite flux through the pathway is only triggered upon infection of the host cell. Also identified in this study were several protein subunits encoded by genes of the adenosine triphosphatase (ATPase) operon. However, chlamydiae with mammalian hosts appear to have lost genes for the F_1 ATPase during evolution (Horn *et al*., 2004). It is thus, unlikely, that these ATPase components are involved in energy generation, especially as most of the components identified resemble the vacuolar ATPases (McClarty and Stephens, 1999).

In addition to components of the glycolytic and pentose phosphate pathways, the alpha and beta subunits of succinyl CoA synthetase from the TCA cycle and pyruvate dehydrogenase, a key linker reaction enzyme were found (**Table 4.2**). Although genetic evidence shows that the TCA cycle is incomplete in *Chlamydia*

(Stephens *et al*., 1998), the presence of these enzymes supports the concept that a modified pathway operates. The ability of cells to synthesise glucose is a major requirement and detected in these datasets was the gluconeogenesis enzyme phosphoenolpyruvate carboxykinase. This enzyme catalyses the conversion of oxaloacetate to phosphoenolpyruvate, bypassing the irreversible glycolytic step of pyruvate kinase (Iliffe-Lee and McClarty, 1999). In the absence of a complete TCA cycle phosphoenolpyruvate carboxykinase provides the only direct link between the TCA cycle and glycolysis.

One of the properties used to differentiate *C. trachomatis* from other species of *Chlamydia* has been the detection of glycogen within its inclusions, using the simple and long- established iodine test (Chiappino *et al*., 1995; Iliffe-Lee and McClarty, 2000). This differentiates on the basis that, unlike *C. trachomatis*, other chlamydial species do not synthesise detectable amounts of glycogen. The *C. trachomatis* genome contains a complete complement of genes for both glycogen biosynthesis and metabolism (Stephens *et al*., 1998). Glycogen is usually stored under conditions of nitrogen starvation, thus it would seem reasonable that glycogen is synthesised in the later stages of development when EBs are forming. Accordingly, this study indicates the presence of the biosynthetic enzyme glycogen synthase in EBs. EBs must also be able to mobilise glycogen as an energy source early in the developmental cycle and presumably the glycogen phosphorylase detected in EBs is synthesised in preparation for this requirement.

4.3.2.2 The chlamydial cell envelope

The chlamydial cell envelope has been the focus of intense study due to its potential significance in vaccine development. An interesting anomaly arising from the genome sequencing project has been the discovery of genes for a complete set of peptidoglycan biosynthetic enzymes in the absence of any detectable cell wall material (Chopra *et al*., 1998). Although the expression levels of such components is likely to be low (Nicholson *et al*., 2003), especially in view of the difficulty in detecting chlamydial cell wall material, this study detected the presence in RBs of UDP-N-acetylmuramoylalanylglutamyl DAP ligase, which is encoded by the *murE* gene. This enzyme is involved in the synthesis of the muramyl-peptide unit, the first stage in assembling peptidoglycan (see **Chapter 5**). The discovery of this enzyme in RBs strongly hints that active peptidoglycan biosynthesis occurs during RB growth and cell division, supporting the notion that peptidoglycan is essential for progression through the developmental cycle (McCoy *et al*., 2003; McCoy and Maurelli, 2006). In addition, peptides from 17 predicted membrane proteins were detected and many of these matched proteins predicted to be located to the *C. trachomatis* outer membrane (Stephens and Lammel, 2001), including the major outer membrane protein (OmpA-CTL0050) and its analogue (PorB-CTL0082), which are both present in EBs and RBs (Kubo and Stephens, 2000). Consistent with previous studies, the 60 kDa cysteine-rich outer membrane protein (OmcB) and the peptidoglycan associated protein (PaI) were found only in EBs (Hatch *et al*., 1986). The complete family of 'Pmp' proteins, which have no homologs in other bacteria, were also detected. Strong evidence suggests that these proteins are autotransporters (Henderson and Lam, 2001) and are believed to be located on the chlamydial surface. Interestingly Pmps B, F and H have so far only been detected in EBs while Pmps C, D, E, G and I were found in both EBs and RBs. By contrast Pmp A was detected only in RBs. These results are in broad agreement with the observation that

Pmps E and especially G and H are abundantly expressed late in the developmental cycle (Mygind *et al*., 2000; Tanzer *et al*., 2001).

4.3.2.3 Type III secretion system (TTSS)

The TTSS in *Chalmydia* is likely to play a role in modifying the host cell processes that may be necessary for host cell invasion, restructuring of the inclusion membrane, or affecting host cell regulatory pathways. Seven proteins associated with the chlamydial TTSS were identified and, while their precise roles remain to be resolved (Subtil *et al*, 2000), they are representative of the cellular compartments associated with these systems. Thus, SctR (CTL0825) and LcrD (CTL0345) are believed to reside in the inner membrane, while SctJ (CTL0822) and SctC (CTL0043) have periplasmic locations and outer membrane locations respectively. Additionally, LcrH_1 (CTL0839), which functions as a cytoplasmic chaperone and a regulator in the TTSS of other bacteria, was identified as were the cytosolic proteins SctN (CTL0038) and SctL (CTL0824). The chlamydial SctJ and SctC proteins were detected in RBs purified at 15 h PI. These observations are consistent with evidence that shows transcription of CTL0822 and CTL0043 occurs in *C. trachomatis* L2 as early as 12 h PI and that SctJ is located to RB membranes. Furthermore, all seven proteins were found in EBs supporting the proposal that EBs contain a fully assembled and functional TTSS (Fields *et al*., 2003).

4.3.2.4 Hypothetical proteins

The largest group of proteins identified in this qualitative study were those designated as 'hypothetical' within the genome sequence. The detection of peptides representing 68 of these predicted proteins substantiates their existence and provides an impetus for more detailed studies to characterise their functions.

4.4 Summary

While the study presented in this chapter has identified proteins from all the major functional classes, including predicted low abundance proteins; components secreted from RBs were not detected. These include the 'Inc' related proteins that have been estimated to constitute up to 12% of the genome coding capacity (Rockey *et al*., 2000). It is likely that such proteins are efficiently exported during infection and hence are only present at reduced levels in EBs or RBs, unless present at high abundance during transit. Despite the absence of these proteins from this dataset, the identification of 321 proteins (~36% of the predicted proteome) of *C. trachomatis* L2 compares well with the current proteome coverage of other obligate intracellular pathogens. Examples include, *Rickettsia felis,* where 11.2% of the proteome was assigned using a combination of 2-DGE and GeLC-MS/MS (Ogawa *et al*., 2007); the foodbourne pathogen *Listeria monocytogenes,* where 245 proteins were identified from 1684 gel spots separated using two dimensional differential gel electrophoresis without apparent host cell contamination (Van de Velde *et al*., 2009); and Tucker *et al*. (2011), who compared *Rickettsia prowazekii* cultured in different host cells identifying between 102 to 178 proteins of the 835 proteins encoded by the genome depending upon the host cell background. The proteomic data captured in this study complements the genomic data, providing biological insights and

an essential framework for quantitative studies of the chlamydial developmental cycle as presented in Chapter 5.

CHAPTER 5

QUANTITATIVE ANALYSIS OF *C. TRACHOMATIS*

5.0 Quantitative analysis of *C. trachomatis*

5.1 Introduction

While the temporal aspects and morphological changes associated with the chlamydial developmental cycle have been well characterized using microscopy (Ward, 1988; Matsumoto, 1982), the underlying mechanisms that regulate and control the transition between the two distinct forms remains unclear. Genome sequencing and transcriptional profiling experiments with *C. trachomatis,* suggest that the cycle is coordinated by defined subsets of genes (Nicholson *et al*., 2003; Belland *et al*., 2003). The previous chapter presented a qualitative analysis of the proteome of *C. trachomatis* L2 using three different technologies, providing protein profiles of both RBs and EBs. Although this information has provided valuable insight into various aspects of *C. trachomatis*, there is a clear need for more quantitative data to obtain information on the temporal expression of proteins during the developmental cycle.

Traditionally, quantitative analysis of proteins has been performed using 2D-GE, providing relative quantification of proteins between different cellular states. However, as discussed in earlier chapters, there are a number of limitations with these gel-based methods, which has in recent years provided an impetus for the development of alternative quantitative technologies. These technologies can be divided into two different types of approach: (i) labelling, and (ii) non-labelling (label-free). Labelling approaches include stable isotope labelling with amino acids in cell culture (SILAC) (Ong *et al*., 2002), isotope-coded affinity tags (ICAT) (Gygi *et al*., 1999b), tandem mass tags (TMT) (Thompson *et al*., 2003), isotopically labeled peptide standards (AQUA) (Gerber *et al*., 2003) and isobaric tags for relative and absolute quantitation (iTRAQ) (Ross *et al*., 2004; Choe *et al*., 2007). Although these approaches have been widely used, the requirement for large amounts of sample, complex sample preparation, non-stoichiometric labeling and the high cost of the associated reagents, also limit these approaches.

To address some of these issues, focus has been directed towards label-free approaches. These have included the exponentially modified protein abundance index emPAI (Ishihama *et al*., 2005), which is an extension of the protein abundance index (Rappsilber *et al*., 2002), and provides the measurement of proteins in absolute amounts. Other approaches include accurate mass tags (AMT), a method that employs an initial qualitative LC-MS/MS analysis to identify peptides and their corresponding retention times, followed by a second LC-MS analysis allowing the quantification of those peptides using ion currents (Lipton *et al*., 2002). More recently, a data independent mode of acquisition termed MS^E, similar to the AMT approach, has been reported (Silva *et al*., 2006a). MS^E uses alternating scans of low and elevated collision energies to obtain ion intensities from both the precursor and the product ions of eluting peptides. This, in combination with reproducible retention times, allows their simultaneous quantitation and identification. An outline of this approach is shown in **Figure 5.1**. Further, the approach also allows the absolute quantification of proteins (Silva *et al*., 2006b). Here, the average intensity of the top three most abundant peptide ions of an internal standard is used to calculate a response factor. This in turn can be used to calculate the absolute amount of

each identified protein by comparison to the mean intensity of the top three most abundant peptide ions from each respective identified protein.

5.2 Quantitative proteomic studies of *C. trachomatis*

There are a limited number of quantitative proteomic studies of *Chlamydia* and even fewer of the LGV biovars (see Proteomics summary **Table 1.3** in Chapter 1). Moreover, these studies can only be considered semi-quantitative at best and the number of proteins assigned within these studies is low. Lundemose *et al*. (1990) characterized the synthesis of early proteins during the intracellular transition from EB to RB using 2-DGE in combination with pulse-labeling experiments, and were able to show the synthesis of seven proteins during the first 8 h after infection before the detection of MOMP at 10 h. Three of these proteins were identified as S1 ribosomal protein, GroEL and DnaK using immunoblotting and were shown to decline during 26-30 h PI, the period when RB are undergoing the second stage of differentiation to the infectious EB form. The same research group also reported the differential expression of several unidentified proteins when investigating the chlamydial response of interferon gamma on *C. trachomatis* serovars A and L2 (Shaw *et al*.,1999).

A more extensive quantitative proteomic analysis of the related human respiratory pathogen *Chlamydophila pneumonia* investigated the global expression changes during the re-differentiation from RB to EB (Mukhopadhyay *et al*., 2006). Although there are distinct differences in the clinical manifestation between the *C. trachomatis* and *C. pneumonia* species, similarities in their biphasic developmental cycle and morphology indicate that there should be core similarities at the molecular level. This study identified 35 proteins whose expression levels were altered during the transition from RB to EB.

At the time of undertaking this project, all quantitative studies of *Chlamydia* have employed 2-DGE in combination with pulse-labeling and pulse-chase experiments. Although such approaches are attractive because they allow the detection of chlamydial proteins in the apparent absence of contaminating host proteins, the technique is restricted to the detection of newly synthesized proteins and does not provide a direct quantitative measure of already existing proteins. If comparisons are to be made between the transcriptome and proteome, then it is critical that alternative approaches are implemented that not only overcome this limitation, but also those limitations associated with 2-DGE. Further, relative quantitation strategies such as 2-DGE, although useful for the comparison of one or more different experimental conditions, only offer information on the direction of change (-up or down-regulation), with amounts expressed as ‘fold’ change. By contrast, absolute quantitation strategies determine the amount of a peptide or protein in terms of their precise molar concentration (e.g., fmol per ml of serum, ng per gram of tissue). As such, the combination of samples that can be compared using absolute quantitation strategies is almost limitless and can equally be used to calculate relative measures of protein expression between samples. However, costly reagents, time-consuming assay development and the limited number of proteins that can be quantified per experiment. Nonetheless, for the integration of proteomic data with other ‘omic’ datasets in the context of systems biology, the quantitation of proteins in absolute amounts is an essential prerequisite.

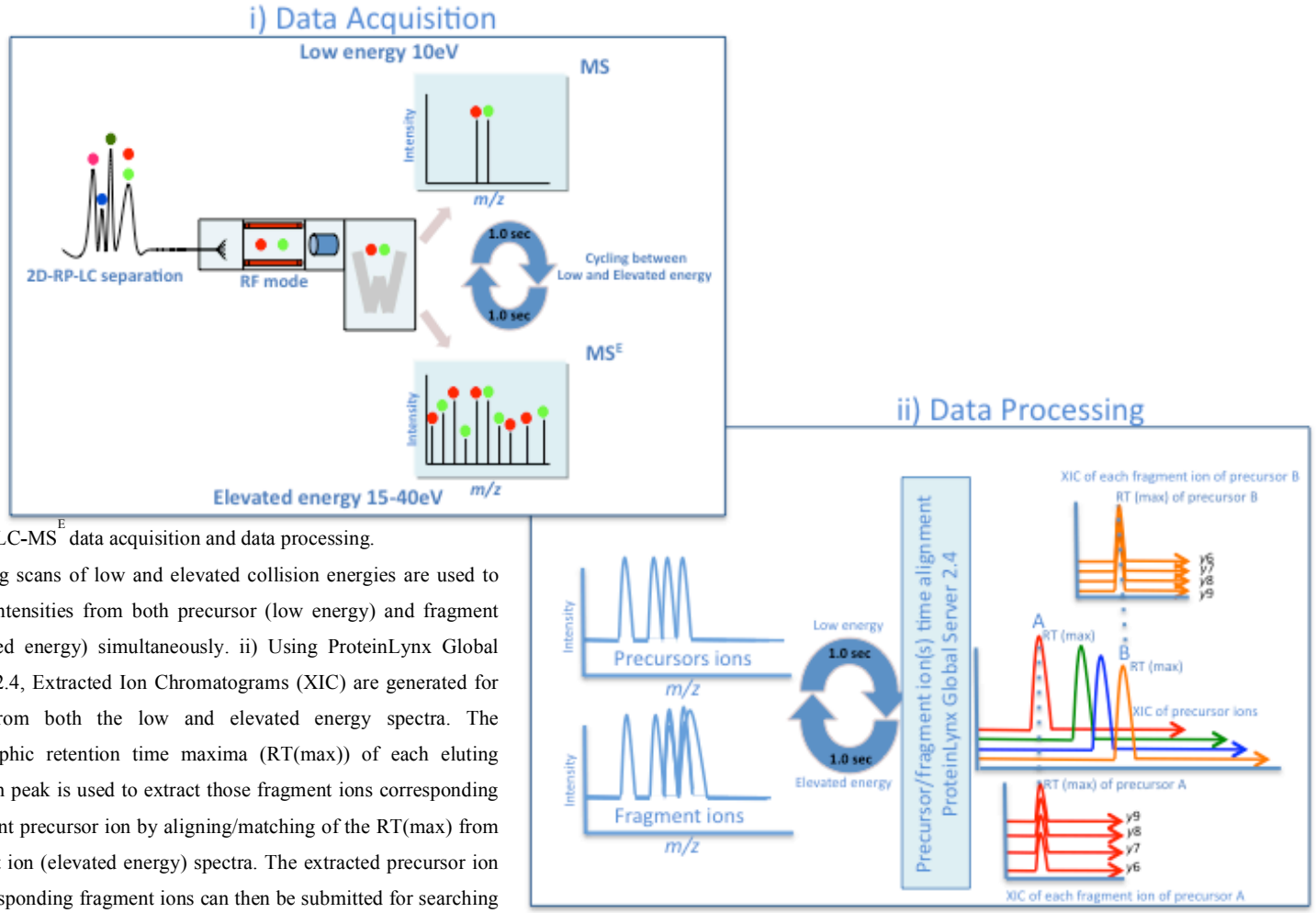


Figure 5.1. LC-MS^{E} data acquisition and data processing.

i) Alternating scans of low and elevated collision energies are used to obtain ion intensities from both precursor (low energy) and fragment ions (elevated energy) simultaneously. ii) Using ProteinLynx Global Server Ver 2.4, Extracted Ion Chromatograms (XIC) are generated for each ion from both the low and elevated energy spectra. The chromatographic retention time maxima (RT(max)) of each eluting precursor ion peak is used to extract those fragment ions corresponding to the relevant precursor ion by aligning/matching of the RT(max) from the fragment ion (elevated energy) spectra. The extracted precursor ion and its corresponding fragment ions can then be submitted for searching

The work in this Chapter describes a quantitative proteomic analysis of EBs and RBs from *C. trachomatis* L2 using two different proteomic approaches to address many of these limitations. Qualitative and quantitative analysis has been performed using: i) MudPIT incorporating iTRAQ tags and ii) label-free analysis using two-dimensional reverse phase chromatography in combination with MS^E with protein expression reported in both relative and absolute amounts.

5.3 Experimental design

Culture of *C. trachomatis L2 and purification of EBs and RBs*

RBs and EBs were cultured and purified in a similar manner to those prepared for the qualitative studies presented in Chapter 4. Four batches of Buffalo Green Monkey Kidney (BGMK) cells were infected with *C. trachomatis* L2/434/Bu as described in *section 2.2.6*. Infected monolayers were harvested at both 15 h and 48 h. RBs were purified at 15 h PI and EBs at 48 h PI by density gradient centrifugation (*section 2.2.7*). To minimize potential errors arising from differences in protein concentration between EBs and RBs, and/or low-level host cell protein contamination, the data was normalized against the number of bacteria determined for each preparation as described in *section 2.8.3.1* (where, 1 genome is equivalent to 1 bacterium). The number of bacteria determined in a typical preparation of purified EBs and RBs was 1.17×10^{12} and 1.50×10^{12} genomes per ml respectively (as calculated by qPCR, *section 2.2.10)*.

Preparation of EB and RB protein lysates

Purified EB and RB preparations (~1-2 mg of cells) were re-suspended in 0.5 M triethylammonium bicarbonate containing 5 mM dithiothreitol (final concentration) and incubated for 2 h on ice. Addition of the reducing agent DTT was found to be necessary for efficient lysis of EBs, but was not required for the osmotically fragile RBs. This requirement for DTT is likely to provide the necessary reduction of the disulphide bridges within the highly cross-linked outer membrane complex of EB, disrupting the outer membrane and facilitating cell breakage. DTT treated cell suspensions were lysed using a combination of ceramic bead maceration and sonication as described in *section 2.3.9.1*.

5.3.1 iTRAQ analysis

Tryptic digests of EBs and RBs were labeled using 4-plex iTRAQ reagents as described in *section 2.3.9.3*. Biological replicates of EB and RB peptides were labeled as illustrated in **Figure 5.2** and sample loading normalized based upon their protein concentration (100 μg per replicate). The combined iTRAQ labeled sample was fractionated using SCX (*section 2.3.9.4*) and each fraction separated and analysed using nanoLC-MS/MS (*section 2.3.9.5*). The chromatogram of the SCX fractionation is shown in **Figure 5.3.**

LC-MS/MS data generated from the analysis of 41 SCX fractions were processed and searched against a protein translation of the *C. trachomatis* L2/434/Bu genome and pL2 plasmid sequence using the Mascot software suite as described in *section 2.7*.2 and *2.8.2.* Identification of peptides, extraction of intensity information from the iTRAQ reporter ions and peptide normalization were performed using Mascot

ver 2.2. Based upon criteria set in Mascot a protein was reported if identified by ≥ 2 unique peptides and achieved an identity threshold of $p<0.03$ (≥98% confidence). To provide an estimate of the false discovery rate (FDR) of the experimental dataset, the same MS/MS dataset was also searched against a randomized database of the *C. trachomatis* L2 translation. The False Discovery Rate of peptides above the identity threshold was estimated to be 3.23%. For the quantification of identified proteins, peptide reporter ion intensities were summed and a weighted protein ratio calculated for each identified protein. Protein ratios were further normalized according to the number of genomes (1 genome representing 1 bacterium) contained within each labelled iTRAQ sample (*section 2.8.3.2*). The mean protein ratio and standard deviation for each protein across the two biological replicates were calculated. Where a ratio was only available for one of the two biological replicates, it was excluded from the quantitative dataset. To complement the qualitative study in Chapter 4, peptides with two unique peptides, but without quantitative information were also reported (**supplementary data Table 5.1**).

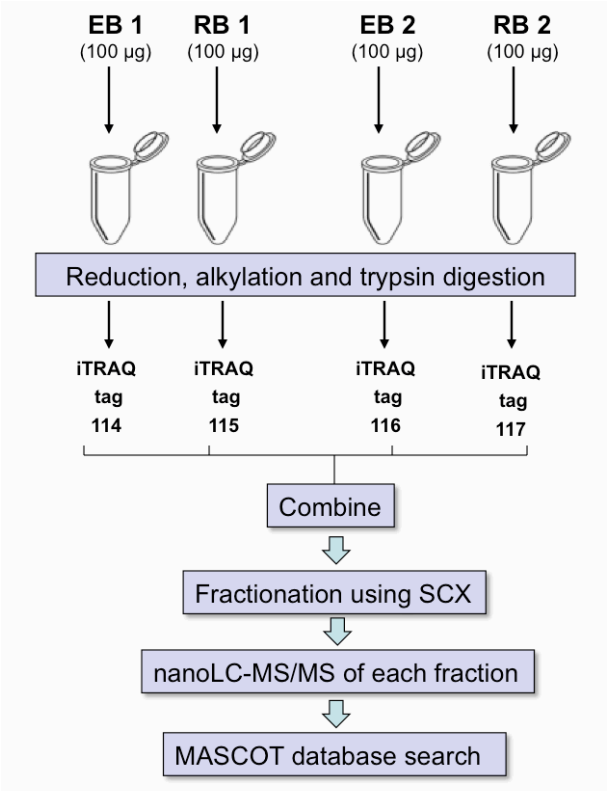


Figure 5.2. Outline workflow for the preparation of EB and RB protein lysates for iTRAQ analysis. 1 and 2 refer to biological replicates of each sample.

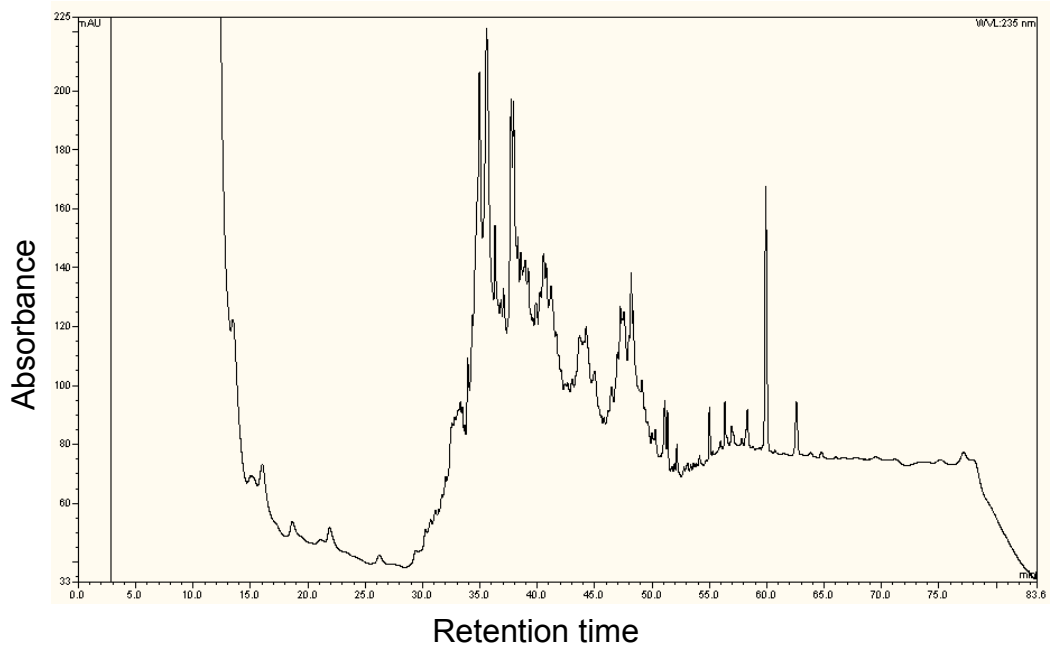


Figure 5.3. SCX fractionation of iTRAQ labeled peptides from *C. trachomatis* L2. UV absorbance was monitored at 235 nm.

5.3.2 2D-RP-UPLC -MS^E (label-free analysis)

Duplicate tryptic digests of two biological preparations (9 µg per replicate) of both EB and RB protein lysates (the same preparations used for the iTRAQ analysis) were separated using an 11 step RP gradient of increasing percentages of acetonitrile buffered at pH 10 using ammonium formate. Each of these 11 percentage cuts were separated in turn using a conventional linear RP gradient at pH 2.0 and analysed online using a Synapt Q-Tof mass spectrometer as described in *section 2.6.4*. **Figure 5.4** provides an outline for the label-free experimental workflow.

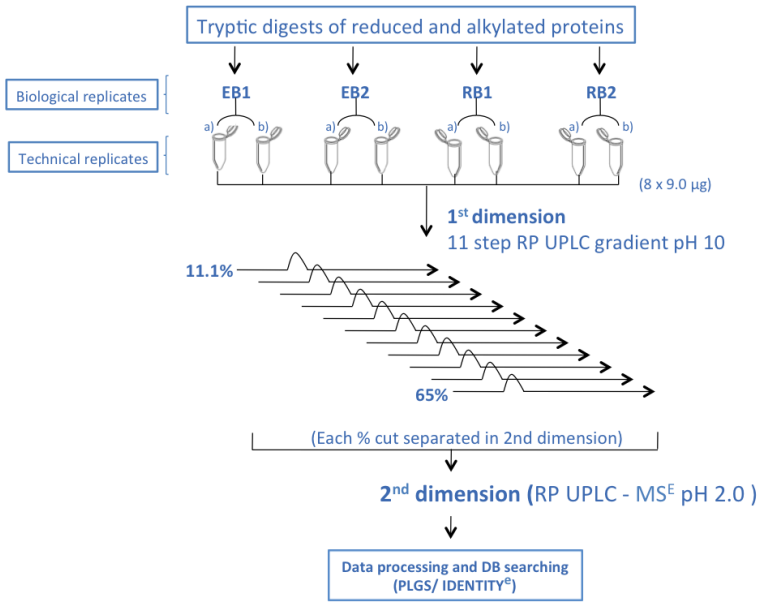


Figure 5.4. Experimental workflow for the analysis of *C. trachomatis* L2 using 2D-RP-RP-LC-MS^E label-free technology.

Data was processed and normalized using the software IDENTITY^E, part of the ProteinLynx Global Server package (*sections 2.7.3 and 2.8.3*). The database-searching algorithm IDENTITY^E uses an iterative process, where several stages of filtering and scoring are applied for protein assignment. Each iterative step increases the specificity, selectivity and sensitivity of each search. There are three major steps within this search strategy. The '1st pass' database search uses physiochemical properties of peptides in the gas-liquid phase to rank peptides, using parameters such as summed product ion intensities; the correct number of product ions according to length, charge state and intensity model; and the presence of preferred fragmentation sites. For the '2nd pass' search, a database based upon the confidently assigned '1st pass' ranked peptides is generated. The remaining data is searched against this database, taking into account peptide modifications, missed cleavages and losses associated with in-source fragmentation. The combined results from the 1st pass and 2nd pass searches are combined. The remaining accurate mass precursor/product ion list is searched to identify multiple modifications using the same assignment criteria used for the '2nd pass' search. All three of these consecutive searches are combined to generate a ranked list of assigned proteins. A false positive discovery rate filter of 4% was applied to the dataset. It is also important to note

that only the intensity of the top three most intense peptide ions from the '1st pass' search are used for the quantitative measurement.

5.5 Results

5.5.1 Protein identifications

iTRAQ analyses of EBs and RBs identified 4534 peptides from 85963 MS/MS spectra. There was a total of 335 non-redundant *C. trachomatis* proteins based upon a minimum of two peptides per protein, of which 169 revealed quantitative data.

In contrast, the label-free analysis assigned a total of 84,877 different chlamydial peptides with 45,465 peptides identified in EBs and 60,641 peptides in RBs. The total number of non-redundant *C. trachomatis* proteins identified using label-free acquisition from both EBs and RBs by ≥ 3 peptides was 580, of which 573 were assigned to the *C. trachomatis* L2 genome and 7 to the genome of the L2 plasmid. After the application of filtering criteria, where a protein must have been assigned in at least two replicates from the same developmental form, quantitative data was obtained for 489 proteins.

To account for possible contaminating homologous host cell peptides, the label-free dataset was also searched against a protein translation of the human genome. A total of 459 assigned chlamydial peptides were found to have human homologs. Of these, a total of 27 were assigned as '1st pass' and 432 as '2nd pass'. Interestingly, the mean peptide length of the 432 homologous '2nd pass' peptides was 3.4 amino acids, whereas the mean length of the remaining 76,015 chlamydial '2nd pass' peptides was 14.7. These short homologous peptides are likely to reflect the increased probability of matching shorter peptides between the two databases. None of the 27 human '1st pass' peptides overlapped with the 'top3' peptides used for quantitation. All overlapping peptides were removed from the chlamydial dataset. (The correlation and removal of homologous peptides between the chlamydial and human datasets was performed using a python script written by Dr. Richard Edwards, Southampton University, UK).

Qualitative and quantitative information for both experimental approaches are summarised in **Table 5.1**. It is important to note, that when comparing these two approaches, that the minimum number of peptides used to assign each protein differ, with a filter of ≥ 2 peptides for iTRAQ and ≥ 3 peptides using the label free approach. The validity of assigning proteins based upon single peptide or so-called 'one hit wonders' in the absence of manually validated spectra has been brought into question (Veenstra *et al*., 2004). In consideration of the large sizes of both datasets, the manual validation of individual spectra was deemed unfeasible. Since there is an increased potential for false-positive identifications when including single peptides, the criteria for the assignment of a protein in the iTRAQ dataset was established as a minimum of two peptides. However, the assignment criteria for the label-free dataset was established as ≥ 3 peptides. This threshold arises from the default quantification requirement, where the intensity of the top three most abundant different peptide ions from each identified protein are required to calculate the concentration of the protein. Proteins identified by iTRAQ and the label-free approach are listed in **appendix I**, **Tables 5.1** and

5.2 respectively. These tables also provide information on the molecular weight, p*I* and the number of peptides used to assign each protein.

Table 5.1. Total protein identifications and quantitative data obtained from both iTRAQ and label-free analysis.

	iTRAQ	**Label-free**
Total number of protein identifications	335	580
Number of assignments with quantitative data	169	489

When comparing the number of peptides used to assign each protein using the two different experimental approaches, clear differences arise. **Figure 5.5** shows the distribution of the number of peptides used to assign a protein using the iTRAQ and label-free approach. The number of peptides in the iTRAQ dataset ranged from 2 to 53 peptides used to assign a peptide with an average of 10 peptides (median =7 peptides) compared to the label-free approach, which ranged from a minimum of 3 to 928 unique peptides per protein assignment with an average of 143 peptides (median = 102 peptides). Thus, the label-free approach resulted in consistently higher protein sequence coverage's. For example, sequence coverage obtained of the 60 kDa cysteine rich protein (OmcB) using the label-free approach was an average of 67% supported by 380 different peptides. In contrast the iTRAQ approach yielded 49% sequence coverage supported by 28 peptides. The higher numbers of peptides and improved sequence coverage obtained through the label-free method significantly increased the confidence in protein identifications.

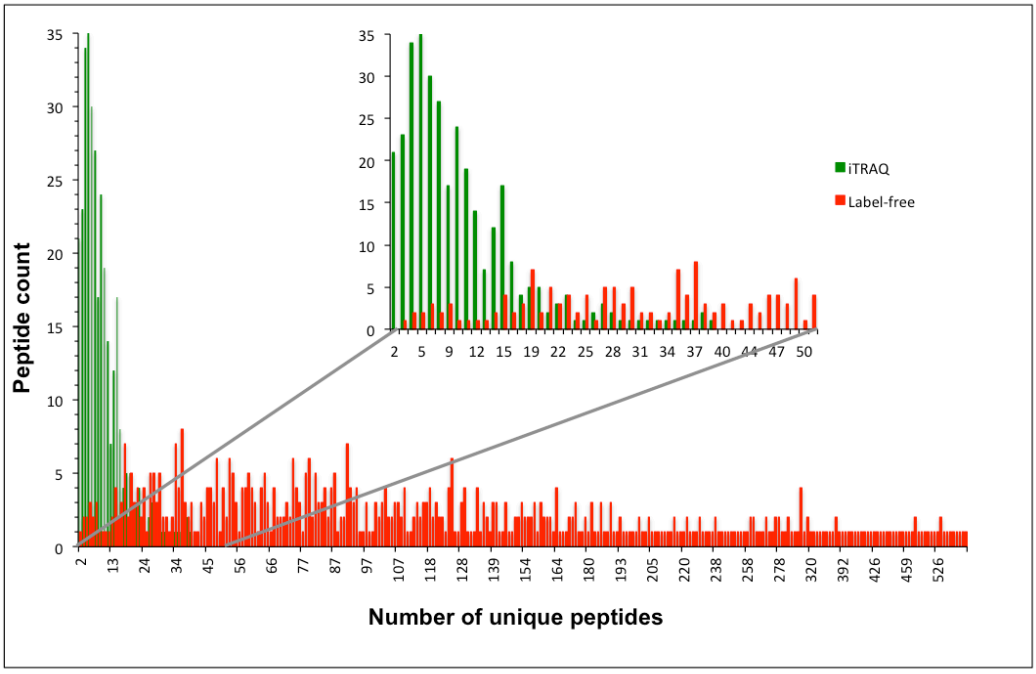


Figure 5.5. The numbers of unique peptides available to assign proteins using the label-free and iTRAQ approaches.

The overlap of protein identifications obtained using the label-free and iTRAQ methods is shown in **Figure 5.6**. Of the 905 predicted genes encoded by the *C. trachomatis* L2 genome and its associated cryptic plasmid, 630 proteins were identified using a combination of the data obtained from the two different approaches, representing ~70% of the theoretical *C. trachomatis* L2 proteome. A significant number of these proteins (~33%) were uniquely identified using the label-free method, while only 6% of the proteins were unique to the iTRAQ approach. The remaining 31% were identified using both technologies.

For comparative purposes, **Figure 5.7** shows the overlap between label-free and iTRAQ identified proteins, but also includes those proteins identified from *C. trachomatis* L2 using the qualitative approaches GeLC-MS/MS and MudPIT previously presented in Chapter 4. Combination of these four datasets realised only an additional 2.0% increase in proteome coverage. Again, in comparison, the label-free method uniquely identified a significantly greater proportion of proteins representing ~29% of the 648 proteins identified, while iTRAQ, GeLC-MS/MS and MudPIT uniquely identified 6%, 2.3% and <1% respectively. The label-free also had a significantly higher proportion of proteins in common with the three alternative technologies. MudPIT and iTRAQ also had a considerable number of overlapping identified proteins between them. Only eleven percent of the proteins were identified using all four technologies. Despite assignment of ~72% of the *C. trachomatis* proteome, ~28% of the proteins encoded by the *C. trachomatis* genome remain to be detected. The possible reasons why these proteins have not been detected using the applied technologies are discussed in Chapter 6.

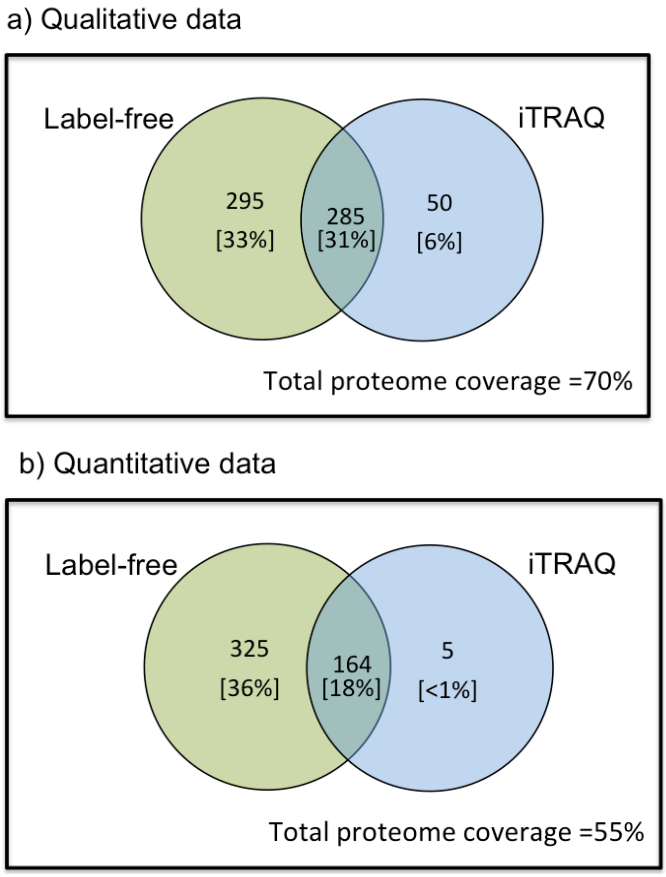


Figure 5.6. Venn diagram illustrating the overlap of proteins assigned using iTRAQ and the label-free approaches. Panel a) represents the qualitative data and panel b) represents the quantitative data.

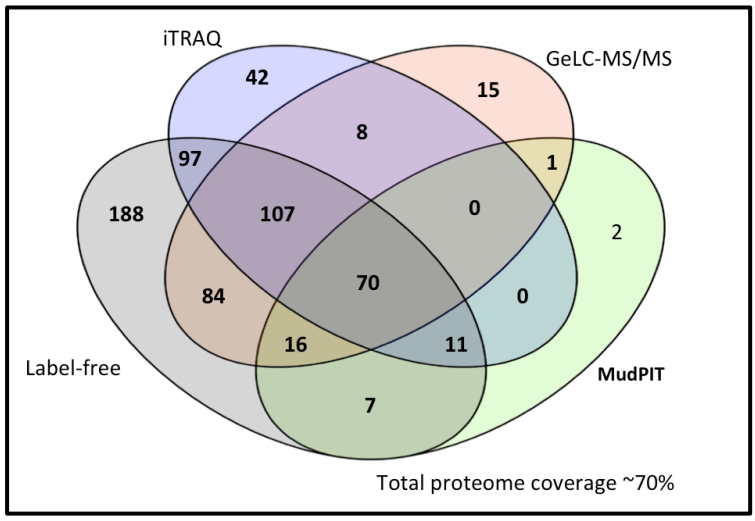


Figure 5.7. Venn diagram illustrating the qualitative overlap of assigned proteins using MudPIT, GeLC-MS/MS, iTRAQ and Label-free approaches.

5.5.2 Quantitation of identified proteins

5.5.2.1 iTRAQ data

Figure 5.8 shows the differential expression profile of proteins from EBs and RBs of *C. trachomatis* L2 as characterised by iTRAQ. This scaled plot of the mean iTRAQ protein ratio plotted against Protein number, highlights the limited range of expression from 0.21 to 1.9. The correlation of the protein ratios across the two biological replicates were in good agreement with an R^2 value of 0.81 as shown in **Figure 5.9**.

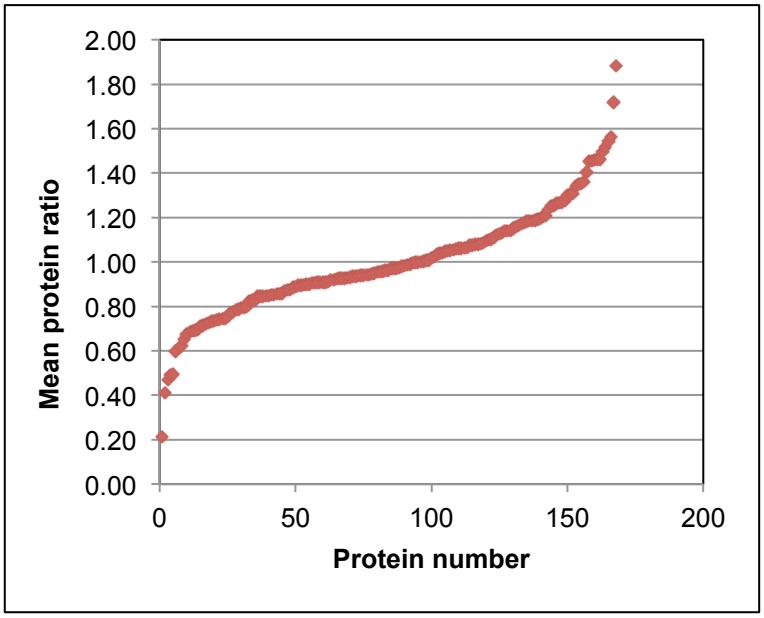


Figure 5.8. The combined iTRAQ protein expression profile of RBs (15 h PI) and EBs (48 h PI) from *C. trachomatis* L2. Protein number is defined as the rank of the protein sorted according to expression ratio.

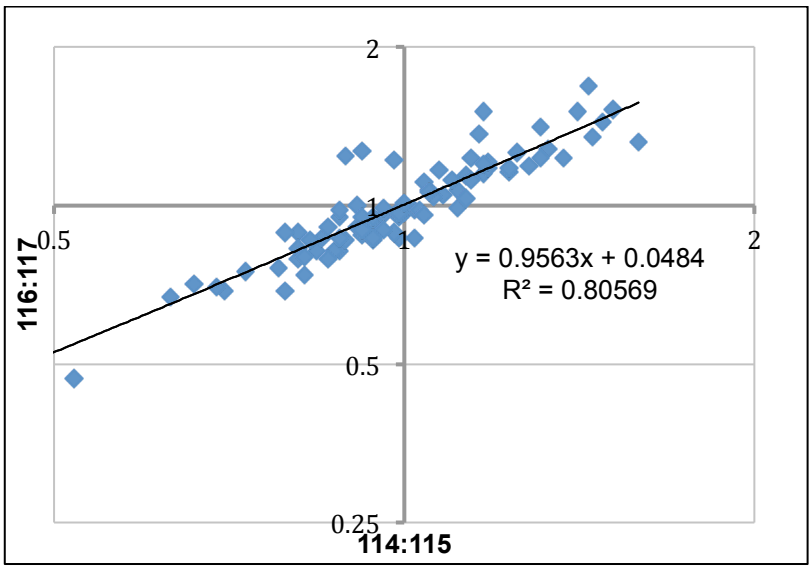


Figure 5.9. Correlation of iTRAQ protein ratios from two biological replicates as measured by the abundance of the reporter tags at *m/z* 114-117. 114:115 = EB:RB protein ratio from biological replicate 1, 116:117 = EB:RB protein ratio from biological replicate 2.

To illustrate a typical peptide identification and relative quantitation result obtained using iTRAQ, **Figure 5.10** depicts a fragmented 4-plex iTRAQ labeled peptide from the 60kDa cysteine-rich outer membrane protein (OmcB). A total of 28 peptides were identified for OmcB representing 49% protein coverage.

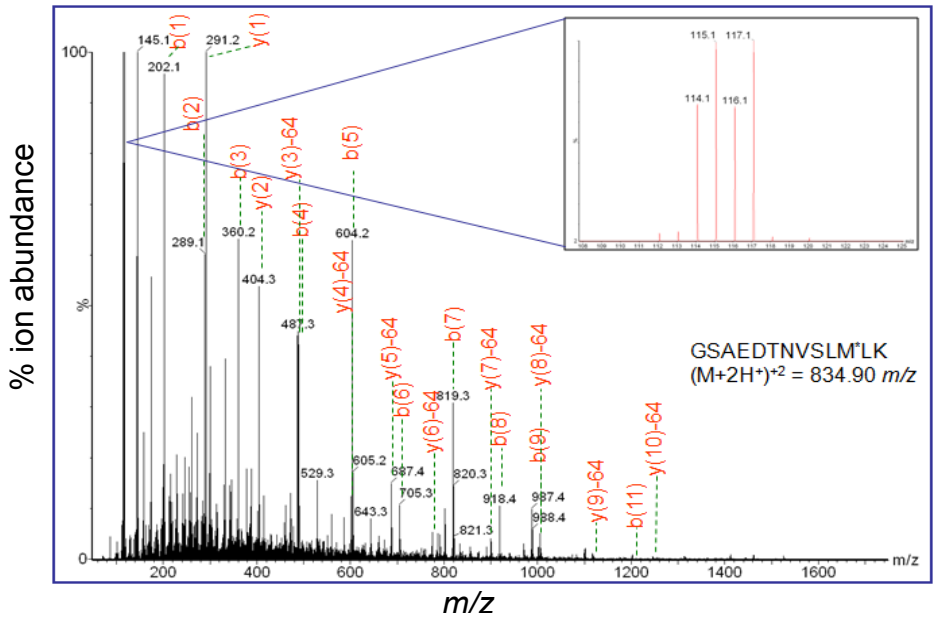


Figure 5.10. A representative mass spectrum from the iTRAQ labelled peptide GSAEDTNVSLM*LK, $(M+2H^{+})^{+2}$ = 834.90 *m/z,* identified as originating from OmcB. The iTRAQ reporter ion intensities in the region from 114 – 117 *m/z* are shown in the inset. * indicates an oxidised methionine. This is indicated in the mass spectrum by the loss of CH_3SOH (-64 amu) indicating the presence of methionine sulphoxide.

iTRAQ is a multiplex technology, i.e. multiple samples are combined within the same experiment, and quantification is dependent upon the presence of a reporter ion. Consequently, if a peptide reporter ion is absent for one of the replicates within the experiment, a protein ratio cannot be calculated. Even when reporter ion data is available for one out of the two biological replicates, this is excluded under the quantification filtering criteria (i.e. datapoints must be present in both biological replicates). Using these criteria, quantitative data were obtained for 169 out of the 335 proteins identified using iTRAQ (**Table 5.2**). These quantitative measurements can be found in **Appendix I**, **Table 5.1**, in addition to those proteins only fulfilling the identification criteria. Because of the relatively subtle nature of the expression changes shown by this experimental dataset, a threshold of 1.4 fold is the cut-off filter for reporting proteins as differentially expressed. This cut-off value was based upon previous iTRAQ studies performed in our laboratory where technical variation was consistently below 30%. This is also consistent with other published studies (Gan *et al*., 2007; Chen *et al*., 2007; Choong *et al*., 2010). A list of these proteins and their expression changes that were equal to or greater than this cut-off threshold for both up- and down-regulation can be found in **Tables 5.2** and **5.3**, respectively.

Table 5.2. Proteins up-regulated in *C. trachomatis* L2 EBs using iTRAQ

Protein name	Gene name	Primary locus	UniProt accession number	p*I*[a)]	Mass (kDa)[a)]	Mean protein ratio	STD dev[b)]	Fold-change
Integration Host Factor Alpha	*ihfA*	CTL0519	B0B7I3	11.0	11.4	1.88	0.02	1.9
hypothetical protein		CTL0322	B0B9H4	9.9	17.9	1.72	0.10	1.7
Transcription antitermination factor	*nusA*	CTL0352	B0B9K5	5.1	48.9	1.56	0.18	1.6
Glucosamine-Fructose-6-P Aminotransferase	*glmS*	CTL0188	B0B945	5.3	67.4	1.54	0.03	1.5
tyrosyl tRNA Synthetase	*tyrS*	CTL0318	B0B9H0	6.6	45.4	1.52	0.00	1.5
hypothetical protein		CTL0036	B0B8P3	4.8	16.5	1.50	0.07	1.5
RECA Protein	*recA*	CTL0018	B0B8M5	7.0	37.8	1.46	0.02	1.5
Elongation Factor TS	*tsf*	CTL0048	B0B8Q5	5.6	30.9	1.46	0.07	1.5
hypothetical protein		CTL0060	B0B8R7	4.9	25.2	1.46	0.11	1.5
oligopeptide Binding Lipoprotein	*oppA4*	CTL0741	B0B853	5.0	77.5	1.46	0.03	1.5
L13 Ribosomal Protein	*rplM*	CTL0380	B0B9N3	10.1	16.8	1.45	0.19	1.5
Thioredoxin Disulfide Isomerase		CTL0149	B0B906	7.9	16.2	1.40	0.07	1.4
S10 Ribosomal Protein	*rpsJ*	CTL0695	B0B808	10.4	11.9	1.36	0.07	1.4
Predicted 1,6-Fructose biphosphate aldolase (dehydrin family)	*dhnA*	CTL0467	B0B9W5	6.3	38.0	1.35	0.03	1.4

a) p*I* and molecular mass were calculated using Mascot server Ver2.2.
b) Calculated standard deviation

Table 5.3. Proteins down-regulated in *C. trachomatis* L2 EBs analysed using iTRAQ

Protein name	Gene name	Primary locus	UniProt accession number	[a)]pI	Mass (kDa)[a)]	Mean protein ratio	STD dev[b)]	Fold-change
Solute Protein Binding Family		CTL0323	B0B9H5	5.0	33.3	0.74	0.02	-1.4
Peptidoglycan-Associated Lipoprotein	*pal*	CTL0863	B0B8H4	7.9	19.0	0.74	0.02	-1.4
Acyl-Carrier UDP-GlcNAc O-Acyltransferase	*lpxA*	CTL0793	B0B8A5	6.0	30.7	0.73	0.02	-1.4
Lon ATP-dependent protease	*lon*	CTL0598	B0B7R3	6.9	91.9	0.73	0.01	-1.4
Putative outer membrane protein B	*pmpB*	CTL0670	B0B7Y3	8.2	183.0	0.72	0.01	-1.4
hypothetical protein		CTL0271	B0B9C3	4.7	26.7	0.71	0.04	-1.4
probable Yop proteins translocation protein C		CTL0043	B0B8Q0	5.4	95.7	0.71	0.01	-1.4
CLP Protease	*clpP*	CTL0690	B0B803	5.2	20.9	0.69	0.01	-1.4
HSP-60	*hsp60*	CTL0365	B0B9L8	5.2	58.1	0.69	0.01	-1.4
Acyl Carrier Protein	*acpP*	CTL0488	B0B7F2	3.8	8.7	0.69	0.03	-1.5
HSP-70 Cofactor	*grpE*	CTL0651	B0B7W5	4.6	21.7	0.69	0.04	-1.5
Candidate inclusion membrane protein		CTL0476	B0B9X4	6.7	29.6	0.67	0.00	-1.5
hypothetical protein		CTL0238	B0B995	5.2	53.6	0.65	0.03	-1.5
Polymorphic outer membrane protein		CTL0255	B0B9A7	5.3	65.8	0.62	0.02	-1.6
hypothetical protein		CTL0272	B0B9C4	6.6	47.7	0.61	0.02	-1.6
ATP Synthase Subunit E	*atpE*	CTL0562	B0B7M6	5.4	22.9	0.60	0.01	-1.7
CHLPN 76kDa Homolog		CTL0886	B0B8J7	4.9	68.9	0.49	0.03	-2.0
hypothetical protein		CTL0540	B0B7K4	8.3	63.5	0.49	0.00	-2.0
60kDa Cysteine-Rich OMP	*omcB*	CTL0702	B0B815	8.0	56.4	0.47	0.02	-2.1
Histone-Like Developmental Protein	*hctA*	CTL0112	B0B8W9	10.6	13.7	0.41	0.01	-2.4
hypothetical protein		CTL0840	B0B8F1	6.5	13.3	0.21	0.03	-4.7

a) p*I* and molecular mass were calculated using Mascot server Ver2.2.
b) Calculated standard deviation.

5.5.2.2 MS^E label-free data

In the label-free system, each sample was analysed separately using an 11 step 2D-RP-RP-UPLC separation in conjunction with the data-independent MS acquisition MS^E, as described briefly above and further in *section 2.6.4*. Eight samples were analysed in total; two biological replicates of EBs, two biological replicates of RBs, all performed in duplicate to provide technical replicates. The excellent technical reproducibility afforded by this approach is highlighted in the scatter plot shown in **Figure 5.11**. Comparison of absolute protein measurements between two technical replicates of an EB protein lysate revealed a close correlation with an R^2 value of 0.96 with mean % CVs between replicates ranging from 12.9 to 16.9 across the dataset. However, there was significant variation of measurements observed between biological replicates with an average coefficient of variation of 45% for RBs and 29% for EBs.

One of the key differences between the two quantitative approaches presented in this thesis is that the label-free system provides absolute quantification in addition to relative quantification (Silva *et al*., 2006b). As described in *section 2.6.4,* the protein alcohol dehydrogenase was spiked into each sample as an internal standard. The average intensity of the top three most abundant peptide ions from this internal standard was used to calculate a response factor. By comparison of this response factor to the average of the top three most

abundant peptide ions obtained from each identified protein, the absolute amounts loaded onto the column in fmol and ng were calculated. Further, the number of cells contained within each preparation was also calculated (see *section 2.8.3.1*), allowing these absolute amounts to be expressed as molecules per cell. **Table 5.4** lists the top 15 most abundant proteins detected in *C. trachomatis* L2 RBs and EBs expressed as absolute amounts ranging from 2728 to 154 molecules per cell.

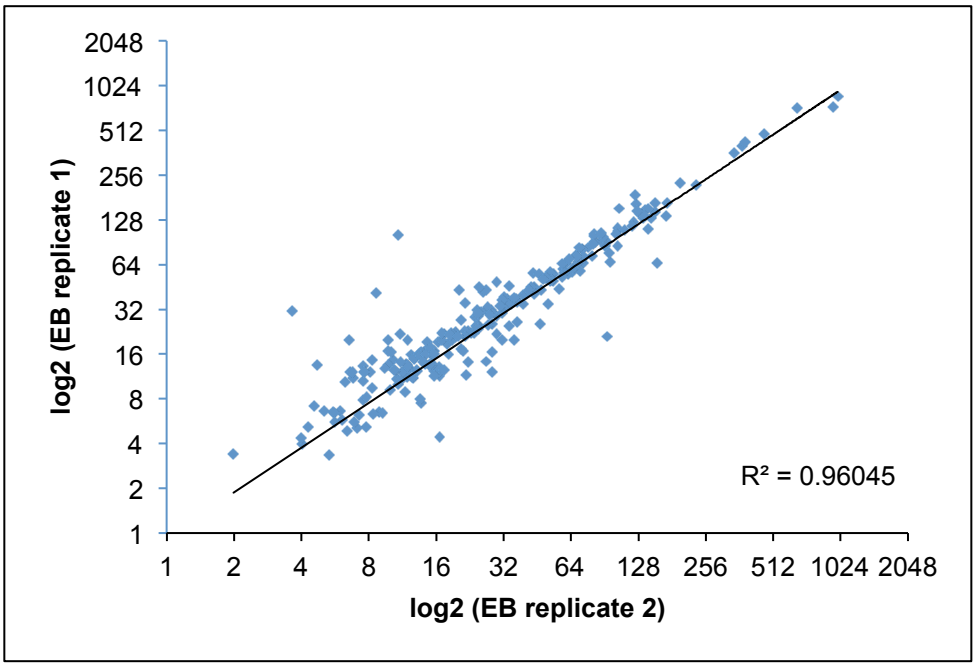


Figure 5.11. Comparison of the absolute protein abundance measurements obtained from the analysis of two technical replicates of EBs from *C. trachomatis* using 2D-LC-MS^{E}. A correlation coefficient of 0.96 was obtained.

Table 5.4. The top 15 most abundant proteins in RBs and EBs from *C. trachomatis* L2

UniProt accession	Locus	Gene name	Protein description	RB[a)]	EB[a)]	EB/RB ratio	Category
RB proteins							
B0B7N8	CTL0574	*tufA*	translation elongation factor Tu	2619	2156	0.82	Translation
B0B8Q7	CTL0050	*ompA*	major outer membrane protein	2041	2728	1.34	Cell envelope
B0B8B5	CTL0803	*mip*	peptidyl-prolyl cis-trans isomerase	1956	1292	0.66	Other categories
B0B9L8	CTL0365	*hsp60*	chaperonin GroEL	1871	1300	0.70	Translation
B0B815	CTL0702	*omcB*	60kD cysteine-rich outer membrane protein	1704	518	0.30	Cell Envelope
B0B8J8	CTL0887		putative exported protein	1464	840	0.57	Exported protein
B0B7W6	CTL0652	*dnaK*	chaperone protein	1424	1660	1.17	Translation
B0B9X4	CTL0476		candidate inclusion membrane protein	968	186	0.19	Cell envelope
B0B9H5	CTL0323		ABC transport protein_ solute binding component	950	379	0.40	Transport and binding proteins
B0B8P0	CTL0033		phosphopeptide binding protein (predicted to be a TTSS protein)	924	341	0.37	Cellular processes
B0B8F8	CTL0847		conserved hypothetical protein	914	1145	1.25	hypothetical protein
B0B9H3	CTL0321		ADP_ATP carrier protein	828	252	0.30	Energy metabolism
B0B952	CTL0195	*htrA*	serine protease	774	399	0.52	Translation
B0B940	CTL0183	*pmpD*	polymorphic outer membrane protein	685	361	0.53	Cell envelope
B0B9A3	CTL0250	*pmpG*	polymorphic outer membrane protein	637	154	0.24	Cell envelope
EB proteins							
B0B8Q7	CTL0050	*ompA*	major outer membrane protein	2041	2728	1.34	Cell envelope
B0B7N8	CTL0574	*tufA*	translation elongation factor Tu	2619	2156	0.82	Translation
B0B7W6	CTL0652	*dnaK*	chaperone protein	1424	1660	1.17	Translation
B0B9L8	CTL0365	*hsp60_1*	chaperonin GroEL	1871	1300	0.70	Translation
B0B8B5	CTL0803	*mip*	peptidyl-prolyl cis-trans isomerase	1956	1292	0.66	Other categories
B0B8F8	CTL0847		conserved hypothetical protein	914	1145	1.25	hypothetical protein
B0B7N2	CTL0568	*rplL*	LSU ribosomal protein L12P (L7/L12)	626	1006	1.61	Translation
B0B8J8	CTL0887		putative exported protein	1464	840	0.57	Exported protein
B0B8I5	CTL0874		conserved hypothetical protein	461	807	1.75	hypothetical protein
B0B7F2	CTL0488	*acpP*	acyl carrier protein	502	662	1.32	Fatty acid & phospholipid metabolism
B0B8Q5	CTL0048	*tsf*	translation elongation factor TS	453	584	1.29	Translation
B0B8G1	CTL0850	*eno*	enolase	525	556	1.06	Energy metabolism
B0B815	CTL0702	*omcB*	60kD cysteine-rich outer membrane protein	1704	518	0.30	Cell envelope
B0B9L9	CTL0366	*groES*	10 kDa chaperonin GroES	622	475	0.76	Translation
B0B7I3	CTL0519	*ihfA*	integration host factor alpha-subunit	204	472	2.32	DNA replication

a) expressed as molecules per cell

Since each sample using the label-free system is analysed in turn, each sample generates an individual set of protein identifications. After application of filtering criteria; where a protein must be present in EBs or RBs in at least two replicates, 48 identified proteins were found to be unique to RBs and 5 unique to EBs. A list of these proteins for RBs and EBs are shown in **Table 5.5**.

Table 5.5. Proteins identified as being unique to a) RBs and b) EBs of *C. trachomatis* L2 using the label-free approach.

UniProt accession	*Locus*	Gene name	Protein description	Average molecules/cell	Replication	Functional category
a) RB Proteins						
B0B8Y4	CTL0127	*murD*	UDP-N-acetylmuramoylalanine--D-glutamate ligase	19	3	Cell Envelope
B0B980	CTL0223		putative integral membrane protein	44	2	Cell Envelope
B0B9W4	CTL0466		candidate inclusion membrane protein	41	2	Cell Envelope
B0B7T4	CTL0619		putative integral membrane protein	49	3	Cell Envelope
B0B8H5	CTL0864		putative soluble transglycosylase	23	3	Cell Envelope
B0B884	CTL0772	*secY*	protein translocase subunit	31	2	Cellular Processes
B0B9J9	CTL0346	*sctU*	type III secretion system inner membrane protein	33	3	Cellular Processes
B0B8D6	CTL0825	*sctR*	type III secretion system_ membrane protein	32	3	Cellular Processes
B0B8E5	CTL0834	*gspE*	general secretion pathway protein E	34	2	Cellular Processes
B0B7P6	CTL0583	*xseA*	exodeoxyribonuclease VII large subunit	29	3	DNA Replication
B0B9C0	CTL0268	*cydA*	cytochrome d ubiquinol oxidase subunit I	117	3	Energy Metabolism
B0B7Q9	CTL0594		2-oxoisovalerate dehydrogenase alpha subunit	26	3	Energy Metabolism
B0B868	CTL0756		putative nucleotide transport protein	70	4	Energy Metabolism
B0B8V9	CTL0102		putative exported protein	33	2	Exported protein
B0B9I1	CTL0329		exported protein	14	2	Exported protein

B0B7M7	CTL0563		putative exported protein	41	2	Exported protein
B0B7S4	CTL0609		putative exported protein	22	4	Exported protein
B0B7V9	CTL0645		putative exported protein	72	2	Exported protein
B0B8E7	CTL0836		putative exported protein	38	2	Exported protein
B0B909	CTL0152		putative exported protein	55	4	Translation
B0B9V6	CTL0458		conserved hypothetical protein	1	2	Fatty Acid & Phospholipid Metabolism
B0B8U4	CTL0087		conserved hypothetical protein	14	2	hypothetical protein
B0B9Q0	CTL0397		conserved hypothetical protein	80	4	hypothetical protein
B0B7H1	CTL0507		conserved hypothetical protein	8	2	hypothetical protein
B0B7H9	CTL0515		conserved hypothetical protein	21	2	hypothetical protein
B0B7J9	CTL0535		conserved hypothetical protein	29	2	hypothetical protein
B0B9L7	CTL0364		conserved hypothetical protein	30	2	hypothetical protein
B0B9X3	CTL0475		candidate inclusion membrane protein	47	3	Other Categories
B0B7H4	CTL0510		putative cysteine desulfurase	17	3	Other Categories
B0B7U4	CTL0629		putative oxidoreductase	25	3	Other Categories
B0B7U9	CTL0635	*phnP*	metal-dependent hydrolase	23	2	Other Categories
B0B831	CTL0718		ribosomal-protein-alanine acetyltransferase	31	2	Other Categories
B0B835	CTL0722	*ispD*	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	25	4	Other Categories
B0B837	CTL0724		hydrolase_ haloacid dehalogenase-like family	68	3	Other Categories
B0B860	CTL0748		methyltransferase	19	3	Other Categories
BOBCM2	pL2-04		Putative uncharacterized protein	39	2	Plasmid
B0B7W4	CTL0650	*hrcA*	Putative transcriptional regulatory protein	63	4	Transcription
B0B7X3	CTL0660		23S rRNA methyltransferase	8	2	Transcription
B0B8T2	CTL0075	*clpP*	ATP-dependent Clp protease	58	3	Translation

B0B9U7	CTL0449	*gcp*	O-sialoglycoprotein endopeptidase	20	2	Translation
B0B848	CTL0736	*pheT*	phenylalanyl-tRNA synthetase beta chain	46	2	Translation
B0B8R6	CTL0059	*dppD*	ABC transport protein_ ATPase component	29	2	Transport and binding Proteins
B0B947	CTL0190		tyrosine-specific transport protein	52	2	Transport and binding Proteins
B0B986	CTL0231		sulfate transporter	47	2	Transport and binding Proteins
B0B9E1	CTL0289		putative membrane transport/efflux protein	15	3	Transport and binding Proteins
B0B7Y5	CTL0672		metal transporter_ metal-binding component	15	2	Transport and binding Proteins
B0B852	CTL0740	*oppB2*	oligopeptide transport system membrane permease	55	3	Transport and binding Proteins
B0B8H3	CTL0862	*tolB*	outer membrane component of membrane transport system	18	3	Transport and binding Proteins
b) EB Proteins						
B0B8J5	CTL0884		conserved hypothetical protein	25	2	hypothetical protein
B0B7Z4	CTL0681		conserved hypothetical protein	46	2	hypothetical protein
B0B9D4	CTL0282	*trmD*	tRNA (guanine-N(1)-)-methyltransferase	28	2	Transcription
B0B9C9	CTL0277	*rpmE*	LSU ribosomal protein L31P	23	2	Translation
B0B880	CTL0768	*rplQ*	LSU ribosomal protein L17P	57	4	Translation

After further filtering; where a protein was required to be present in at least 3 replicates, 436 proteins were found to be common to both EBs and RBs. The relative protein expression ratios for each of these proteins were calculated from their respective mean absolute quantity values in > 3 replicates. **Figure 5.12** shows the relative expression profile for these regulated proteins. A cut-off value of regulation was considered to be 30% fold-change, an average fold-change between -0.30 and 0.30 on a natural log scale (±1.3 fold-change), a value that is 2-3 times the estimated error on the intensity measurement (Vissers *et al*., 2007). A list of all proteins along with their quantitative information can be found in **Appendix 1, Table 5.2**.

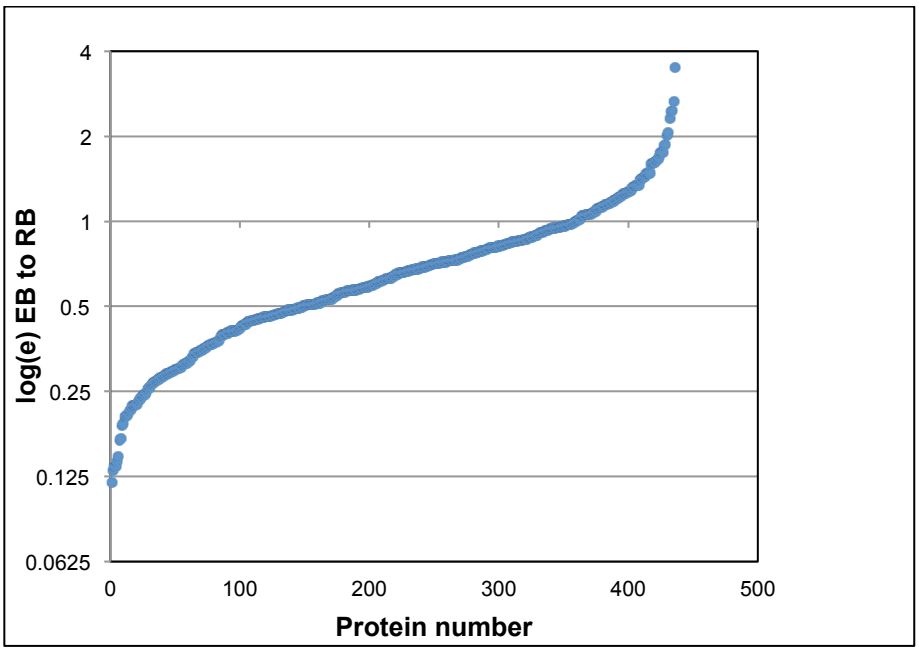


Figure 5.12. Label-free protein expression profile of RBs (15 h PI) and EBs (48 h PI) from *C. trachomatis* L2. A log(e) scaled plot of the mean protein ratios from EB to RB. Protein number is defined as the rank of the protein sorted according to expression ratio.

Figure 5.6 showed the distribution of quantitative data obtained using the two approaches. In contrast to iTRAQ where quantitative information was obtained for only 50% of the proteins assigned (169 out of the 335 proteins), the Label-free approach provided extensive proteome coverage, with quantitative data for ~84% of its identified proteins, representing 489 proteins.

A comparison of the expression ratios obtained for the overlapping quantitative data between the two different approaches showed a poor correlation with an R^2 value of only 0.060 (**Figure 5.13**). Interrogation of the raw data for the label-free approach showed good correlation between technical replicates as shown earlier in **Figure 5.11**. However, closer examination of the iTRAQ reporter ion ratios obtained for peptides assigned to a specific protein often showed disparity. **Figures 5.14** and **5.15** represent the distribution of peptide reporter ion ratios obtained for a selection of both the up- and down-regulated proteins. These graphs highlight the significant variation observed between iTRAQ peptide ratios used to calculate a quantitative value for a specific protein. The factors likely to contribute to these variations include; the possible post-translational modification of a peptide(s), co-fragmentation of peptides or the mis-assignment of a peptide to a protein by the Mascot algorithm. For abundant proteins where there are a large number of iTRAQ peptides assigned to a protein, there is a general correlation between the iTRAQ and label-free data, although the iTRAQ expression changes generally appear lower in magnitude.

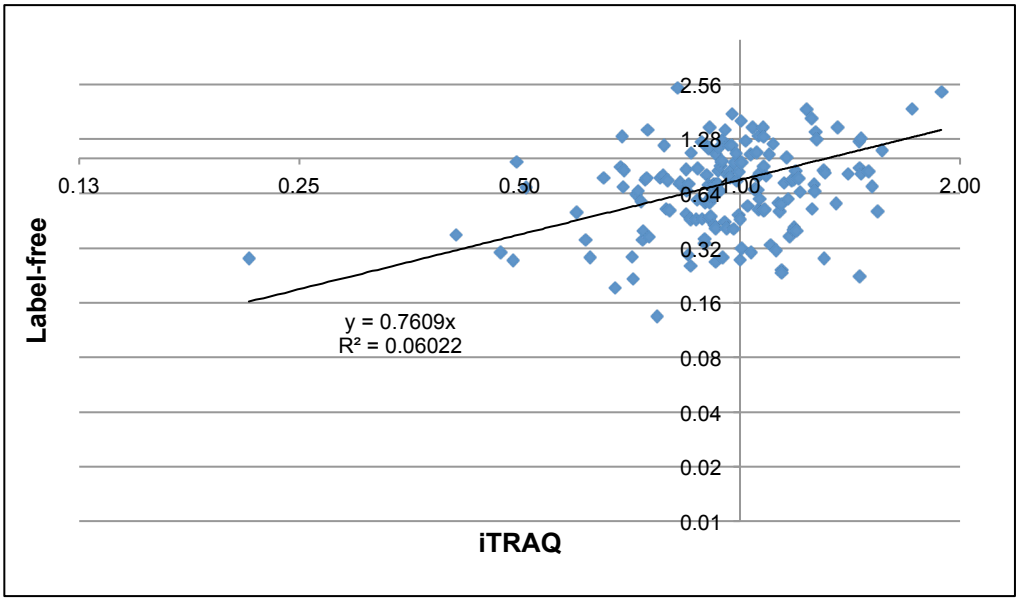


Figure 5.13. Correlation of protein expression ratios obtained from iTRAQ and the label-free method. Data has been transformed onto a log(e) scale.

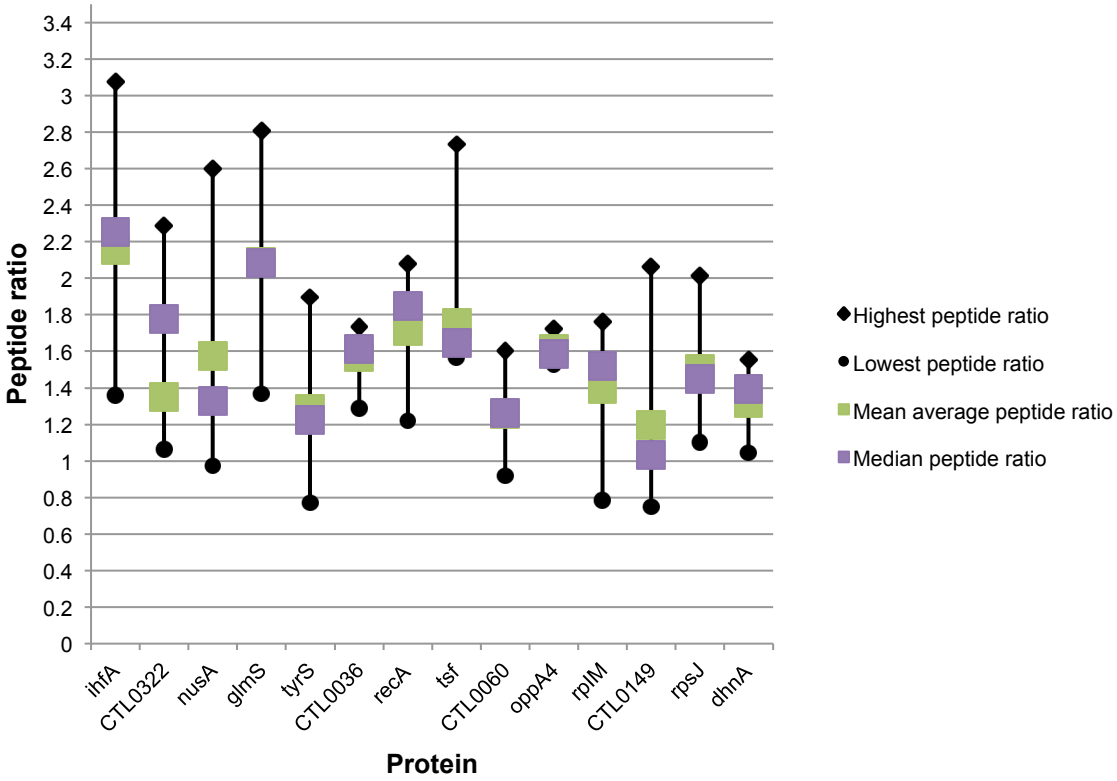


Figure 5.14. The distribution of iTRAQ peptide ratios obtained for proteins up-regulated in EBs when compared to RBs (**Table 5.2**). Ratios represented are a mean average of two biological replicates.

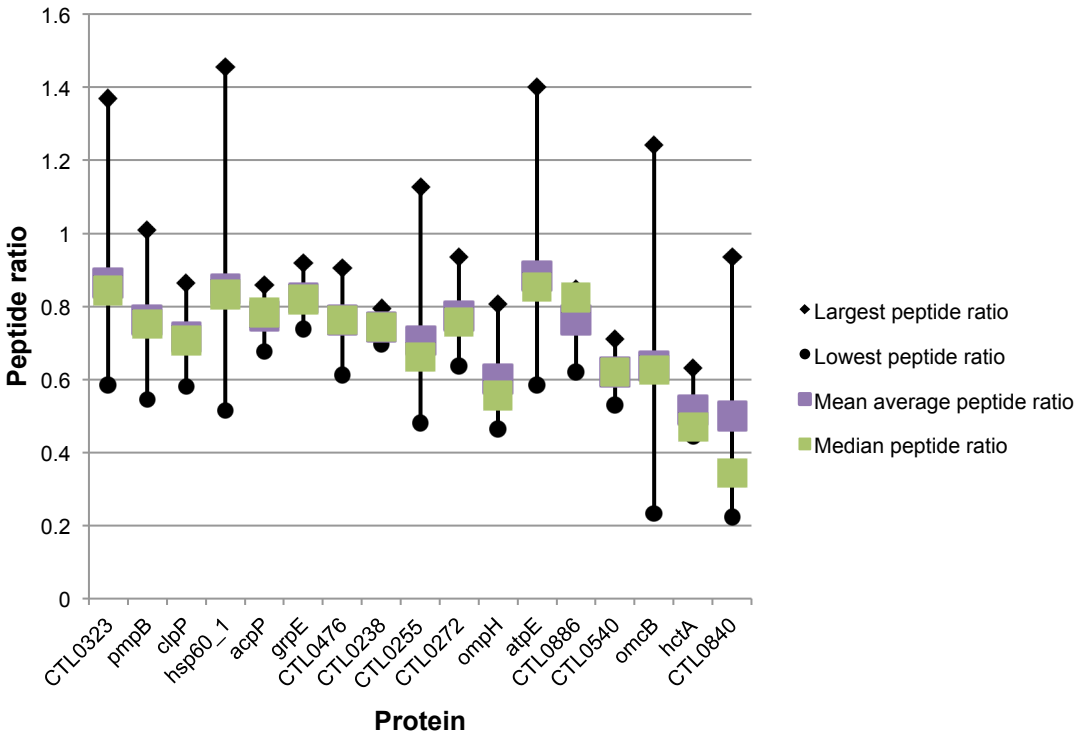


Figure 5.15. The distribution of iTRAQ peptide ratios obtained for proteins down-regulated in EBs when compared to RBs (see **Table 5.3**). Ratios represented are a mean average of two biological replicates.

5.5.3 *In silico* characterisation of proteins

Identified proteins from both iTRAQ and the label-free approach were categorized according to cellular function (**Figure 5.16**) as described in *section 2.8.4*. Further comparison of the functional distribution between these proteins assigned using label-free, iTRAQ and the qualitative techniques; MudPIT and GeLC-MS/MS presented in Chapter 4, are also shown in **Figure 5.17**. The correlation of predicted charge (p*I*) and molecular mass profiles of the predicted chlamydial proteome with the distribution profiles obtained from the experimentally assigned proteins for the iTRAQ data are shown in **Figure 5.18a** and **Figure 5.18b** respectively. The p*I* profile was typical of the bimodal arrangement observed for other bacterial proteomes, with the majority of proteins focused between p*I* 4-7 and p*I* 9-11 (Schwartz *et al*., 2001) (similar profiles were also obtained for the label-free approach, data not shown). These profiles and the highly similar distribution of proteins across the functional categories between technologies, including those presented in Chapter 4, support the idea that peptide sampling is relatively unbiased using these two quantitative approaches.

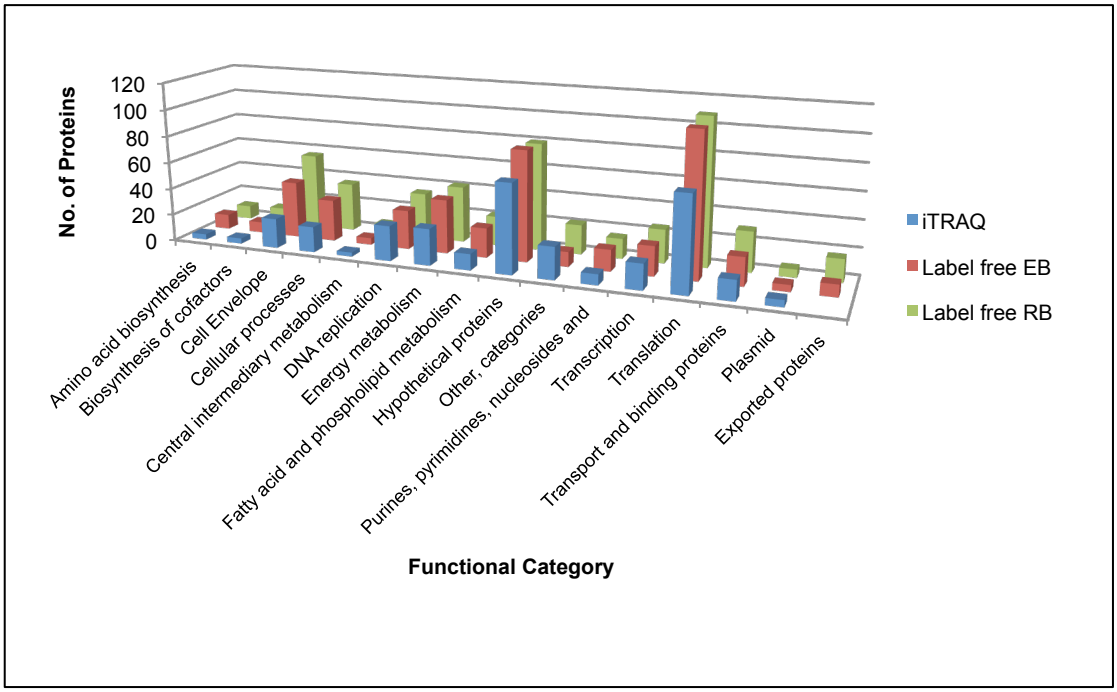


Figure 5.16. Comparison of the distribution of identified proteins between the label-free and iTRAQ approaches according to functional category.

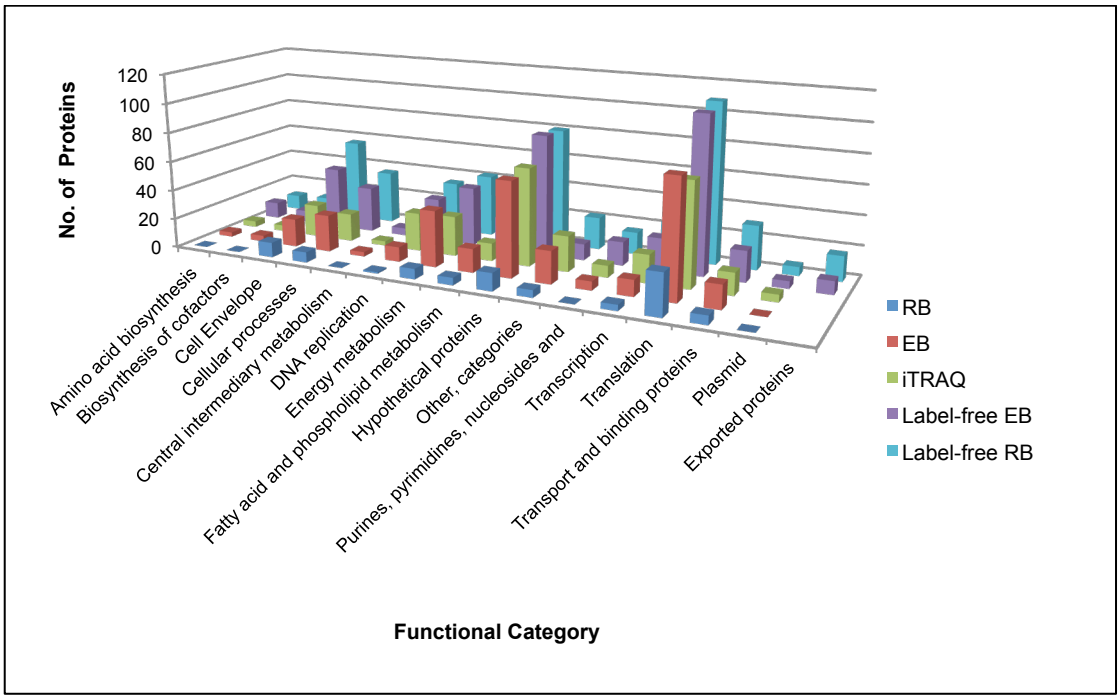
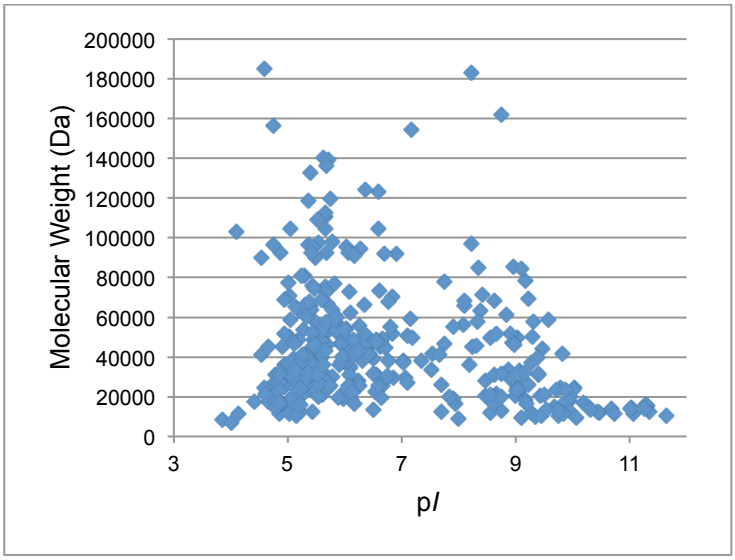


Figure 5.17. Comparison of the distribution of proteins assigned from the label-free, iTRAQ and qualitative approaches presented in Chapter 4 according to functional category. Qualitative data consists of GeLC-MS/MS, MudPIT and 2-DGE from both EB and RB of *C. trachomatis* L2. Both qualitative and quantitative data was included from both the label-free and iTRAQ approaches.

a) iTRAQ experimental data



b) Theoretical data

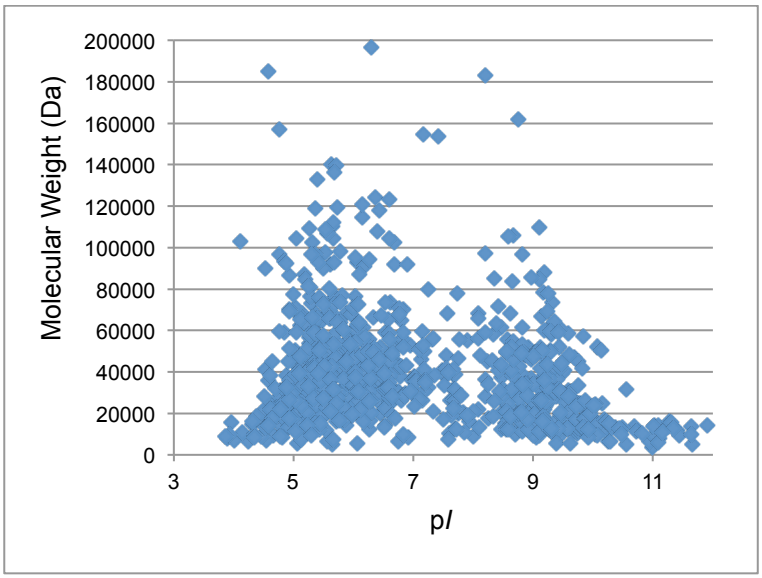


Figure 5.18. Molecular weight and p*I* distributions for iTRAQ data from *C. trachomatis* L2. Graph a) shows the molecular weight and p*I* profile for the experimental iTRAQ data and graph b), the theoretical molecular mass and p*I* distribution for the *C. trachomatis* L2/434/Bu proteome.

The measurement of protein quantity in absolute amounts (molecules/cell) offered the opportunity to calculate, to a first approximation, the total amount of energy expended in synthesizing a specific protein within a cell. It has been reported that 60 kcal/mol are required to extend a nascent polypeptide chain by a single amino acid (Szaflarski and Nierhaus, 2007). This energy expenditure arises from (i) the synthesis of a codon encoding a specific amino acid, (ii) the charging of the tRNA by its synthetase with the cognate amino acid, and (iii) the incorporation of the amino acid into the nascent polypeptide chain. The required energy for this process is generated from the hydrolysis of 10 energy-rich bonds in the form of either ATP or GTP, each with an energy content of about ΔG = -6 kcal/mol. Therefore, by calculating the number of constituent amino acids that make-up a specific protein and using the avogadro constant, the amount of energy required to

synthesize a single molecule of that protein can be calculated. Knowing the total number of molecules synthesised during the transition from RB to EB has permitted the amount of energy expended in synthesizing specific proteins to be calculated. This energy expenditure has been expressed in terms of functional distribution and is shown in **Figures 5.19** and **5.20.**

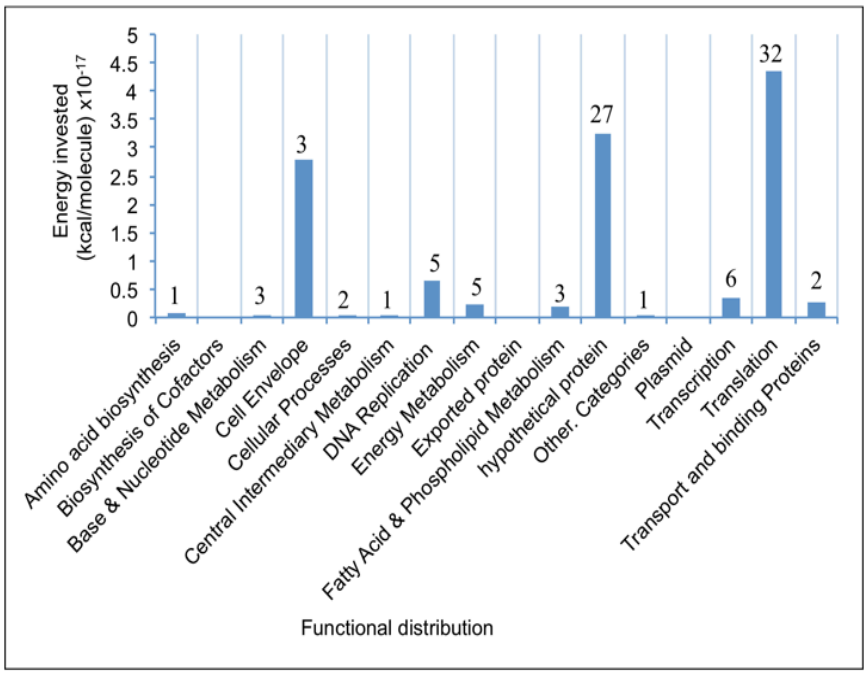


Figure 5.19. Energy expended synthesizing proteins in *C. trachomatis* L2 during the transition from RB to EB represented as functional category. The number of proteins for each category are indicated above each column.

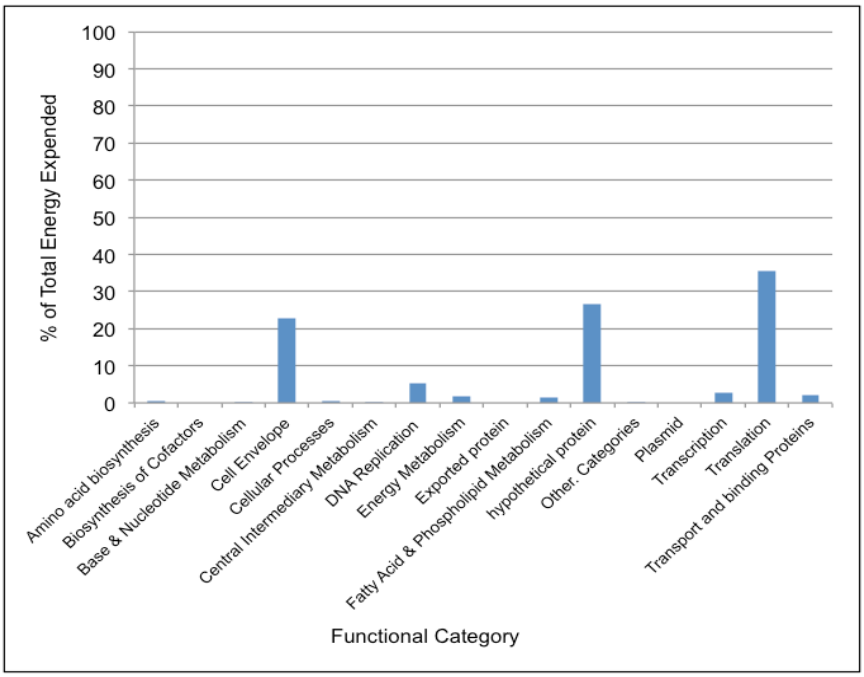


Figure 5.20. Percentage of the total energy expended synthesizing proteins during the transition from RB to EB according to functional category.

The energy in kcal/mol required to synthesise the total number of molecules from each different protein expressed in *C. trachomatis* L2 was calculated for each protein that was up-regulated in EBs. Proteins were grouped according to functional category and the energy values for each protein within that functional category summed. The majority of the energy expended synthesizing new proteins fell into three functional categories, cell envelope, hypothetical proteins and protein translation (see **Figures 5.19** and **5.20**).

5.6 Discussion

This chapter presents a quantitative analysis of the *C. trachomatis* L2 proteome during the late stages of the developmental cycle when RBs re-organise to EBs. The study has exploited two quantitative technologies; the multiplex isobaric labeling technology iTRAQ in combination with multidimensional liquid chromatography; and a label-free approach combining 2D–RP-RP UPLC and the data independent acquisition method MS^E. To the best of the author's knowledge, both captured datasets represent the largest quantitative proteomic studies of chlamydiae to date.

5.6.1 Comparison of technologies

Although iTRAQ and the label-free approach allow qualitative and quantitative data to be collected concurrently, the multiplex nature of iTRAQ allowed the simultaneous analysis of 4 samples, whereas samples using the label-free methodology were analysed consecutively. The demands on MS instrument time for both technologies are high and comparable, with the label-free and iTRAQ taking 88 and 82 hrs, respectively, for four samples. However, recently, the iTRAQ approach has been extended to allow the analysis of up to 8 samples within the same experiment (Choe *et al*., 2007), doubling the sample throughput without increasing MS instrument time. Although the same number of samples could be analysed using the label-free approach, this would require double the amount of MS instrument time (176 hrs). But, at the expense of MS machine time, there is no limit on the number of samples that can be analysed and compared by the MS^E technology *per se*.

Between the two technologies, the amount of protein required to generate an adequate dataset varied considerably, with only 9.0 µg of protein required for the analysis of each label-free sample, compared to ~100 µg for each individual iTRAQ labelled sample. These lower sample requirements of the label-free system offer a clear advantage, particularly when sample amounts are limiting. The lower sample requirements of the label-free approach may allow additional replicate analyses to be performed, thereby improving confidence in the collected datasets (Chong *et al*., 2006) or allowing the analysis of biological samples where previously low sample yield would have negated their analysis.

A potential disadvantage of the label-free approach is its reliance on reproducible high-resolution chromatographic separation and exact mass measurement of samples to allow precursor/fragment ion correlation and peak/peptide matching from sample to sample. Nonetheless, the technology showed excellent reproducibility between technical replicates, with 87% of identified proteins providing quantitative data. By contrast, although the iTRAQ technology showed good peptide replication between samples, a characteristic intrinsic to multiplex technologies, only ~50% of identified proteins yielded quantitative data after appropriate filtering.

This significant difference can be attributed to a number of factors: (i) the number of peptides used to assign proteins. The average number of peptides used to assign a protein using iTRAQ was 10, compared to an average of 46 for the label-free approach, resulting in significantly lower protein sequence coverage for the former data. (ii) the number of peptides with sufficient reporter ion information. For example, the confident assignment of the Cysteine-Rich protein (OmcB), which was confidently assigned with 28 peptides, provided only 9 peptides with sufficient iTRAQ reporter ion information to calculate the proteins expression ratio. (iii) Further compounding this problem are the observed variations in peptide reporter ion ratios, which are particularly variable when the number of peptides used to assign, and hence quantify a particular protein, are low. For example, where a protein assignment is based upon only two peptides and there are significant differences between the observed reporter ion ratios for each, it is not possible to differentiate which peptide ratio represents the true value of the expression change and which is the outlier. As such, those proteins with higher numbers of assigned peptides with sufficient reporter ion information will provide a more precise quantitative measurement through improved statistical averaging. However, as shown earlier in this chapter, the number of peptides used to assign proteins using iTRAQ, were significantly lower than those obtained for the label-free approach and the correlation of the expression changes between the two technologies was poor. When comparing the correlation of expression changes between the technologies for abundant proteins, the direction of change correlates reasonably well, although the magnitude of the iTRAQ change generally appears notably less than the label-free data. As iTRAQ has become more adopted as a quantitative proteomic technology, there has been some concern on the accuracy of the relative abundance estimates obtained. Towards the end of this study, more in-depth studies evaluating the precision and accuracy of iTRAQ have been reported (Karp *et al*., 2010; Ow *et al*., 2009; Wang *et al*., 2012). These studies demonstrated that iTRAQ could provide quantitative data within two orders of magnitude in simple mixtures, but suffers from variance in peptide ion ratios in low signal peptides. As these datasets are dominated by low signal peptides, there is a propensity towards one ratio, leading to an under-estimation of the relative abundance. This under-estimation arises from contamination in precursor ion selection of the peptides for MS/MS. However, since this cross-talk is consistent across the experiment, there is a linear relationship between the expected and observed ratios, opening-up the opportunity for the development of a corrective software solution.

As shown in this Chapter, the label-free technology provided extensive proteome coverage compared to iTRAQ (**Table 5.6**). These improvements are attributed largely to the LCMS^E mode of data acquisition employed. Unlike Data-Directed-Acquisitions used in traditional LC-MS/MS, the limitation on the number of peptides that can be isolated by the Q-Tof's quadrupole for fragmentation during the time frame of an eluting chromatographic peak, is essentially removed when performing LC-MS^E. Acquiring exact mass MS data using an alternating scan function, switching between low and elevated collision energies, provides a data independent mode of acquisition allowing thousands of ions to be de-convoluted into a list of precursor and their associated time-resolved fragment ions that can be searched against an appropriate database. Additionally, since there is no disruption to MS data collection resulting from switching of the quadrupole onto selected ions, the low energy spectra can be used to make multiple precursor ion intensity measurements

across a chromatographic peak, generating accurate reproducible intensity measurements that can be used for quantitation.

One of the most important distinguishing characteristics between iTRAQ and the label-free technology is the ability to use the latters intensity measurements to provide what is considered the gold-standard of quantitative proteomics, absolute quantitation. By employing the relationship reported by Silva *et al*. (2006b), the average intensity of the top three peptides from the internal standard alcohol dehydrogenase was used as a response factor that was applied to all other proteins within each chlamydial protein lysate, thereby allowing an estimate of the number of protein copies per cell to be determined. As discussed in Chapter 1, several approaches using stable isotopes in combination with separation and MS analysis have been described (Gygi *et al*.,1999a; Conrads *et al*., 2001). However, such techniques offer only a targeted approach to absolute quantitation because they require the synthesis of one or more isotopically labeled peptides per protein to be quantified, a technique that is limited for the analysis of highly complex mixtures both in terms of cost and time. The ability of this technology to utilize native proteins as internal standards to calculate absolute concentration measurements of any characterized protein within a complex mixture is extremely powerful.

Therefore, in the light of the absolute quantitation data generated using the label-free technology and because of the poor correlation with the obtained iTRAQ data, further discussion will focus on the data obtained using the label-free approach.

Table 5.6. A comparison of the iTRAQ and Label-free approaches using information from the datasets generated.

	iTRAQ (4-plex)	**Label-free**
Sample requirements (µg)	100 µg per iTRAQ label (400 µg total loading)	9.0 µg per sample (36 µg for 4 samples/replicates)
Total analysis time	~6 days (per 4-plex)	~4 days (4 samples)
Total instrument time	82 h (per 4-plex)	88 h(4 samples) 22 h per sample
Size of data file	41 fractions x ~272 MB (11.15 GB per 4-plex)	11 fractions x ~572 MB (6.29 GB per sample; 4 samples = 25.2 GB)
No. of proteins confidently identified	335	580
No. of proteins with quantitative data	169	489
Type of quantitation	Relative	Absolute
Average No. of peptides per protein	10	46
Average protein sequence coverage	26%	64%

5.6.2 Biological insights

Six hundred and fifteen *C. trachomatis* proteins were identified from 16 functional categories using a combination of both iTRAQ and label-free, representing ~68% of the total proteins predicted to be expressed by the chlamydial genome. Over-represented categories in both RBs and EBs included functions such as cell envelope proteins, cellular processes, energy metabolism, hypothetical proteins, DNA replication and protein translation. Using the label-free technology, estimates of the average number of protein copies per cell were obtained for ~55% of the proteome in both RBs and EBs. However, the significant variation observed in this study between biological replicates and in particular those from RB preparations, has limited in certain instances, the reporting of precise –fold change to observed trends of differential expression. Purification of the osmotically fragile RB is challenging. The variation observed between the two biological replicates of RBs, is likely to reflect the difficulties in their isolation and purification. Although time-consuming and costly, the analysis of an increased number of biological replicates could help improve the measurement of this variation.

5.6.2.1 Outer membrane proteins

The structure of the chlamydial cell envelope is basically similar to that of other gram-negative bacteria with an outer membrane containing lipopolysaccharide, a periplasm and an inner membrane. Nonetheless, there are two distinct features that are unique to the chlamydiae. These are the absence or, at the very most, low levels of peptidoglycan and the prescence of disulphide-bond-cross-linked-proteins making up the outer membrane.

DNA condensation and the formation of this highly cross-linked outer membrane complex are a hallmark event in the transition from RB to EB. Genes previously characterized as being expressed during these late stage processes include, an integration host factor (*ihfA*), *recA* and the histone-like proteins (*hctA* and *hctB*), implicated in mediating chromosomal condensation (Brickmann *et al.,* 1993; Barry *et al*., 1993). The Major Outer Membrane Protein (MOMP), the two cysteine-rich proteins OmcA and OmcB and predicted thioredoxin disulphide isomerases are recognizable key proteins associated with the formation of the highly cross-linked outer membrane complex. Of the nucleoid-associated proteins detected in this study, IfhA showed a significant up-regulation of 2.3-fold which is consistent with gene expression studies (Nicholson *et al*., 2003; Zhong *et al*., 2001b). The histone-like protein HctB was not detected, however, the protein encoded by the *hctA* gene was detected and surprisingly showed a marked 2.7-fold decrease in expression. Although the detection of this protein has not been reported in previous proteomic studies, this observation is contrary to gene expression data where the transcript has been shown to increase in expression from 36 hpi (Nicholson *et al*., 2003). A possible explanation for this disparity may be attributed to the function of the histone-like proteins. If these proteins perform similar functions to those of other characterized histones, HctA maybe complexed with DNA forming a higher–ordered structure like chromatin. Such structures may make it inaccessible to proteolysis, thus preventing or reducing the detection

of HctA derived peptides, resulting in an artfactual decrease of observed protein abundance. Alternatively, it is also possible that this dense DNA-histone complex may have been partially depleted during the sample preparation process.

The highly abundant major outer membrane protein (MOMP) showed only a marginal increase in protein expression between 15 and 48 h PI, suggesting the translation of this protein occurs earlier in the developmental cycle. This is consistent with mRNA and other protein studies, where the expression of the MOMP transcript was maintained from 15 h to 48 h (Belland *et al*., 2003; Hatch *et al*., 1984). The largest observed expression changes for the COMC associated proteins was the 60 kDa cysteine-rich protein (OmcB), which was detected in both RBs and EBs. Surprisingly this protein was only detected in EBs in the qualitative studies presented in Chapter 4. Similar to the conflicting results obtained for the histone-like protein (HctA), this protein also showed a 3.3-fold decrease, contrary to previous observations of an up-regulation of mRNA transcripts late in the developmental cycle. Hatch *et al*. (1986) showed that although MOMP monomer could be released from the COMC, the 60 kDa cysteine-rich protein could not be reduced to monomers even in the presence of reducing agents such as dithiothreitol. It is therefore plausible that peptides from OmcB are inaccessible or not efficiently released from the COMC leading to an apparent decrease in OmcB peptide/protein expression.

5.6.2.2 Polymorphic Outer membrane proteins

Another unique set of genes associated with the chlamydial outer membrane, and that have no homologue in other bacteria, are the polymorphic outer membrane proteins (Pmps). This study provides a quantitative measure of all nine putative polymorphic outer membrane proteins encoded by the C. *trachomatis* L2/434/Bu genome. The expression of all nine Pmp proteins was down-regulated from 15 to 48 h PI, ranging between 1.6 and 4.8–fold. This is in agreement with previous gene expression studies where transcripts orthologous to all nine of these *pmps* in L2 were detected early in the developmental cycle at 8 h PI (Lindquist and Stephens., 1998) and more recently by Nunes *et al*. (2007), who showed that all nine transcripts were detected as early as 2 h PI. The expression of these transcripts peaked at ~18 h PI, where they were stable until 36/48 h PI when the expression levels decreased. Of all the Pmps, PmpA and PmpF showed the lowest level of protein expression at 15 h PI and further decreased at 48 h PI, a result that is consistent with the maximal transcript expression of *PmpA* observed before 12 h PI, supporting a role for this protein early in the developmental cycle (Nunes *et al*., 2007). PmpD showed the highest level of protein expression in both developmental forms, albeit lower in EBs. By contrast, *pmpF* showed the highest mRNA levels, but the lowest protein expression levels indicating possible post-transcriptional regulation of this gene, an observation that is in accordance with the mRNA expression disparities observed for *pmpF* and *pmpE* between reference strain L2/434 and clinical strains (Nunes *et al*., 2007). The role of Pmps in chlamydial biology and disease pathogenesis is unknown. However, there are some serovar-specific differences between the Pmps that may result in differences in virulence and tissue tropism (Longbottom *et al*., 1998; Gomes *et al*., 2006; Stothard *et al*., 2003). They are characterized by a C-terminal phenylalanine, GGAI motifs and cleavable signal peptides, suggesting they are located in the outer membrane (Struyve *et al*., 1991). The Pmps are considered to be autotransporter proteins, part of the type V secretion pathway and such a function that

has been experimentally confirmed for PmpD in *C. trachomatis* and Pmp21, a *C. pneumoniae* orthologue of PmpD from *C. trachomatis* (Kiselev *et al*., 2007). Shaw *et al.* (2002a), detected peptides mapping to the C-terminal part of the PmpD sequence, suggesting that this protein had been processed in keeping with their role as autotransporters. In contrast, we detected peptides mapping the N-terminal, C-terminal and internal fragments, with the exception of the signal sequence (**Figure 5.21**). Indeed, coverage of the N-terminal, C-terminal and internal fragments were detected for all the Pmp proteins. However, in keeping with their role as autotransporters, they may still be processed and the detection of N-terminal and C-terminal fragments could simply reflect the high sensitivity of the technology used and represent proteins yet unprocessed. This observation highlights an advantage in using 2D gel technology, where isoforms can be easily resolved with quantitation at the protein level in comparison to other technologies performing quantitation at the peptide level, e.g., label-free and iTRAQ approaches.

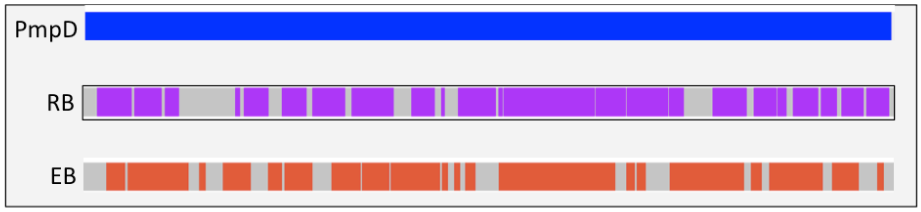


Figure 5.21. Diagram indicating the peptide coverage obtained for the protein PmpD from *C. trachomatis* L2 in EBs and RBs using label-free. The ORF of PmpD is indicated in blue. Regions highlighted in purple and orange indicate the detection of PmpD peptides in RBs and EBs, respectively.

5.6.2.3 The Chlamydial cell wall anomaly

In most bacteria, the cell wall or peptidoglycan (PG) layer is essential, providing structural strength to resist osmotic stresses. Previous attempts to detect PG components such as N-acetylmuramic acid (MurNAc) in chlamydiae have been unsuccessful, suggesting that the chlamydial cell wall does not contain PG, or if it does, in only small amounts. The osmotic stability of EBs is afforded by the highly cross-linked cysteine–rich proteins present in the cell envelope, which likely replaces the requirement for a peptidoglycan layer. The osmotically fragile RB possess fewer cross-linked proteins. However, they are protected from osmotic stresses by the membrane-bound inclusion within the host cell and so may not require the structural strength provided by a PG layer. As discussed in Chapter 1, the discovery of a complete set of biosynthetic genes for peptidoglycan synthesis in the genome sequence was therefore a surprise and raised the question whether these enzymes and their products are synthesized or are they simply genetic remnants.

Peptidoglycan biosynthesis in other bacteria occurs in three compartments. The precursor, UDP-MurNAc pentapeptide is synthesized in the cytoplasm by six enzymes (MurA to MurF). This precursor is subsequently transferred to the lipid carrier undecaprenyl phosphate, catalyzed by MraY, to form the first membrane bound intermediate, Lipid I. Catalysed by MurG, Lipid II is synthesized by the addition of UDP-GlcNAc to Lipid I, followed by translocation into the peptidoglycan structure (**Figure 5.22**). Gene expression studies have shown the expression of mRNA from this set of genes shows a marked increase at 16-18 h PI

(Nicholson *et al*., 2003; Belland *et al*., 2003). At the translational level, only a few PG synthesis enzymes have been detected. Previous studies have detected MurG and MurC in EBs (Montigiani *et al*., 2002; Vandahl *et al*., 2001) and in the qualitative studies presented in Chapter 4, MurE was detected in RBs only. The datasets presented in this Chapter and summarized in **Table 5.7** provide further evidence for the translation of these enzymes. Specifically, these include MurA, MurC, MurD, and the two penicillin binding proteins, whose expression has previously been detected using radiolabelled penicillin (Barbour *et al*., 1982).

When chlamydiae–infected cells are treated with inhibitors of cell wall synthesis such as penicillin, chlamydial replication is halted with the formation of large aberrant RBs, suggesting a role for PG in RBs. Since the RB form is most like other gram-negative bacterium, and undergoes division, the presence of PG would require remodeling during division and would therefore be susceptible to penicillin treatment. At the same time, the osmotically protective niche of the inclusion may mean that, if PG has a role in maintaining structural integrity, it is only minimal. With the exception of one unsuccessful attempt to directly detect PG in RBs (Barbour *et al*., 1982), all others have been in EBs since the osmotic fragility of RBs has previously prevented their isolation. This is also true for the detection of enzymes involved in PG. However, evidence for a peptidoglycan-like structure in RBs has been obtained using immune detection (Brown & Rockey, 2000). The quantitative data reported here suggests that these enzymes are detected in RBs, and if detected in EBs, their expression levels are low, indicated by their presence in only a single replicate compared to RBs. Where detected in the label-free dataset, the amounts of these enzymes were low (MurC = 42±28 molecules/RB and 20±10 molecules/RB). Assuming normal catalytic efficiencies, this suggests that PG would only be produced in small amounts. This decrease is concomitant with the maintenance in RBs and EBs of the glycoside hydrolase, muramidase, which breaks down PG by hydrolyzing the 1,4 beta-linkages between N-acetyl-D-glucosamine and N-acetylmuramic acid. So what is the role of PG? Chlamydiae have reduced their genome size to become an efficient intracellular organism; therefore, why expend energy synthesizing these enzymes unless PG is required? It has previously been proposed that PG plays a role in cell division (Chopra *et al*., 1998) and that PG in the absence of the FtsZ homologue (a conserved tubulin-like protein that polymerizes into ring-structures at the mid-cell of dividing cells) provides a replacement role. The data presented shows that maximal expression of these enzymes occurs in RB at 15 h PI, albeit at low levels, supporting an alternative role for PG in RBs such as cell division.

Another possible role for PG may be in host cell signaling. Since PG is unique to bacteria, innate immunity systems of the host detect liberated PG fragments via specific receptors and subsequently initiate an immune response. The major family of receptors that sense PG and other bacterial components include the Toll-like receptors (TLRs) and Nod-like receptors (NLRs) (Chen *et al*., 2009; Takeda *et al*., 2003; Girardin *et al*., 2003). As such, bacterial systems have evolved mechanisms to evade or subvert such pathways to allow replication and minimize non-advantageous host responses. These mechanisms include additional structures on the cell surface, to reduce PG break down, and modification of the PG structure to prevent degradation by host cell enzymes. Several studies have also shown that signaling of the NLR receptor, Nod1, can be altered by arresting the release of Nod1 signaling ligands by modifying PG and preventing its degradation (Boneca *et al*., 2007; Shimada *et al*., 2010). There is also some evidence that signaling and initiating a modified host

response can be beneficial to the bacterium. For example, initiating a response that recruits certain host cells to the site of infection can provide a replicative niche for intracellular pathogens (Girardin *et al*., 2003). Could it be that Chlamydiae produce a PG with modified structures that once released, initiates an altered immune response that benefits this pathogen? This could be either through evading certain unfavorable pathways or by recruiting host cells to initiate the next round of infection, thereby ensuring the shortest time in the harsh environment outside the host cell. The higher expression of the PG enzymes in RBs, compared to EBs, could indicate that synthesis of a modified PG structure occurs in RBs. This structure may be prepared in readiness for hydrolysis in EBs by muramidase, and the resulting fragments could subsequently initiate a modified host response that is beneficial to chlamydiae. However, the definitive role(s) of PG and its associated biosynthetic enzymes still remains unclear. The recent development of a genetic transformation system for *C. trachomatis* (Wang *et al*., 2011) provides the opportunity to resolve the role of PG in *Chlamydia*.

Table 5.7. Peptidoglycan biosynthetic enzymes expressed in *C. trachomatis* L2

Protein	Observed in RB	Observed in EB	Technique
MurA	Yes*	Yes*	iTRAQ (qualitative)
MurB	No	No	n/a
MurC	Yes (42 molecules/cell) ND	Yes (22 molecules/cell) Yes	Label-free Previously detected (Vandahl *et al*., 2001)
MurD	Yes (20 molecules/cell)	No	Label free iTRAQ (qualitative)
MurE	Yes	No	Qualitative (chapter 4)
MurF	No	No	n/a
MurG	No	No	Detected in Ebs from previous study (Montigiani *et al*., 2002)
PBP	No	Yes	Label-free(qualitative)
PBPB	Yes*	Yes*	iTRAQ (qualitative)

* Unable to determine whether detected in EBs or RBs due to multiplex nature of iTRAQ.

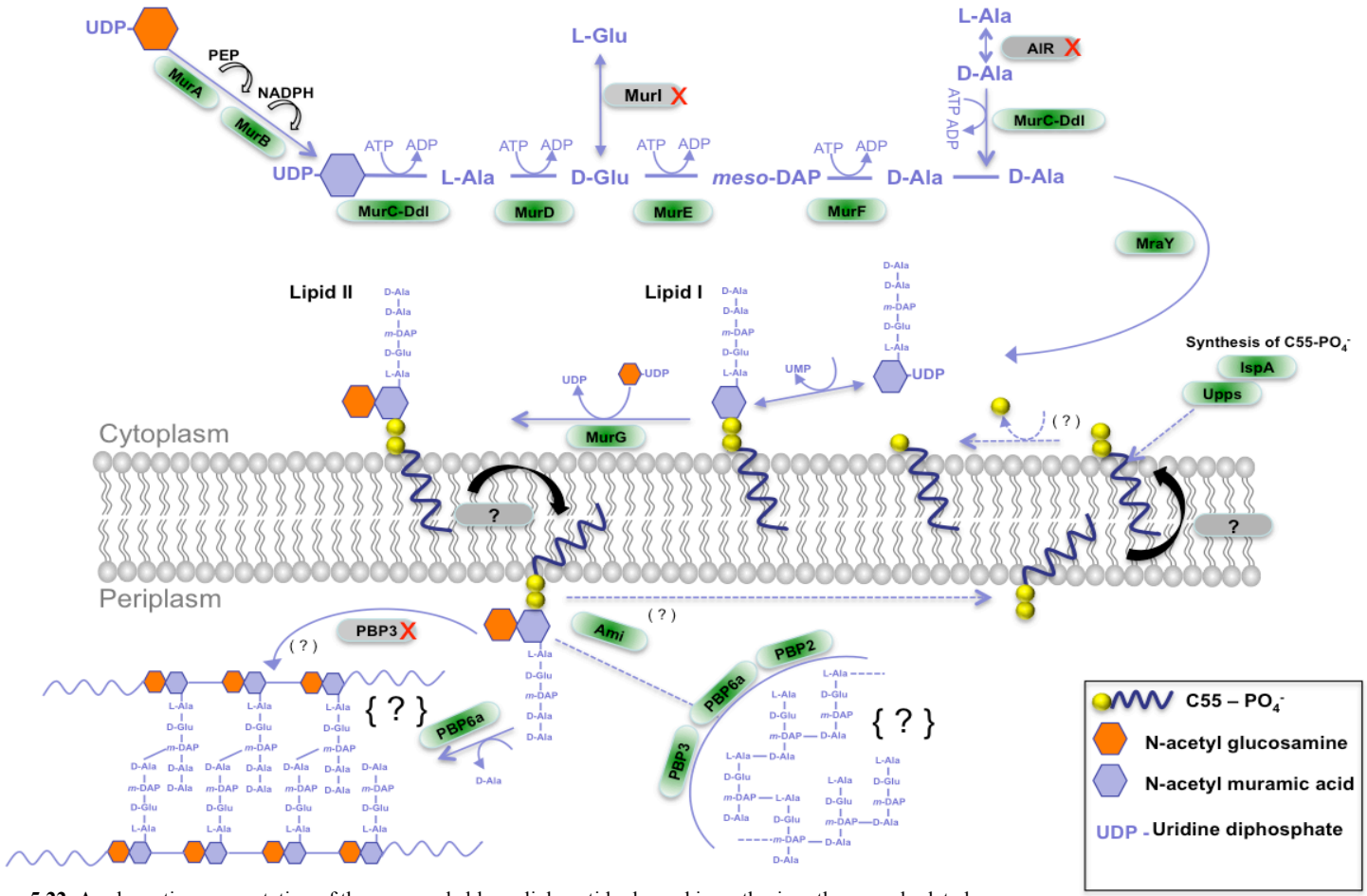


Figure 5.22. A schematic representation of the proposed chlamydial peptidoglycan biosynthesis pathway and related genes.
The precursor, UDP-MurNAc pentapeptide is synthesized in the cytoplasm by six enzymes (MurA to Mur F). This precursor is subsequently transferred to the lipid carrier undecaprenyl phosphate catalyzed by MraY to form the first membrane bound intermediate, Lipid I. Catalysed by MurG, Lipid II is synthesized by the addition of UDP-GlcNAc to Lipid I, followed by translocation into the peptidoglycan structure.

5.6.2.3 Type III secretion

Chapter 4 identified seven predicted proteins associated with the chlamydial TTSS. An additional 37 proteins were identified in this quantitative study, providing further support for the idea that this organism expresses a functional TTSS. The *Chlamydiaceae* encode between 20 to 30 genes for structural proteins and chaperones of the TTSS depending upon the species. However, the minimal set of proteins that constitutes a functional apparatus still remains unclear (Peters *et al*., 2007; Hefty and Stephens, 2007). Fifteen structural proteins (**Table 5.8**), 4 chaperones and 25 effector proteins of the TTSS were detected and quantified in this study. TTSS chlamydial effector proteins unlike in other pathogens are distributed throughout the genome. The considerable efforts to determine exactly how many there are and the elucidation of their functions, is still ongoing work (Subtil *et al*., 2001; Fields *et al*., 2003; Fields *et al*., 2005; Lugert *et al*., 2004; Clifton *et al*., 2005; Subtil *et al*., 2005; Jamison and Hackstadt, 2008; Kleba and Stephens, 2008; reviewed by Beeckman and Vanrompay, 2010). Nonetheless, several elegant studies have demonstrated the secretion and localization of predicted effector proteins. These include the inclusion membrane protein IncA, which was shown to localize to the cytoplasmic side of the inclusion membrane in both *C. trachomatis* and *C. psittaci* (Bannatine *et al*., 1998; Rockey *et al*., 1997) and the predicted *C. pneumoniae* proteins IncA, IncB, IncC, Cpn0809, Cpn1020, SctN, SctW and LcrH1. These have all been indicated to be secreted into the chlamydial inclusion (reviewed by Beeckman and Vanrompay 2010).

Electron microscopic studies revealed a patch of type-III-like structures on the chlamydial surface. In *C. psittaci*, the mean number of these projections was 45 in RB at 10 h PI, decreasing to 20 in late RBs (20 h PI) and leveling to about 18 later in EBs at 48 h PI (Matsumoto, 1982). The expression levels of the predicted structural TTSS proteins shown in **Table 5.8**, with the exception of SctW, were generally expressed at reduced levels in the late stage (48 h PI) of infection and in some cases, were only present in RBs at 15 h PI. This observation is consistent with an observed decrease in the type-III-like apparatus later in the cycle. Similarly, the predicted effector proteins, including IncA, IncE, IncG, IncC, and 7 additional predicted inclusion membrane proteins (not detected in the qualitative studies presented in Chapter 4) were also detected and, where quantitative data was available, they showed decreased levels of expression in EBs, indicating that they may have been secreted. The exception to this was the effector chaperone Mcsc, which was equally abundant in both RBs and EBs and the effector protein CADD, which was more abundant in EBs.

A proposed hypothesis consistent with these observations is that these type-III-like projections are induced/or activated upon contact of the RBs with the juxtaposed inclusion membrane. Activation during the early transition from EB to RB, allows the delivery of effector proteins, until detachment of RBs from the inclusion and subsequent deactivation of the type-III apparatus later in the developmental cycle (Bavoil and Hsia., 1998). In addition to the Inc proteins, the predicted type III effector proteins CopD, CopB CopN and Pkn5 were also expressed in both EBs and RBs, with the exception of Pkn5, which was only detected in RBs. The translocation of these chlamydial effector proteins has been demonstrated by the *Salmonella*

Typhimurium type III secretion systems (Ho and Starnbach, 2005) and CopD has previously been shown to be abundant in EBs (Fields *et al*., 2003; Shaw *et al*., 2002a). Interestingly, the effector protein CADD (*Chlamydia* protein associating with death domains) has previously been shown to be expressed late in the developmental cycle and to modulate host cell apoptosis (Stenner-Liewen *et al*., 2002). In keeping with this, an increased expression of this TTSS protein was observed in EBs at 48 h PI. This data confirms the expression of the predicted effector proteins CopD, CopB, CopN, Pkn5 and CADD in *Chlamydia.*

A functional TTSS is also supported at the initial stages of infection by the presence of a predicted type-III effector protein, TARP. This protein is translocated into the host cell and phosphorylated, directly nucleating actin filament formation, promoting rapid filament polymerization essential for EB entry (Jewett *et al*., 2006). It is likely that there are many other proteins that are also translocated during the early stages to assist in EB entry, avoiding the innate immune responses and modulating maturation of the EB-containing endosome. It has been hypothesized that these early stage effector proteins such as TARP are synthesized at the later stages of the cycle and are pre-packaged into EBs and ready-to-go for translocation into the host upon attachment. Contrary to this, we observed a 2.5 -fold decrease in the levels of TARP from 15 to 48 h PI. However, it is unknown whether TARP functions at any other point within the developmental cycle and the observed fluctuations could simply reflect a modulation of protein levels according to demand. As discussed later in this Chapter, another possible hypothesis is that EBs are not as metabolically inert as originally proposed. In addition to certain proteins being synthesized and ready-to-go during the later stages, there may be a low level of protein synthesis in the extracellular form.

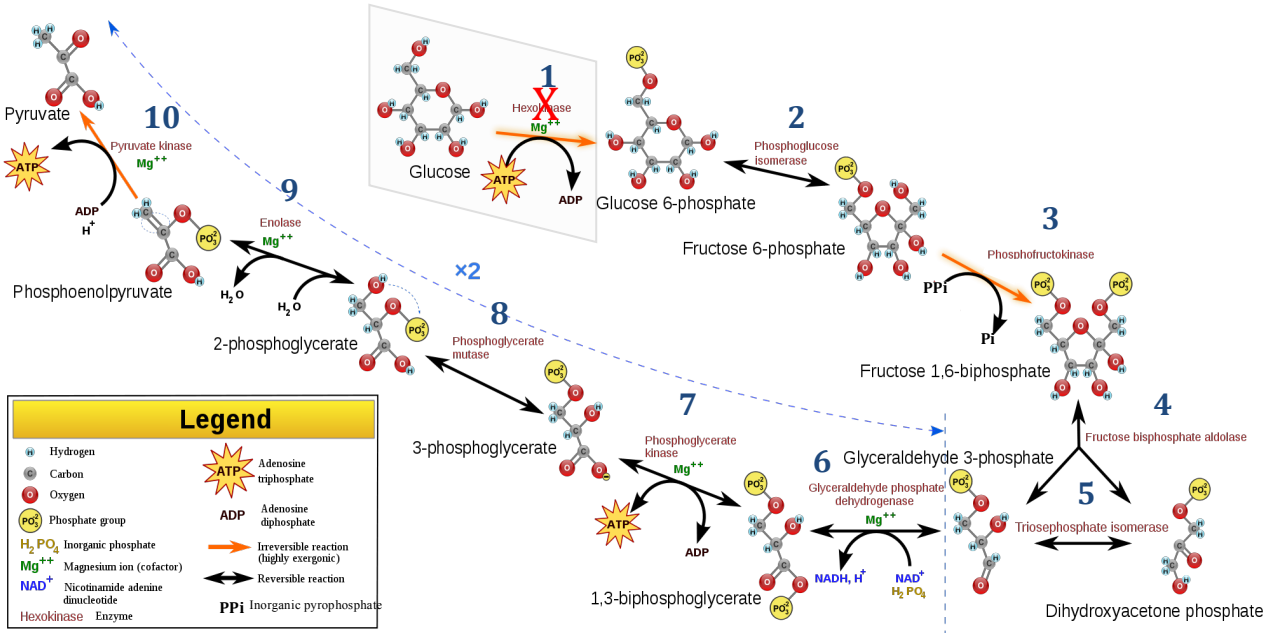
Table 5.8 Structural proteins of the Type III secretion apparatus identified.
IM: inner membrane; OM: outer membrane; CP: cytoplasmic; MA: membrane associated; HCM: host cell membrane; TL: translocon component. Nomenclature adapted from Beeckman and Vanrompay, (2010).

Locus	Protein name	RB molecules/cell	EB molecules/cell	Location
CTL0345	SctV/LcrD/CdsV	252±132	65±20	IM
CTL0344	SctW/LcrE/CopN	54±11	42±18	Secreted
CTL0038	SctN/CdsN	91±11	45±18	CP:MA
CTL0041	SctQ/CdsQ	162±68	54±10	IM
CTL0825	SctR/CdsR	32±16	ND	IM
CTL0826	SctS/CdS	ND	ND	IM
CTL0827	ScT/CdsT	ND	ND	IM
CTL0036	SctU/CdsU	33±10	ND	IM
CTL0043	SctC/CdsC	260±125	75±12	OM
CTL0033	SctD/CdsD	924±441	341±7	IM
CTL0035	SctF/CdsF	ND	ND	Needle
CTL0822	SctJ/CdsJ	566±249	228±24	IM & OM
CTL0824	SctL/CdsL	159±35	115±27	CP
CTL0841	CopB	73±8	31±5	HCM
CTL0842	CopD	134±21	76±12	HCM TL
CTL0238	LcrV	70±14	55±28	HCM TL

5.6.2.4 Energy metabolism

Genome sequencing has changed our view of central metabolism in the chlamydiae. Originally thought to be energy parasites, incapable of synthesizing their own ATP and dependent on ATP and high-energy metabolites from the host cell (Moulder *et al*., 1962), it was a surprise when genes for an intact glycolytic pathway were discovered. But, like the peptidoglycan anomaly, it has not been clear whether these predicted genes are expressed at the level of translation. Chapter 4 reported the detection of an entire set of predicted components of the glycolytic pathway, in addition to the predicted enzymes of a partial TCA cycle and a complete pentose phosphate pathway. Vandahl *et al*. (2001) also reported the detection in EBs of seven enzymes involved in glycolysis and three in each of the pentose phosphate pathway and partial TCA cycle. A number of these recombinant chlamydial enzymes, including pyruvate kinase (PK), glyceraldehyde-3-

phosphate dehydrogenase (GAPDH), phosphoglycerate kinase (PK) and glucose-6-phosphate dehydrogenase (ZWF) have also been cloned, expressed and shown to be functional in *E. coli* (Iliffe-Lee and McClarty, 1999). The quantitative measurements in this study indicate that these glucose metabolism enzymes, including the ADP/ATP translocase, are at their most abundant in the metabolically active RB form (15 h PI), and show a general trend for decreased levels of expression in EBs (48 h PI). A summary of the expression levels of these glycolytic enzymes is shown in **Figure 5.23**. Despite the higher levels of these enzymes in RBs, the abundance of these enzymes in EBs is still significant, indicating that chlamydiae have the capability to generate ATP via substrate-level phosphorylation throughout most of their developmental cycle. This is consistent with previous RT-PCR results, which showed maximal expression of the genes PK, GAPDH, PGK and ZWF in RBs (Iliffe-Lee and McClarty, 1999). It has been demonstrated that EBs contain large pools of ATP (Tipples and McClarty, 1993) but, unlike RBs, they are unable to obtain ATP from the host cell via the ATP/ADP translocase (Hatch *et al*., 1982). The expression of these enzymes for glucose metabolism strongly suggests that both RBs and EBs are capable of producing ATP. This supports the hypothesis that chlamydia may build up and store ATP using pre-existing enzymes in the extracellular stage. This in turn may aid survival outside the host and/or be of importance during the initial stages of infection and subsequent differentiation into RBs. On initiation of infection, chlamydiae begin their transition to the metabolically active RB, with a higher energy requirement for replication. ATP demands therefore increase and hence the levels of glucose metabolism enzymes increases with further fueling, using host cell ATP obtained via the ADP/ATP translocase. While these results still do not answer whether chlamydiae can survive independently of host ATP, they question the dogma of a metabolically inert extracellular form.



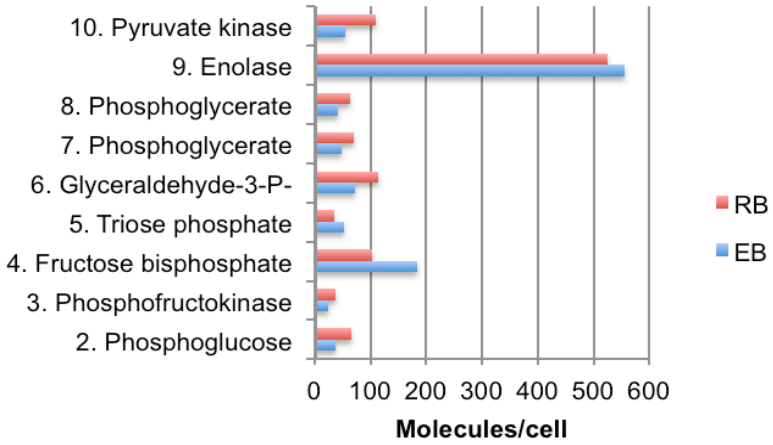


Figure 5.23. Representation of the glycolytic pathway in *C. trachomatis* L2 and associated expression levels of each glycolytic enzyme from 15 to 48 h PI. There is no Hexokinase homolog in the *C. trachomatis* L2 genome. Adapted from Wikipedia. Retrieved 10th August, 2011 from http://wikipedia.org/wiki/glycolysis

184

5.6.2.4 Global proteome expression profile

In light of the clear phenotypic differences between the two transitional forms of *C. trachomatis,* differences in protein expression between the two forms are to be expected. Mukhopadhyay *et al*. (2006) previously demonstrated using 2-DGE combined with pulse-labeling, the differential expression of 35 proteins in the respiratory pathogen *C. pneumoniae* during the transition from RB to EB, with a range of expression from a 4.6 down- to 14.8 fold up-regulation. By comparison, the range of expression observed for *C. trachomatis* L2 was 8.3 down- to 3.5 fold up-regulation. Although the expression differences between the two experimental datasets could be attributed to strain variation between *C. trachomatis* L2 and *C. pneumoniae,* it is also probable that they are as a consequence of the intrinsic differences between the technologies used in each study. Since pulse-labeling 2-DGE techniques do not allow the detection of proteins prior to the addition of radiolabel, they are unable to provide a measure of the abundance of any pre-existing proteins. Hence, in the case of the *C. pneumoniae* study, proteins synthesized earlier in the developmental cycle prior to radiolabeling are not represented. Because of this constraint, if the rates of synthesis and degradation for a particular protein are similar, although a pulse-labeling 2-DGE experiment may indicate an apparent increase in protein abundance, the net change when taking into account the abundance of a protein prior to the pulse may only be subtle. Since the label-free technology is not reliant upon the metabolic incorporation of a label, this constraint is circumvented and is likely to be more representative of the protein changes occurring.

As highlighted by the shotgun studies of the two developmental forms presented in Chapter 4, both RBs and EBs were highly represented by proteins involved in protein metabolism, transcription and translation. The presence of these functional classes of proteins in the metabolically active replicating RBs was not surprising, but their detection in metabolically inert EBs was unexpected. Albeit, in the absence of quantitative data, the levels at which they exist could simply reflect low-level residual RB proteins. The functional categories of the proteins assigned in this quantitative study reflect those obtained in the qualitative studies presented in Chapter 4. However, when considering the differential expression or further, the proteins that *Chlamydia* expends energy synthesizing during the transition from RB to EB, the profile is quite different.

The availability of absolute quantitation data has allowed, to a first approximation, an estimation of the amount of energy required to synthesize a specific protein within a chlamydial cell. The functional distribution of this expended energy falls into three main categories, cell envelope formation, hypothetical proteins and proteins involved in protein translation. Interestingly, the majority of the energy expended in cell envelope formation can be attributed to the Major Outer Membrane Protein. Although the increased expression of MOMP in EBs is relatively low (1.3 fold), the high abundance of MOMP ($\geq$2000 molecules/RB and $>$2700 molecules/EB) equates to a greater number of molecules synthesized per cell, representing a significant energy burden upon chlamydia. The infectious EB exists as a tough spore-like particle whose role is to provide protection in the harsh conditions outside the host cell. As such, the energy expended in synthesizing MOMP to maintain the outer cell envelope is likely to be an essential requirement.

Previous studies have reported EBs to be metabolically inert (reviewed by Moulder, 1991). The observed trend for the down-regulation of proteins in EBs was therefore not surprising and likely reflects the period when chlamydia begins shutting down its metabolic processes in preparation for survival in the harsh environment outside the host cell. However, the synthesis of proteins attributed mainly to protein translation (32) and hypothetical proteins (27) were unexpected (**Figure 5.19**).

The hypothetical proteins showed some of the largest changes in protein expression observed in this study, ranging from 3.6 fold down-regulation (CTL0255) to the up-regulation of 27 proteins ranging from 1.3 to 3.5 fold (CTL0869) in EBs. A number of these proteins were also highly abundant. By the same token that energy expended synthesizing MOMP is important for survival outside the host cell, using energy synthesizing these uncharacterized chlamydial proteins also suggests that they too may play an essential role in EB maturation, survival outside the host cell and/or infection. Further characterization of these proteins will clearly be an important step in understanding the roles they play in chlamydial biology.

The subtle increased expression or maintenance of ribosomal proteins in EBs (**Figure 5.20**) also presents a significant energy burden. This highlights the necessity for chlamydiae to have an expressed protein translation system in EBs. Why? Interestingly, the mRNA expression profiles of *C. trachomatis* indicate the expression and maintenance of nearly all transcripts from 16 h onwards to the end of the developmental cycle. Subsets of these transcripts are 'carried over' into infectious EBs and others have proposed that these "maternal mRNAs" are being stored and are 'ready-to-go' with translation initiated upon infection (Plaunt & Hatch, 1988). Further support for this proposal comes from the surprisingly low number of new genes expressed early in the developmental cycle (<6 h) when EBs reorganize to RBs. The immediate translation of these stored transcripts would clearly be advantageous to chlamydia's rapid and effective invasion of its host cell. However, such a system would require the necessary molecular machinery for the immediate translation of these stored transcripts. The data presented in this thesis suggests that EBs contain pre-packaged proteins that could provide the necessary molecular machinery for the immediate translation of early genes, essentially priming chlamydiae at the mRNA and protein level so that they can quickly capitalize on contact with a host cell.

Analogous storage and 'ready-to-go' systems have been proposed for other organisms such as those reported for spores of the yeast *Saccharomyces cerevisiae* (Xu and West, 1992). Germinating yeast spores were shown to initiate protein synthesis within 20 min and transcription later at 70 min after the addition of glucose, indicating the presence of mRNA available for immediate translation. However, a more recent study of mRNA turnover during yeast sporulation reported the initiation of transcription and translation at considerably earlier time points than the previously reported 70 min (Brengues *et al*., 2002). Moreover, newly synthesized molecules were detected within spores previously thought to be metabolically inert. They postulated that these spores maintained a basal level of transcriptional and translational activity, with low levels of turnover, which were highly boosted upon spore germination, a process possibly important for spore

survival. Like the yeast spores, EBs are considered metabolically inert (Hatch *et al*., 1985). However, is it possible that EBs are not quiescent cells, but maintain a basal level of metabolic activity similar to yeast?

The survival of *Chlamydia* ultimately depends upon attachment and uptake into the host cell to create a niche suitable for replication. As such it requires rapid and efficient mechanisms to effect uptake and to subvert host cell defenses. The question therefore remains whether chlamydiae create all their necessary molecular machinery for infection *de novo* or have systems primed and 'ready-to-go'. Such systems have already been proposed for *C. trachomatis* at the level of transcription. However, the proteomic studies presented in this thesis provide both qualitative and quantitative evidence that chlamydiae may indeed be primed and 'ready-to-go' at the protein level.

5.6.3 Addendum

During the final stages of preparation of this thesis, Saka *et al*. (2011) reported a label-free quantitative proteomic analysis of *C. trachomatis* L2/434/BU, comparing RBs and EBs. This has offered the opportunity to compare, where possible, the label-free proteomic dataset presented in this chapter with this recently published study. However, it is important to note, as Saka *et al*. do themselves, that the purified RBs and EBs only represent a single time point and may not necessarily reflect RBs and EBs at earlier or later stages. As such, a caveat to this comparison is that this thesis compares RBs at 15 h PI and EBs at 48 h PI cultured in BGMK cells, whereas, Saka *et al.* compare the later time point of 18 h PI for RBs and EBs earlier at 44 h PI in HeLa cells.

Saka *et al.* identified 485 (54% of the proteome) chlamydial proteins of which they obtained quantitative information for 373. By comparison, the study presented in this chapter identified 580, of which 489 had quantitative data. These differences in proteome coverage are likely to reflect the differences in the separation methodology employed. Both studies used 2D-RP-RP-LC-MS/MS, using a high pH RP first dimension followed by low pH RP second dimension. Based upon the high complexity of EB and RB lysates observed in initial LC-MS^E scoping experiments, an 11 step first dimension fractionation was used in this study (*Section 2.6.4*), compared to 5 steps used by Saka *et al*.

Interestingly, both studies found RBs were primed for high metabolic activity, with the expression of proteins required to address the high demand for nutrients, ATP generation and increases in cellular mass. Both studies also showed that the 'inert' EB form express the necessary glycolytic enzymes to metabolize glucose and generate ATP, independent of the host cell. Although there are many similarities between these two datasets, there are also some distinct differences in the observed protein expression profiles. The two main differences are associated with energy metabolism and type III secretion (TTS) and these are now discussed.

5.6.3.1 Type III secretion

Similar to the data presented in this thesis, Saka *et al*. identified the majority of the structural components making up the TTSS, 25 predicted TTSS effectors and 7 TTSS chaperones. Surprisingly, they found that the majority of the structural TTSS proteins were either in higher abundance, or detected exclusively in the EB form. The exception to this was SctD, which was equally abundant in both RBs and EBs. The TTSS chaperones were also markedly absent in the RB form and the authors suggest a reduced TTSS capacity, or a limited number of active TTSS apparatus in RBs. Notably, the absence of the C-ring components of the TTSS basal body, SctQ and the ATPase, SctN in RBs, leads the authors to suggest substitutes for these components in the RB form. By contrast, the data presented in this Chapter indicate higher expression levels of the TTSS components in RBs with reduced expression in EBs, a trend that is consistent with the decrease of the TTS-like projections per bacterium, observed during the transition from RB to EB. Additionally, the C-ring components, SctQ and SctN were expressed at 162 ±68 and 91 ±11 molecules/cell in RBs, respectively, and 54 ±10 and 45 ±18 molecules/cell in EBs, respectively. The authors postulate that the expression of SctQ and SctN are ramped-up to pre-pack future EBs, but the results in this chapter indicate the expression of these TTS components in RBs at 15 h PI or earlier.

Known as the 'chlamydial paradox', the detectable expression of the TTS-specific genes occurs between 8-12 h PI, indicating the expression of the TTS apparatus mid-cycle. However, the translocation of chlamydial effector proteins are required early after infection, before *de-novo* synthesis of a TTS apparatus (Reviewed by Beeckman and Vanrompay, 2010). Nevertheless, it has been shown in these studies and by others, that EBs possess all the components of a TTS apparatus (Vandahl *et al*., 2001). As such, Fields *et al*. (2003) proposed that EBs have a TTSS which translocate proteins early in the developmental cycle, with the TTS apparatuses being replenished mid-cycle in RBs, providing apparatus for the subsequent EB progeny. Although the data presented in this thesis supports such a hypothesis, the results of Saka *et al*. suggests *de-novo* synthesis of the TTS apparatus late in the developmental cycle, with a progressive reduced TTS capacity during the transformation from early EBs to RBs at 18 h PI. This result is surprising when considering the expression of the TTS-specific genes occurs mid-cycle; and is concomitant with an increased number of type-III like projections; and at a time when the intravacuolar environment is being heavily modified by predicted TTS substrates, such as the Inc proteins. Nevertheless, the data in this thesis does confirm the expression of TTSS components in EBs, albeit at lower levels than RBs, supporting the hypothesis that EBs are pre-loaded with TTSSs (Peters *et al*., 2007). These observed differences between the TTSS expression profiles in RBs, may reflect the stages when RBs were harvested and purified. However, it is clear that further experimental data on the expression levels of these TTSS components at different stages throughout the developmental cycle are required.

5.6.3.2 Energy metabolism

This thesis identified almost all components encoding enzymes for glycolysis, tricarboxylic acid and the pentose phosphate pathway in both RBs and the metabolically inert EBs, suggesting that Chlamydiae are capable of utilizing glucose to generate ATP throughout their developmental cycle. 'Paradoxically', Saka *et*

al. only identify 6 out of the 43 proteins represented in the 'Energy metabolism' category in RBs, but 39 in EBs. The authors suggest that ATP synthesis occurs via glucose catabolism in EBs, but switches to ATP synthesis generated by ion gradients in the RB form. However, the data presented in this Chapter, although consistent with glucose catabolism using pre-existing pathways in EBs, also indicates that RBs also have the functional capacity to generate ATP via the glycolytic pathway, an observation that is supported by the maximal expression of the genes *PK, GAPDH, PGK, ZWF* in RBs (Iliffe-Lee and McClarty, 1999).

The biological insights acquired from these two studies, highlight the complexity of the chlamydial life cycle and underline the need for more proteomic studies, at multiple stages, to further our understanding of chlamydial infection at the molecular level.

CHAPTER 6

GENERAL DISCUSSION

6.0 GENERAL DISCUSSION

6.1 Introduction

The study presented in this thesis has provided a comprehensive proteomic analysis of the two distinctive developmental forms of the obligate intracellular pathogen *C. trachomatis* L2. Using qualitative and quantitative mass spectrometry strategies presented in Chapters 4 and 5 has confirmed the expression of ~72% of the predicted *C. trachomatis* L2 ORFs. Further, using two-dimensional reverse-phase chromatography in combination with MS^E, estimates of the average copies per cell of ~54% of the proteins present in RBs and EBs have been determined. This coverage compares well to other studies in bacteria such as those of the spirochete *Leptospira interrogans*, where estimates of protein copies per cell were obtained for 51% of the predicted proteome (Malmstrom *et al*., 2009). However, it should be noted that the significant proteome coverage achieved is not solely attributable to the separation and mass spectrometric strategies employed.

Importantly, the comprehensive proteome analysis of an intracellular pathogen requires the purification of the pathogen from the host cell with high purity to avoid the masking of bacterium-derived peptides by host cell components. The separation of the spore-like elementary body of *Chlamydia* from host cells, although not without it's difficulties, can be achieved without lysis of the bacterial cell. However, the purification of osmotically fragile RBs has proven more challenging and has previously limited the proteomic analysis of RBs in a purified form. In collaboration with Professor Ian Clarke (Microbiology, Southampton General Hospital), the development of a purification strategy to obtain RBs in high purity with a low level of host cell components has been pivotal to achieving the proteome coverage presented in this thesis. However, despite the relatively extensive coverage, ~28% of the predicted proteome of *C. trachomatis* L2 is as yet undetected. A key question therefore still remains as to why it has not yet been possible to identify these missing proteins. There are likely to be a number of possible explanations, but does inspection of the proteins forming this unidentified dataset allow one to rule out some of these possibilities? Some of these possibilities are discussed below.

6.1.1 Pseudogenes

Non-functional genes or pseudogenes present a problem in defining the limits of an expressed proteome simply because their prediction can be difficult (Rouchka and Cha, 2009). Pseudogenes are thought to originate through the same mechanisms as normal protein-coding genes, but have become non-functional through the accumulation of disabling mutations such as deletions, stop-codons and frameshifts. The frequency of these pseudogenes usually depends upon the rates of gene duplication and loss, but typically, the number of pseudogenes present in bacteria is relatively low, i.e. the majority of bacterial genomes encode for ~90% protein and structural RNAs. Nonetheless, in some instances it has been shown that bacteria can posses high numbers of pseudogenes, e.g. approximately 25% of the *Rickettsia prowazekii* genome is non-coding for proteins (Anderson *et al*., 1998). The intracellular lifestyle of intracellular pathogens such as the

chlamydiae to some extent provides a protective environment from mobile genetic elements such as bacteriophage and transposons. Chlamydiae show a remarkable conservation in gene content and gene order; and their small genome sizes reflect evolutionary gene loss and genome streamlining (Thomson *et al*., 2008). As a consequence, in comparison to free-living bacteria, they possess relatively small numbers of pseudogenes (Moran and Wernegreen, 2000). The availability of the genome sequence of *Chlamydia trachomatis* L2/434/Bu has allowed whole genome comparisons between L2 and other members of Chlamydiaceae. Consistent with previous findings, they are very similar in genome size, gene content and gene order (Thomson *et al*., 2008). Although there was no apparent gene acquisition, gene loss and small mutations were a defining characteristic that may explain the differences in host adaptation and tissue tropism observed for this LGV strain.

The major region of variation between coding sequences of the different chlamydial species is the region termed the plasticity zone (PZ). This is indicated in the circular representation of the *C. trachomatis* L2 genome represented in **Figure 6.1**. This variant region is principally due to the loss of the cytotoxin gene(s), which have almost entirely been deleted from *C. trachomatis* L2 leaving two remnants CTL0420 and CTL0421 (Belland *et al*., 2001); and 4 encoded phospholipase genes, CTL0409, CTL0411, CTL0413 and CTL0414. However, although CTL0409 and CTL0414 have acquired multiple frameshift mutations and deletions, CTL0411 and CTL0413 appear intact in UW-3 (serovar D), Har-13 (serovar A) and L2. Visualising the peptides from RBs and EBs identified in Chapter 5 mapping to the L2 genome, there is a striking absence of peptides detected across the PZ. This would be expected for the predicted pseudogenes and validates their assignment as pseudogenes. However, the clear and notable absence of peptides mapping to other predicted 'functional' genes able to encode proteins, such as the phospholipase genes, CTL0411 and CTL0413, spanning this same region, does raise the question whether these genes are expressed or whether they too are non-coding either at the level of transcription or translation. Additional pseudogenes were also assigned outside the PZ and these are shown in **Table 6.1** along with those assigned in the PZ.

Of the proteins not identified in this study (**Figure 6.2**), the hypothetical proteins, or proteins of unknown function are quite notable, representing >10% of the predicted *C. trachomatis* L2 genome. By comparison the remaining (~19%) unidentified proteins were fairly evenly distributed across the remaining 13 functional categories. Interestingly, in a previous study focused on improving psuedogene assignment, using 11 genomes from 4 bacterial genera, the number of pseudogenes ranged from 27 in *Staphylococcus aureus* MW2 to 337 in *Yersinia pestis* CO92. Over half of these pseudogenes identified were previously annotated as 'hypothetical' (Lerat and Ochman, 2005). Considering the high representation of 'hypotheticals' within this dataset, could some of these also represent unassigned pseudogenes; or are they characteristically atypical preventing their detection; or are they simply not expressed under the conditions of measurement? Whatever the reason, it is clear that accurate prediction of pseudogenes is required in defining our understanding of what represents a complete proteome.

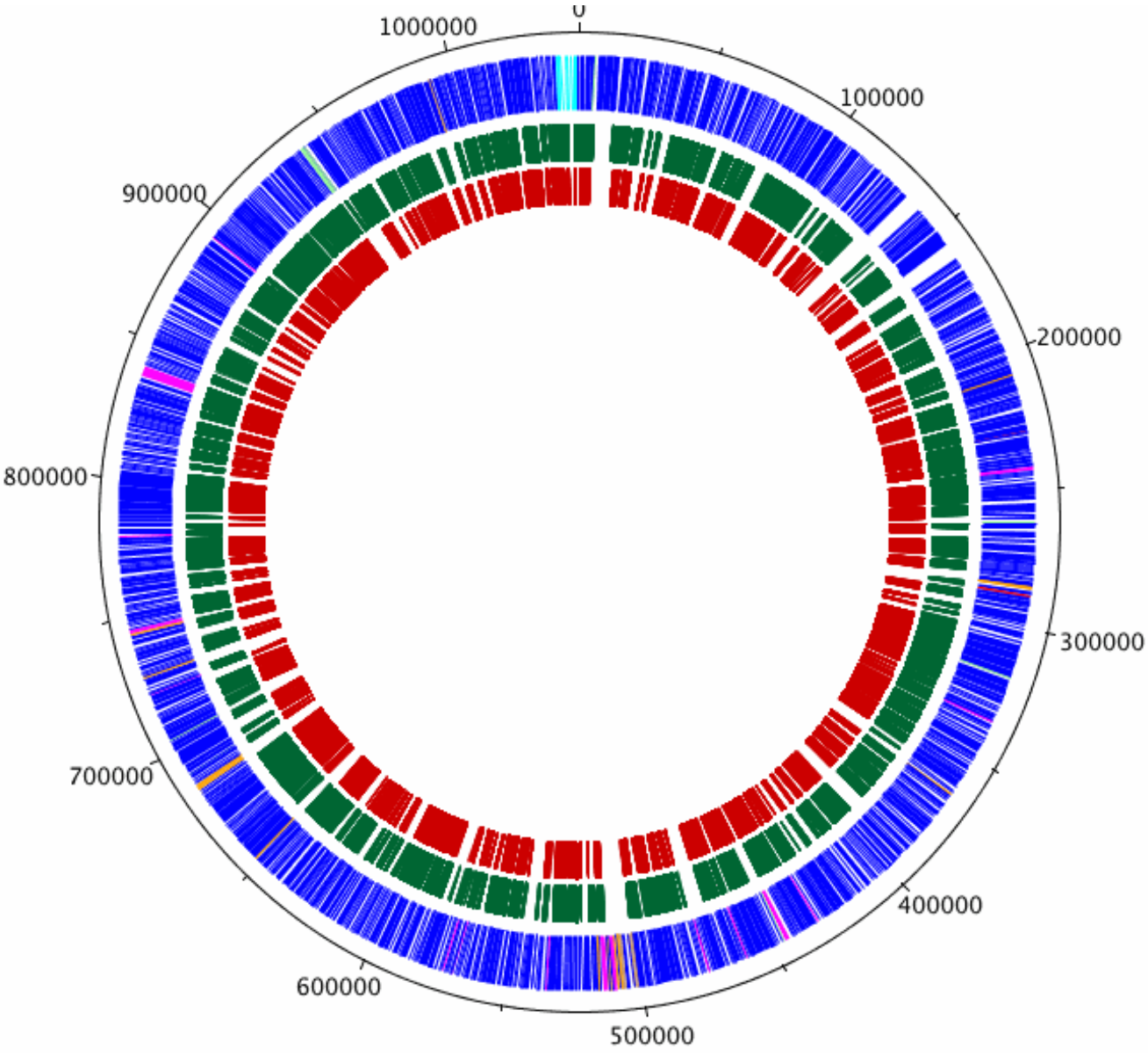


Figure 6.1. Circular representation of the *C. trachomatis* L2 chromosome and the mapping of peptides assigned from both EBs and RBs to their corresponding CDS. The outer scale shows the size in bp. From the outside in, circle 1 shows the position of the CDS. Circles 2 (green) and 3 (red) indicate the CDSs of the peptides assigned in RBs and EBs respectively. Using the published gene predictions for *C. trachomatis* strains UW-3 and Har-3, the strain L2 CDSs have been colour coded depending on whether the are: (blue) predicted and intact in all isolates; (pink) predicted and intact in L2 and UW-3; (green) predicted and intact in L2 and Har-13; (orange) defunct in L2, predicted and intact in Har-13 and UW-3; (red) unique to L2; (brown) defunct in all isolates. The region spanning the plasticity zone (PZ) is indicated.

Table 6.1. Pseudogenes in *C. trachomatis* L2/434/Bu identified by whole genome comparisons with *C. trachomatis* strains UW-3 (serovar D) and Har-13 (serovar A).

L2 *locus*	Protein description
CTL0161	Conserved hypothetical protein
CTL0228	Fumarate hydratase (FumC)
CTL0292	Conserved hypothetical protein
CTL0409	Phospholipase D protein
CTL0414	Phospholipase D protein
CTL0415	Conserved hypothetical protein
CTL0418	Putative membrane protein
CTL0420	Cytotoxin (adherence)
CTL0421	Cytotoxin (adherence)
CTL0426A	Conserved hypothetical protein
CTL0552	Putative integral membrane protein
CTL0578	Conserved hypothetical protein
CTL0612	Inner membrane protein
CTL0627	Pyruvoyl-dependent arginine decarboxylase
CTL0856	Succinate dehydrogenase (sdhC)

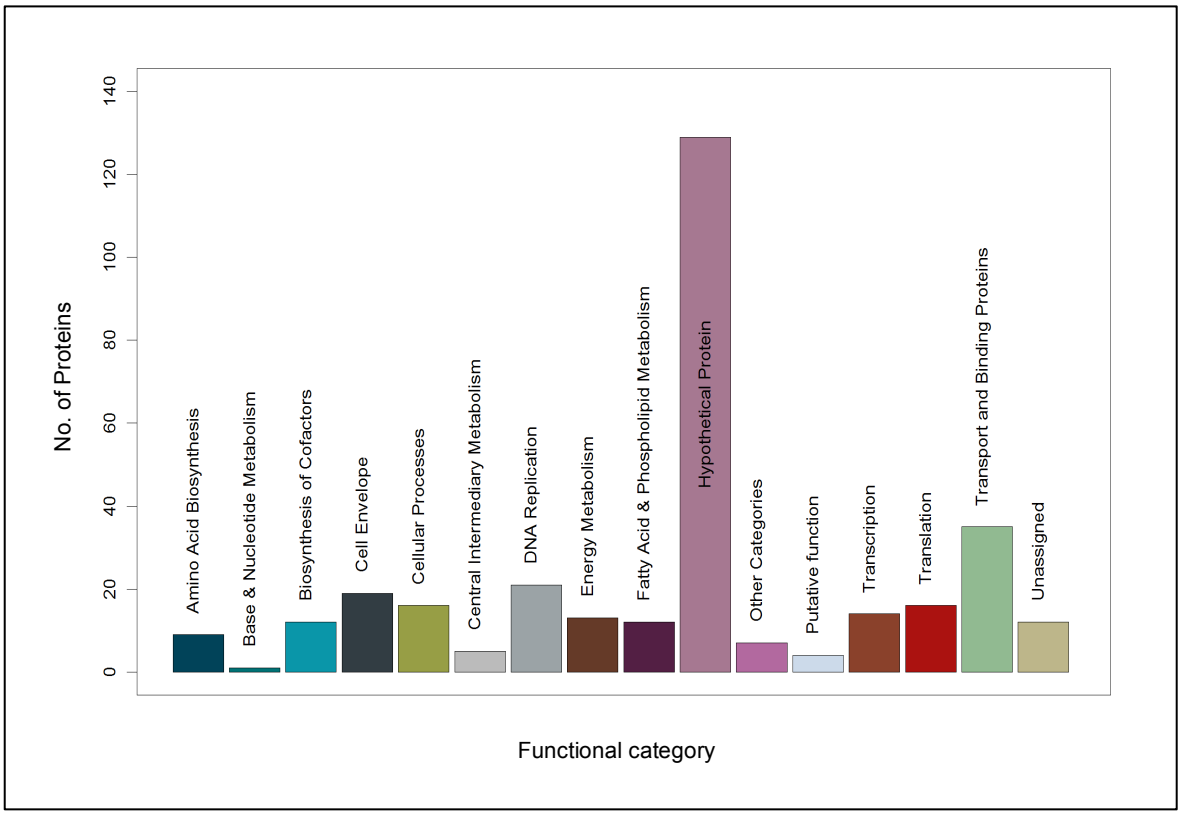


Figure 6.2. *C. trachomatis* L2 proteins not yet identified in this study distributed according to functional category.

6.1.2 Physio-chemical properties of the non-detected proteins

Efforts to increase proteome coverage have focused on the development of innovative fractionation technologies and improved methods of peptide assignment. However, irrespective of such improvements, parts of the proteome have proven to be refractory to detection using the applied technology, either because of their physio-chemical properties and/or their low abundance.

At first inspection of the data presented in Chapter 5, the molecular weight, isoelectric point and hydrophobicity profiles of the proteins unidentified, do not appear atypical of those already detected (**Figure 6.3**), encompassing molecular weights from 5.2 to 196.7 kDa, isoelectric pH values ranging from 3.93 to 12.65 and hydrophobicity values (GRAVY) from -1.45 to +1.31. However, closer inspection of these data, show some distinct sampling bias. The molecular masses of proteins < 8 kDa (**Figure 6.4**) represents ~6% of the proteins not detected, compared to only ~0.6% of proteins > 8 kDa identified, suggesting a possible sampling issue of proteins in this mass range.

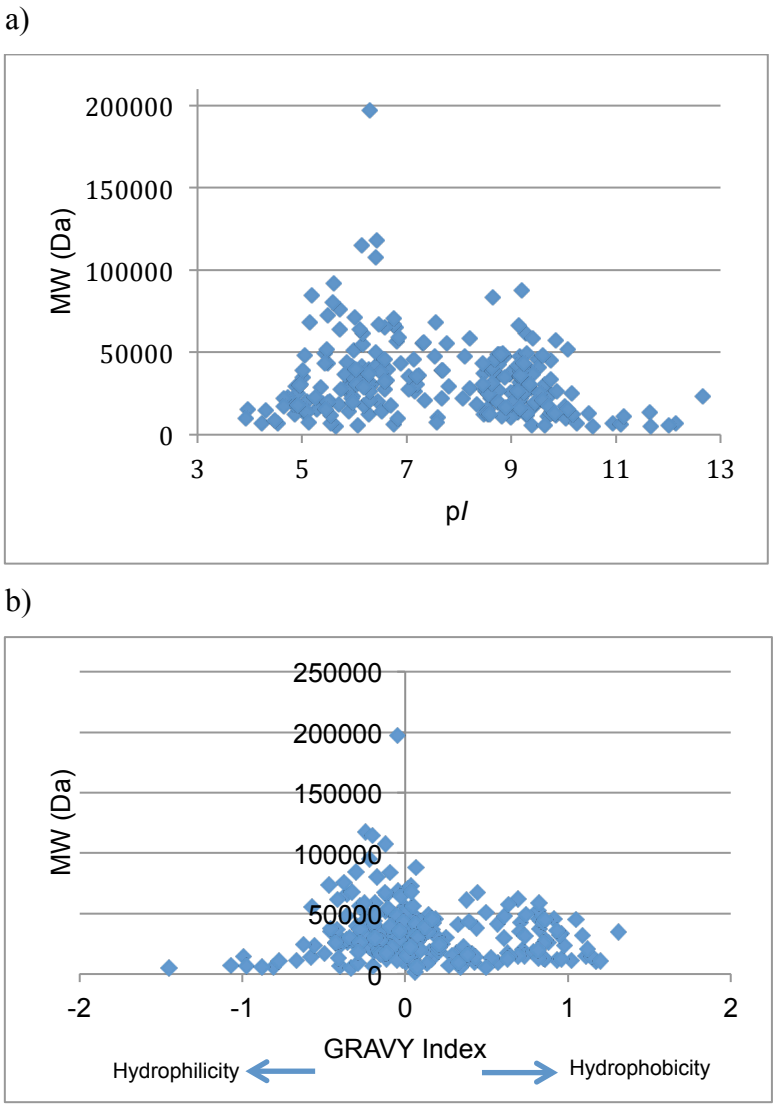


Figure 6.3. Distribution of proteins unidentified in accordance with their a) molecular weight vs. isoelectric point (pI) and b) their molecular weight vs. Gravy index.

High-throughput proteomics is currently reliant on 'bottom-up approaches', where proteins must be reduced to more manageably- sized peptides that can be ionised and measured in the mass spectrometer. This is usually achieved through chemical or more often than not, proteolytic digestion using a site-specific protease. In this study, the protease trypsin was employed. **Figure 6.5** shows the proportion of proteins according to the theoretical number of tryptic peptides they contain for both the identified and unidentified proteins in the *C. trachomatis* proteome. The data points to a clear underrepresentation of small proteins containing less than ~10 tryptic peptides. This is probably not surprising, since the likelihood of identifying a protein increases with the number of peptides generated from a protein, which tends to be low for small proteins and are therefore likely to be underrepresented independent of the method used for analysis. These low molecular weight proteins also typically exhibit low levels of expression, reducing their detection (see *section 6.3*, Low abundance and expression). Since the validation of the theoretical proteome and understanding an organism at the systems level requires the analysis of all protein classes, methods for enriching these low molecular weight proteins will be essential to incorporate into the current experimental set-up. This may include using molecular weight membrane filtration devices, gel-filtration and employing alternative proteases or even digestion using a combination of different proteases. More recently, an alternative 'top-down' strategy has been used to characterise both low and high molecular weight proteins on a proteome-wide scale (Tran *et al*., 2011). Here, intact proteins are separated and identified by subjecting intact protein molecular ions to gas-phase fragmentation and subsequently searching/interpreting their fragmentation spectra. However, as a consequence of the highly complex spectra obtained from multiply charged product ion species, very effective protein separation and costly high-resolution mass spectrometers (i.e., FT-ICR) are required. As such, the uptake of 'top-down' strategies has lagged behind 'bottom-up' proteomic approaches. Nontheless, using such an approach for the identification of low molecular weight proteins or large peptide fragments (middle-down proteomics), such as for those not detected in this study, could be a very attractive alternative identification strategy.

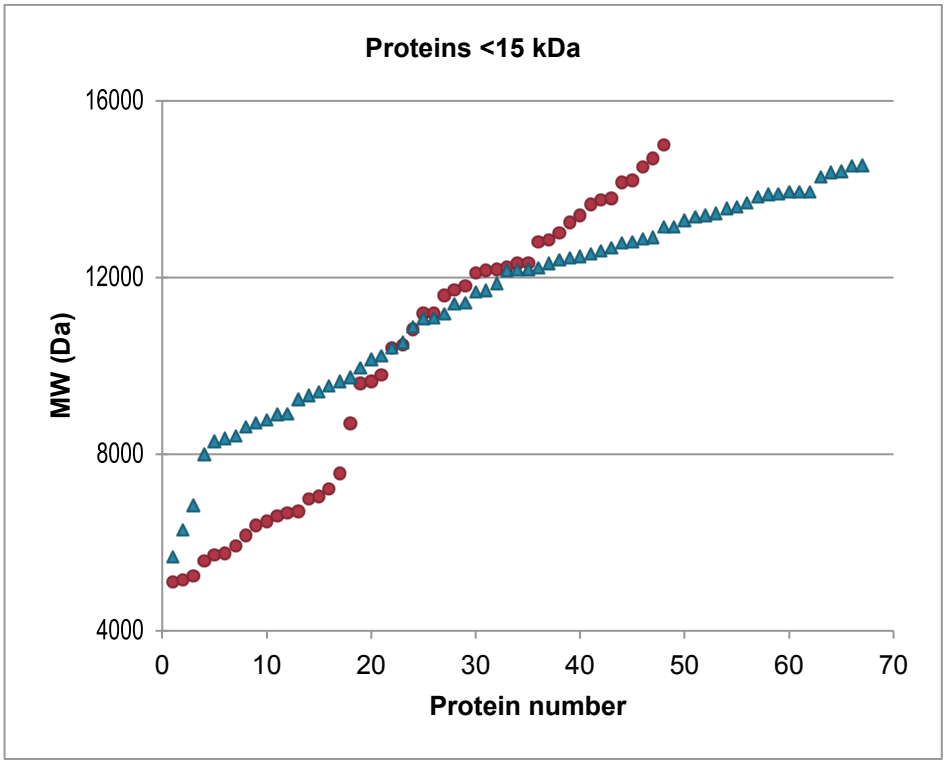


Figure 6.4. Molecular weight profile of proteins identified (blue) and unidentified (red) below 15 kDa.

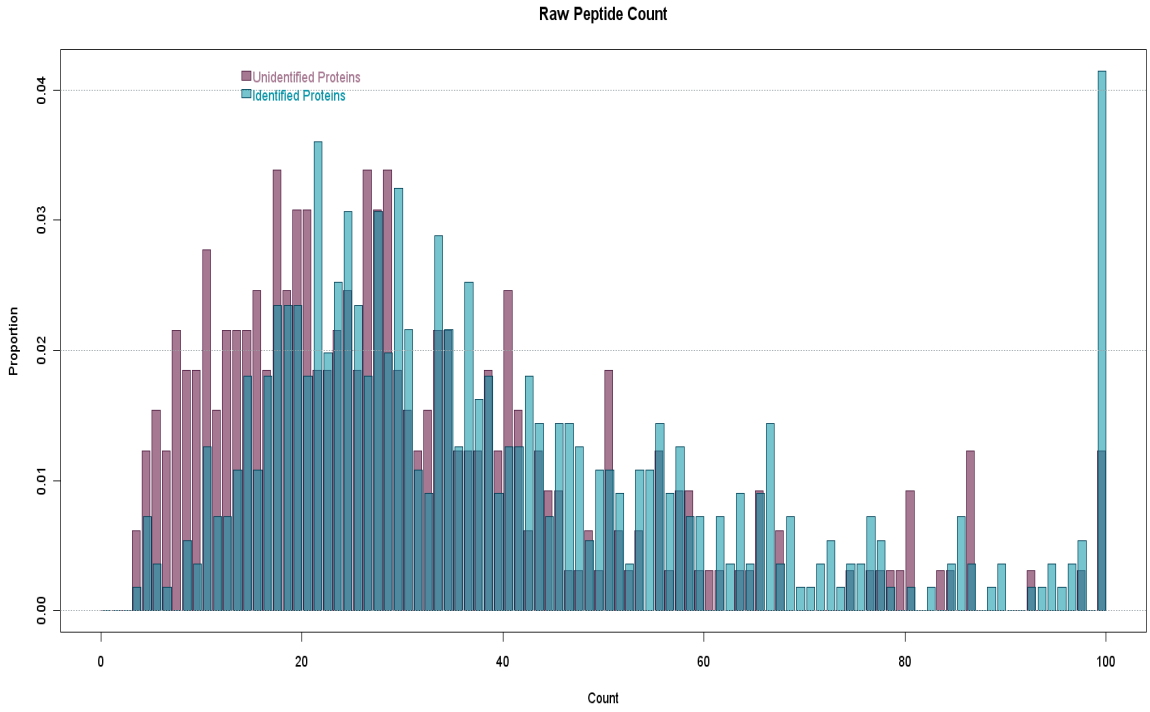


Figure 6.5. Histogram representing the proportion of proteins in the *C. trachomatis* predicted proteome according to their theoretical tryptic peptide count for both identified and unidentified proteins.

6.1.3 Low abundance and expression

The detection of proteins at low cellular concentrations is challenging and presents a major hurdle in the goal of obtaining complete proteome coverage. However, they represent a highly desirable set of proteins, not only from the perspective of understanding their role in chlamydial biology, but also their potential as therapeutic targets. With low cellular concentrations, they offer the possibility of modulating protein/enzyme activity at low therapeutic doses, improving drug efficacy and reducing toxicity dose effects.

Optimization of environmental conditions to ensure that all encoded proteins are expressed under the conditions of measurement is also essential. For many microorganisms, such as *E. coli*, complete proteome coverage has been hampered by the lack of a defined set of *in-vitro* conditions that allows the expression of all of its encoded genes (Chang *et al*., 2004; Conway and Schoolnik, 2003; Tao *et al*., 1999). However, for the chamydiae, transcriptomic studies would appear to suggest that complete proteome coverage is achievable, with the expression of almost every chromosomal and plasmid gene mid-cycle, with the exception of a small subset of 28 genes that are expressed late in the developmental cycle (Belland *et al*., 2003).

The detection limits of the label-free quantitation used in this study provided lower limits of quantification at ~ 10-20 copies per cell, based upon the assumption that genome number is equivalent to the number of cells. Although the copy number of these unidentified proteins may be below this detection limit, there may also be alternative reasons for the absence of these proteins in these data.

One possible explanation is the secretion of these proteins into the host cell cytoplasm. Detailed understanding of how and what proteins are secreted remains unclear, but may include the release of pre-existing proteins from EBs to facilitate chlamydial entry and the secretion of proteins from RBs to support chlamydial development and modulate host immune responses. For example, characterised by a bilobed hydrophobic motif, there are 46 predicted candidate inclusion membrane proteins of which, at least 10 have been reported as being secreted and localised to the inclusion membrane (Bannantine *et al*., 2000; Bannantine *et al*., 1998a; Bannantine *et al*., 1998b; Rockey *et al*., 2002). Additionally, two proteins lacking this motif have also been shown to be associated with the inclusion membrane implying there may be many more exported (Fields and Hackstadt, 2000; Fling *et al*., 2001). In this study, the Inc proteins A,C,E and G were detected, but in addition to these annotated proteins, six additional putative inclusion membrane proteins, CTL0466, CTL0476, CTL0478, CTL0480, CTL0481 and CTL0540 were also detected. These were all detected in RBs, but all showed either a decrease or were below the limit of detection in EBs. Although these proteins may be down regulated via cellular protein turnover, the data could also support their predicted secretion. The lack of detection of the remaining 40 or more predicted inclusion membrane proteins in either RBs (15 h PI) or EBs (48 h PI) may indicate that these proteins are required for early inclusion membrane development and are therefore secreted prior to the sampling point at 15 h PI. The undefined secretome of

Chlamydia clearly represents a significant proportion of the ~270 yet undetected proteins and will therefore require effective methodologies to allow their detection.

6.1.4 Detection of unidentified proteins

There is a requirement for a generic, highly sensitive strategy for the detection and accurate quantitation of proteins that are not easily detected using high-throughput strategies, such as low abundance proteins or those present in high protein backgrounds masking their measurement. As discussed and to some extent developed in Chapter 3, the AQUA approach proposed by Gerber *et al*. (2003) offers a highly sensitive targeted mass spectrometric methodology, that is able to discriminate between one protein and another in highly complex mixtures and provide accurate measurements of quantification. By comparison to high-throughput strategies such as the label-free approach presented in Chapter 5, the throughput is relatively low and requires the synthesis of isotopically labelled peptides corresponding to each peptide measured. However, the approach is highly sensitive and offers excellent quantitative precision. Use of approaches such as AQUA can complement high-throughput proteomic strategies, by providing highly sensitive targeted analyses for the detection and quantitation of proteins not identified using shotgun approaches. However, this is of course dependent upon whether these proteins are expressed under the conditions of measurement. These approaches also offer the opportunity to implement targeted high-throughput assays for the measurement of cellular concentrations of proteins involved in specific pathways and/or processes with high sensitivity. A valuable resource generated as part of these studies, is a database of proteotypic peptide data for the proteins of *C. trachomatis*, providing information on the *m/z* of the observed peptides, their charge state, corresponding fragment ions, collision energy and chromatographic retention times. Utilising this information, methodologies can be devised, based upon these recorded peptide transitions, to provide optimal measurement of specific sub-sets of proteins. Because of the specificity and sensitivity of such assays, the levels of proteins can potentially be measured in highly complex backgrounds such as a host cell lysate. For example, by using several peptide transitions observed for each glycolytic enzyme of chlamydiae, quantitative measurement of this pathway could be made from the initial stages of infection, where cellular concentrations are likely to be low due to the limited number of chlamydiae, through progressive time points in the developmental cycle without the requirement of purifying chlamydiae from the host cell. In the absence of a genetic system for *C. trachomatis*, proteomics has been an essential tool for advancing our understanding of *Chlamydia* at the molecular level. The combination of genetic transformation and proteomic technologies offers the potential to advance our molecular understanding of Chlamydiae even further.

6.2 Future Work

Although major advances in characterising the proteome of *C. trachomatis* L2 have been achieved, the assignment of yet undetected proteins and validation experiments falling outside the time limits and scope of this study would ideally form the basis of future studies. Three key future experiments are described briefly below:

6.2.1 Validation of > 5 proteins using MRM technology

The AQUA (or targeted MRM) approach is currently considered to be the 'gold standard' for the determination of the absolute concentration of a protein. To provide validation and further confidence in the absolute quantitation measurements made for the proteins of *C. trachomatis* using the high-throughput label-free technology in this study, a panel of proteins (>5) should be validated using the AQUA approach. This selected panel of proteins should consist of proteins spanning the dynamic range of the measurements obtained in the high-throughput experiments and represent proteins from different functional categories.

6.2.2 Absolute measurement of specific pathways during the developmental cycle

Using the proteotypic peptide data already obtained, MRM methodologies for specific sub-sets of proteins, i.e., glycolytic pathway enzymes, peptidoglycan pathway biosynthetic enzymes, plasmid proteins, Pmps, etc, could be developed. These methodologies can subsequently be applied to the routine measurement of these protein sub-sets at different stages in the developmental cycle (including the initial stages, where there is an absence of proteomic studies because of the limited numbers of developmental forms), under different treatments and even in cell lysates containing both chlamydiae and host cell components, eliminating the requirement for purification of RBs and EBs.

6.2.3 Enrichment of proteins associated with the inclusion membrane.

As discussed above, screening of predicted genome sequences of *C. trachomatis* and *C. pneumoniae* revealed >40 sequences that contain the characteristic bi-lobed hydrophobic domain that is unique to the inclusion membrane proteins (Bannantine *et al*., 2000). It has also been shown that at least ten of these proteins that share this secondary motif are associated with the inclusion membrane. However, not all proteins containing this motif are necessarily localised to the inclusion membrane (Fields and Hackstadt, 2000; Fling *et al*., 2001). Experimental confirmation of which inc proteins and other proteins are localised to the inclusion membrane would be a major advance, helping to inform on the interactions with the host cell and environmental sensing occurring along the inclusion membrane. To achieve this, some form of enrichment and purification strategy is required to isolate the inclusion membrane from host and pathogen cellular components so that proteins associated with the inclusion membrane can be identified. However, the isolation of such structures is challenging.

An elegant approach that uses iTRAQ in combination with analytical density centrifugation described by Dunkley et al. (2004), allows the simultaneous assignment of proteins to multiple organelles and even sub-cellular compartments within organelles such as the golgi, plasma membrane, mitochondrion, ER and vacuolar localisation (Sadowski *et al*., 2008). This technology termed LOPIT, or Localisation of organelle proteins by isotopic tagging has been used to show the distribution of proteins associated with a specific organelle. Here, cells are disrupted and their contents layered and centrifuged using self-forming iodixanol density gradients to separate out the multiple organelles and subcellular compartments. Fractions from these density gradients are extracted, digested into peptides and each fraction labelled with one of the eight

possible iTRAQ tags. Once labelled, the fractions are pooled and analysed using LC-MS/MS to obtain the identities and an abundance profile of every assigned protein from each fractionation position. Using multivariate statistical techniques, such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), proteins are clustered according to similarities in their protein abundance profiles. By comparing the abundance profiles of proteins known to be associated with a particular organelle, other proteins associated with that organelle or sub-cellular compartment can be assigned.

The proposed experiment would use this strategy to isolate and assign proteins associated with the vacuolar inclusion membrane. However, the success of this experiment would depend on the assumption that the intracellular membrane would exist after lysis as a distinct structure that either reforms as a vesicle, or is characteristically distinct to allow separation from contaminating components of other host and pathogen organelles, etc. If this criterion is met, by using proteins already known to be associated with the inclusion membrane as location markers (e.g., IncA), new proteins interacting with the inclusion membrane could be assigned. This could provide an important tool for elucidating proteins associated with the inclusion membrane that are likely to play a crucial role in the molecular and cellular interactions that facilitate growth and inclusion development. Equally, it could also potentially provide important information on immunoprotection, translating into potential vaccine targets.

6.3 Conclusion

The study of *C. trachomatis* is challenging because of its obligate intracellular nature and until recently, the absence of a genetic system for manipulation. As such, alternative molecular techniques such as genomics, transcriptomics and proteomics remain essential tools in advancing the field of Chlamydial research. This thesis has provided a deep insight into the proteome of the Lymphogranuloma Venereum strain of *C. trachomatis* L2, a serovar that is distinctive in both tissue tropism and pathogenicity from the other genital and ocular strains. The study has focused on the development, implementation and application of proteomic methodologies to the analysis of the two unique stages of its life cycle, to validate the accuracy of the predicted proteome.

Notably, a large proportion of the detected proteome was expressed at increased cellular concentrations in the replicative form, reflecting a time of greatest metabolic activity. Challenging the original dogma that chlamydiae are solely energy parasites, the expression of a complete glycolytic pathway in both forms suggest they are capable of synthesising their own intracellular ATP, not only in metabolically active RBs, but also in extracellular EBs and even possibly in the absence of a host cell. Despite the absence of any detectable cell wall material, the detection of enzymes for an almost complete peptidoglycan biosynthesis pathway in RBs suggests an alternative role in replication and/or host cell modulation.

The secretome of chlamydiae still remains uncharacterised, however, the expression of the type III translocation apparatus in both developmental stages supports the translocation of proteins into the host cell cytoplasm that may play a role in regulating the developmental cycle. As such, identifying these secreted proteins, their interacting partners and their localisation will be key in understanding the importance and relationship of signalling between chlamydiae and the host cell.

By contrast, the extracellular stage has a decreased expression of a large proportion of the detected *C. trachomatis* proteome. Strikingly, against this background of reduced protein synthesis, *C. trachomatis* expends a considerable amount of energy maintaining the outer cell envelope, protein translational machinery and a set of hypothetical proteins. The energy expended in maintaining this functional protein translational system implies that this is important for survival and that EBs are primed and ready-to-go upon infection and/or even maintain a low level of protein synthesis during this extracellular stage, challenging the dogma of the quiescent infectious bacterial cell. Further support for this hypothesis was recently observed in the EBs of the related organism *Protochlamydia amoebophila*, where almost a complete set of ribosomal proteins was detected (Sixt *et al*., 2011) and is further supported by the demonstration of extended metabolic activity in extracellular EBs after release from the host cell using Confocal Raman microspectroscopy (Haider *et al*., 2010). The function of the 'expressed hypothetical proteins' are unknown, but their expression at these late stages are likely to be important in either the formation of EBs; survival in the extracellular stage of it's lifecycle or for initiating infection.

One important legacy of this study is the generation of a database containing proteotypic peptide data for the proteome of *C. trachomatis* L2. This information provides a platform that assists both validation and future targeted quantitative proteome studies. It will facilitate the measurement of cellular concentrations of proteins through the developmental cycle. The chlamydial response to perturbations, such as immune challenges, environmental stimuli, therapeutics and in different cellular states such as persistence can be measured against such a database.

REFERENCES

AbdelRahman, Y.M., and Belland, R.J. (2005). The chlamydial developmental cycle. FEMS Microbiology Reviews ***29***, 949-959.

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature ***422***, 198-207.

Aitken, I.D., Clarkson, M.J., and Linklater, K. (1990). Enzootic abortion of ewes. Veterinary Record ***126***, 136-138.

Albrethsen, J., Knol, J.C., Piersma, S.R., Pham, T.V., de Wit, M., Mongera, S., Carvalho, B., Verheul, H.M., Fijneman, R.J., Meijer, G.A., *et al.* (2010). Subnuclear proteomics in colorectal cancer: identification of proteins enriched in the nuclear matrix fraction and regulation in adenoma to carcinoma progression. Molecular & Cellular proteomics ***9***, 988-1005.

Almeida, J.L., Hill, C.R., and Cole, K.D. (2011). Authentication of African green monkey cell lines using human short tandem repeat markers. BMC biotechnology ***11***, 102.

Anderson, L., and Seilhamer, J. (1997). A comparison of selected mRNA and protein abundances in human liver. Electrophoresis ***18***, 533-537.

Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H., and Kurland, C.G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature ***396***, 133-140.

Andren, P.E., Emmett, M.R., and Caprioli, R.M. (1994). Micro-Electrospray - Zeptomole-Attomole Per Microliter Sensitivity for Peptides. Journal of the American Society for Mass Spectrometry ***5***, 867-869.

Ansong, C., Purvine, S.O., Adkins, J.N., Lipton, M.S., and Smith, R.D. (2008). Proteogenomics: needs and roles to be filled by proteomics in genome annotation. Briefings in functional genomics & proteomics *7*, 50-62.

Arcari, C.M., Gaydos, C.A., Nieto, F.J., Krauss, M., and Nelson, K.E. (2005). Association between *Chlamydia pneumoniae* and Acute Myocardial Infarction in Young Men in the United States Military: The Importance of Timing of Exposure Measurement. Clinical Infectious Diseases ***40***, 1123-1130.

Azzouzi, N., Elhataoui, M., Bakhatar, A., Takourt, B., and Benslimane, A. (2005). [Part of *Chlamydia pneumoniae* in atherosclerosis and exacerbation of chronic obstructive pulmonary disease and asthma]. Annales de Biologie Clinique ***63***, 179-184.

Bannantine, Griffiths, Viratyosin, Brown, and Rockey (2000). A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane. Cellular microbiology ***2***, 35-47.

Bannantine, J.P., Rockey, D.D., and Hackstadt, T. (1998a). Tandem genes of *Chlamydia psittaci* that encode proteins localized to the inclusion membrane. Molecular Microbiology ***28***, 1017-1026.

Bannantine, J.P., Stamm, W.E., Suchland, R.J., and Rockey, D.D. (1998b). *Chlamydia trachomatis* IncA is localized to the inclusion membrane and is recognized by antisera from infected humans and primates. Infection and Immunity ***66***, 6017-6021.

Barbour, A.G., Amano, K., Hackstadt, T., Perry, L., and Caldwell, H.D. (1982). *Chlamydia trachomatis* has penicillin-binding proteins but not detectable muramic acid. Journal of Bacteriology ***151***, 420-428.

Barr, J.R., Maggio, V.L., Patterson, D.G., Jr., Cooper, G.R., Henderson, L.O., Turner, W.E., Smith, S.J., Hannon, W.H., Needham, L.L., and Sampson, E.J. (1996). Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. Clinical Chemistry ***42***, 1676-1682.

Barrett, S., and Taylor, C. (2005). A review on pelvic inflammatory disease. International Journal of STD and AIDS ***16***, 715-721.

Barron, A.L.e. (1988). Microbiology *of Chlamydia* (CRC press).

Barry, C.E., III, Brickman, T.J., and Hackstadt, T. (1993). Hc1-mediated effects on DNA structure: a potential regulator of chlamydial development. Molecular Microbiology ***9***, 273-283.

Batteiger, B.E. (1996). The major outer membrane protein of a single *Chlamydia trachomatis* serovar can possess more than one serovar-specific epitope. Infection and Immunity ***64***, 542-547.

Batteiger, B.E., Fraiz, J., Newhall, W.J., Katz, B.P., and Jones, R.B. (1989). Association of recurrent chlamydial infection with gonorrhea. Journal of Infectious Diseases ***159***, 661-669.

Batteiger, B.E., Newhall, W.J., and Jones, R.B. (1985). Differences in outer membrane proteins of the lymphogranuloma venereum and trachoma biovars of *Chlamydia trachomatis*. Infection and Immunity ***50***, 488-494.

Bavoil, P.M., Hsia, R., and Ojcius, D.M. (2000). Closing in on *Chlamydia* and its intracellular bag of tricks. Microbiology ***146 (Pt 11)***, 2723-2731.

Bavoil, P.M., and Hsia, R.C. (1998). Type III secretion in *Chlamydia*: a case of deja vu? Molecular Microbiology ***28***, 860-862.

Bax, C.J., Mutsaers, J.A., Jansen, C.L., Trimbos, J.B., Dorr, P.J., and Oostvogel, P.M. (2003). Comparison of serological assays for detection of *Chlamydia trachomatis* antibodies in different groups of obstetrical and gynecological patients. Clinical and Diagnostic Laboratory Immunology ***10***, 174-176.

Bax, C.J., Quint, K.D., Peters, R.P.H., Ouburg, S., Oostvogel, P.M., Mutsaers, J.A.E.M., D√∂rr, P.J., Schmidt, S., Jansen, C., van Leeuwen, A.P.*, et al.* (2011). Analyses of multiple-site and concurrent *Chlamydia trachomatis* serovar infections, and serovar tissue tropism for urogenital versus rectal specimens in male and female patients. Sexually Transmitted Infections ***87***, 503-507.

Beatty, W.L., Belanger, T.A., Desai, A.A., Morrison, R.P., and Byrne, G.I. (1994b). Tryptophan depletion as a mechanism of gamma interferon-mediated chlamydial persistence. Infection and Immunity ***62***, 3705-3711.

Beatty, W.L., Morrison, R.P., and Byrne, G.I. (1994a). Persistent chlamydiae: from cell culture to a paradigm for chlamydial pathogenesis. Microbiological Reviews ***58***, 686-699.

Beavis, R.C., and Chait, B.T. (1996). Matrix-assisted laser desorption ionization mass-spectrometry of proteins. Methods in Enzymology ***270***, 519-551.

Beeckman, D.S., and Vanrompay, D.C. (2010). Bacterial secretion systems with an emphasis on the chlamydial Type III secretion system. Current issues in molecular biology ***12***, 17-41.

Beem, M.O., and Saxon, E.M. (1977). Respiratory-tract colonization and a distinctive pneumonia syndrome in infants infected with *Chlamydia trachomatis*. New England Journal of Medicine ***296***, 306-310.

Belland, R.J., Ouellette, S.P., Gieffers, J., and Byrne, G.I. (2004). *Chlamydia pneumoniae* and atherosclerosis. Cellular Microbiology ***6***, 117-127.

Belland, R.J., Scidmore, M.A., Crane, D.D., Hogan, D.M., Whitmire, W., McClarty, G., and Caldwell, H.D. (2001). *Chlamydia trachomatis* cytotoxicity associated with complete and partial cytotoxin genes. Proceedings of the National Academy of Sciences of the United States of America ***98***, 13984-13989.

Belland, R.J., Zhong, G., Crane, D.D., Hogan, D., Sturdevant, D., Sharma, J., Beatty, W.L., and Caldwell, H.D. (2003). Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. Proceedings of the National Academy of Sciences of the United States of America ***100***, 8478-8483.

Beynon, R.J., and Pratt, J.M. (2005). Metabolic labeling of proteins for proteomics. Molecular & Cellular proteomics ***4***, 857-872.

Biemann, K. (1988). Contributions of mass spectrometry to peptide and protein structure. Biomedical and Environmental Mass Spectrometry ***16***, 99-111.

Biemann, K., and Papayannopoulos, I.A. (1994). Amino acid sequencing of proteins. Accounts of Chemical Research ***27***, 370-378.

Bini, L., Sanchez-Campillo, M., Santucci, A., Magi, B., Marzocchi, B., Comanducci, M., Christiansen, G., Birkelund, S., Cevenini, R., Vretou, E*., et al.* (1996). Mapping of *Chlamydia trachomatis* proteins by immobiline-polyacrylamide two-dimensional electrophoresis: spot identification by N-terminal sequencing and immunoblotting. Electrophoresis ***17***, 185-190.

Birkelund, S., Lundemose, A.G., and Christiansen, G. (1988). Chemical cross-linking of *Chlamydia trachomatis*. Infection and Immunity ***56***, 654-659.

Birkelund, S., Morgan-Fisher, M., Timmerman, E., Gevaert, K., Shaw, A.C., and Christiansen, G. (2009). Analysis of proteins in *Chlamydia trachomatis* L2 outer membrane complex, COMC. FEMS Immunology and Medical Microbiology ***55***, 187-195.

Bjellqvist, B., Ek, K., Righetti, P.G., Gianazza, E., Gorg, A., Westermeier, R., and Postel, W. (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. Journal of Biochemical and Biophysical Methods ***6***, 317-339.

Bondar, O.P., Barnidge, D.R., Klee, E.W., Davis, B.J., and Klee, G.G. (2007). LC-MS/MS quantification of Zn-alpha2 glycoprotein: a potential serum biomarker for prostate cancer. Clinical Chemistry ***53***, 673-678.

Boneca, I.G., Dussurget, O., Cabanes, D., Nahori, M.A., Sousa, S., Lecuit, M., Psylinakis, E., Bouriotis, V., Hugot, J.P., Giovannini, M*., et al.* (2007). A critical role for peptidoglycan N-deacetylation in Listeria evasion from the host innate immune system. Proceedings of the National Academy of Sciences of the United States of America ***104***, 997-1002.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry ***72***, 248-254.

Brengues, M., Pintard, L., and Lapeyre, B. (2002). mRNA decay is rapidly induced after spore germination of Saccharomyces cerevisiae. Journal of Biological Chemistry ***277***, 40505-40512.

Brenner, S.E. (1999). Errors in genome annotation. Trends in Genetics ***15***, 132-133.

Brickman, T.J., Barry, C.E., III, and Hackstadt, T. (1993). Molecular cloning and expression of hctB encoding a strain-variant chlamydial histone-like protein with DNA-binding activity. The Journal of Bacteriology ***175***, 4274-4281.

Brown, P.O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. Nature Genetics ***21***, 33-37.

Brown, W.J., and Rockey, D.D. (2000). Identification of an antigen localized to an apparent septum within dividing chlamydiae. Infection and Immunity ***68***, 708-715.

Brunham, R.C., Binns, B., McDowell, J., and Paraskevas, M. (1986). *Chlamydia trachomatis* infection in women with ectopic pregnancy. Obstetrics and Gynecology ***67***, 722-726.

Burton, M.J., and Mabey, D.C. (2009). The global burden of trachoma: a review. PLoS neglected tropical diseases ***3***, e460.

Byrne, G.I. (2003). *Chlamydia* uncloaked. Proceedings of the National Academy of Sciences ***100***, 8040-8042.

Caldwell, H.D., Wood, H., Crane, D., Bailey, R., Jones, R.B., Mabey, D., Maclean, I., Mohammed, Z., Peeling, R., Roshick, C*., et al.* (2003). Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. Journal of Clinical Investigation ***111***, 1757-1769.

Campbell, L.A., Lee, A., and Kuo, C.C. (2006). Cleavage of the N-linked oligosaccharide from the surfaces of *Chlamydia* species affects infectivity in the mouse model of lung infection. Infection and Immunity ***74***, 3027-3029.

Carabeo, R.A., Grieshaber, S.S., Fischer, E., and Hackstadt, T. (2002). *Chlamydia trachomatis* induces remodeling of the actin cytoskeleton during attachment and entry into HeLa cells. Infection and Immunity ***70***, 3793-3803.

Carabeo, R.A., and Hackstadt, T. (2001). Isolation and characterization of a mutant Chinese hamster ovary cell line that is resistant to *Chlamydia trachomatis* infection at a novel step in the attachment process. Infection and Immunity ***69***, 5899-5904.

Carlson, J.H., Porcella, S.F., McClarty, G., and Caldwell, H.D. (2005). Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. Infection and Immunity ***73***, 6407-6418.

Centers for Disease, Control and Prevention (2001).
http://www.cdc.gov/drugresistance/healthcare/problem.htm.

Centers for Disease, Control and Prevention (2010). Sexually Transmitted Disease Surveillance, 2010. Atlanta, GA: U.S. Department of Health and Human Services, 2011.

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S., *et al.* (2004). Carbon nutrition of Escherichia coli in the mouse intestine. Proceedings of the National Academy of Sciences of the United States of America ***101***, 7427-7432.

Chen, G., Shaw, M.H., Kim, Y.G., and Nunez, G. (2009). NOD-like receptors: role in innate immunity and inflammatory disease. Annual review of pathology ***4***, 365-398.

Chen, J., Canales, L., and Neal, R.E. (2011). Multi-Segment Direct Inject nano-ESI-LTQ-FT-ICR-MS/MS For Protein Identification. Proteome science ***9***, 38.

Chen, Y., Choong, L.Y., Lin, Q., Philp, R., Wong, C.H., Ang, B.K., Tan, Y.L., Loh, M.C., Hew, C.L., Shah, N., *et al.* (2007). Differential expression of novel tyrosine kinase substrates during breast cancer development. Molecular & Cellular proteomics ***6***, 2072-2087.

Chenau, J., Michelland, S., Sidibe, J., and Seve, M. (2008). Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling. Proteome science ***6***, 9.

Cheng, D., Hoogenraad, C.C., Rush, J., Ramm, E., Schlager, M.A., Duong, D.M., Xu, P., Wijayawardana, S.R., Hanfelt, J., Nakagawa, T., *et al.* (2006). Relative and absolute quantification of postsynaptic density proteome isolated from rat forebrain and cerebellum. Molecular & Cellular proteomics ***5***, 1158-1170.

Cheng, H., Macaluso, M., Vermund, S.H., and Hook, E.W., III (2001). Relative accuracy of nucleic acid amplification tests and culture in detecting *Chlamydia* in asymptomatic men. Journal of Clinical Microbiology ***39***, 3927-3937.

Chiappino, M.L., Dawson, C., Schachter, J., and Nichols, B.A. (1995). Cytochemical localization of glycogen in *Chlamydia trachomatis* inclusions. The Journal of Bacteriology ***177***, 5358-5363.

Choe, L., D'Ascenzo, M., Relkin, N.R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P., and Lee, K.H. (2007). 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. Proteomics ***7***, 3651-3660.

Chong, P.K., Gan, C.S., Pham, T.K., and Wright, P.C. (2006). Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections. Journal of Proteome Research ***5***, 1232-1240.

Choong, L.Y., Lim, S., Chong, P.K., Wong, C.Y., Shah, N., and Lim, Y.P. (2010). Proteome-wide profiling of the MCF10AT breast cancer progression model. PLoS One ***5***, e11030.

Chopra, I., Storey, C., Falla, T.J., and Pearce, J.H. (1998). Antibiotics, peptidoglycan synthesis and genomics: the chlamydial anomaly revisited. Microbiology ***144 (Pt 10)***, 2673-2678.

Christoforidis, S., McBride, H.M., Burgoyne, R.D., and Zerial, M. (1999). The Rab5 effector EEA1 is a core component of endosome docking. Nature ***397***, 621-625.

Clausen, J.D., Christiansen, G., Holst, H.U., and Birkelund, S. (1997). *Chlamydia trachomatis* utilizes the host cell microtubule network during early events of infection. Molecular Microbiology ***25***, 441-449.

Clifton, D.R., Dooley, C.A., Grieshaber, S.S., Carabeo, R.A., Fields, K.A., and Hackstadt, T. (2005). Tyrosine phosphorylation of the chlamydial effector protein Tarp is species specific and not required for recruitment of actin. Infection and Immunity ***73***, 3860-3868.

Clifton, D.R., Fields, K.A., Grieshaber, S.S., Dooley, C.A., Fischer, E.R., Mead, D.J., Carabeo, R.A., and Hackstadt, T. (2004). A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. Proceedings of the National Academy of Sciences of the United States of America ***101***, 10166-10171.

Cocchiaro, J.L., Kumar, Y., Fischer, E.R., Hackstadt, T., and Valdivia, R.H. (2008). Cytoplasmic lipid droplets are translocated into the lumen of the *Chlamydia trachomatis* parasitophorous vacuole. Proceedings of the National Academy of Sciences of the United States of America ***105***, 9379-9384.

Comisarow, M.B., and Marshall, A.G. (1974a). Fourier transform ion cyclotron resonance spectroscopy. Chemical Physics Letters ***25***, 282-283.

Comisarow, M.B., and Marshall, A.G. (1974b). Frequency-sweep fourier transform ion cyclotron resonance spectroscopy. Chemical Physics Letters ***26***, 489-490.

Conrads, T.P., Alving, K., Veenstra, T.D., Belov, M.E., Anderson, G.A., Anderson, D.J., Lipton, M.S., Pasa-Tolic, L., Udseth, H.R., Chrisler, W.B.*, et al.* (2001). Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling. Analytical Chemistry ***73***, 2132-2139.

Conrads, T.P., Anderson, G.A., Veenstra, T.D., Pasa-Tolic, L., and Smith, R.D. (2000). Utility of accurate mass tags for proteome-wide protein identification. Analytical Chemistry ***72***, 3349-3354.

Conway, T., and Schoolnik, G.K. (2003). Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. Molecular Microbiology ***47***, 879-889.

Corkery, L.J., Pang, H., Schneider, B.B., Covey, T.R., and Siu, K.W. (2005). Automated nanospray using chip-based emitters for the quantitative analysis of pharmaceutical compounds. Journal of the American Society for Mass Spectrometry ***16***, 363-369.

Cox, B., Kislinger, T., and Emili, A. (2005). Integrating gene and protein expression data: pattern analysis and profile mining. Methods ***35***, 303-314.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. Journal of Proteome Research ***10***, 1794-1805.

Craig, R., and Beavis, R.C. (2004). TANDEM: matching proteins with tandem mass spectra. Bioinformatics ***20***, 1466-1467.

Darougar, S., Jones, B.R., Kinnison, J.R., Vaughan-Jackson, J.D., and Dunlop, E.M. (1972). Chlamydial infection. Advances in the diagnostic isolation of *Chlamydia*, including TRIC agent, from the eye, genital tract, and rectum. British Journal of Venereal Diseases ***48***, 416-420.

Daugaard, L., Christiansen, G., and Birkelund, S. (2001). Characterization of a hypervariable region in the genome of *Chlamydophila pneumoniae*. FEMS Microbiology Letters ***203***, 241-248.

Davis, C.H., and Wyrick, P.B. (1997). Differences in the association of *Chlamydia trachomatis* serovar E and serovar L2 with epithelial cells in vitro may reflect biological differences in vivo. Infection and Immunity ***65***, 2914-2924.

Davis, M.T., Spahr, C.S., McGinley, M.D., Robinson, J.H., Bures, E.J., Beierle, J., Mort, J., Yu, W., Luethy, R., and Patterson, S.D. (2001). Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. II. Limitations of complex mixture analyses. Proteomics ***1***, 108-117.

Dawson, C.R., Daghfous, T., Messadi, M., Hoshiwara, I., and Schachter, J. (1976). Severe endemic trachoma in Tunisia. British Journal of Ophthalmology ***60***, 245-252.

de Godoy, L.M., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Frohlich, F., Walther, T.C., and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature ***455***, 1251-1254.

DeSouza, L., Diehl, G., Rodrigues, M.J., Guo, J., Romaschin, A.D., Colgan, T.J., and Siu, K.W. (2005). Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. Journal of Proteome Research ***4***, 377-386.

Dethy, J.M., Ackermann, B.L., Delatour, C., Henion, J.D., and Schultz, G.A. (2003). Demonstration of direct bioanalysis of drugs in plasma using nanoelectrospray infusion from a silicon chip coupled with tandem mass spectrometry. Analytical Chemistry ***75***, 805-811.

Dilbeck, P.M., Evermann, J.F., Crawford, T.B., Ward, A.C., Leathers, C.W., Holland, C.J., Mebus, C.A., Logan, L.L., Rurangirwa, F.R., and McGuire, T.C. (1990). Isolation of a previously undescribed rickettsia from an aborted bovine fetus. Journal of Clinical Microbiology ***28***, 814-816.

Dille, B.J., Butzen, C.C., and Birkenmeyer, L.G. (1993). Amplification of *Chlamydia trachomatis* DNA by ligase chain reaction. Journal of Clinical Microbiology ***31***, 729-731.

Dowell, J.A., Frost, D.C., Zhang, J., and Li, L. (2008). Comparison of two-dimensional fractionation techniques for shotgun proteomics. Analytical Chemistry ***80***, 6715-6723.

Dunkley, T.P., Watson, R., Griffin, J.L., Dupree, P., and Lilley, K.S. (2004). Localization of organelle proteins by isotope tagging (LOPIT). Molecular & Cellular proteomics ***3***, 1128-1134.

Ebers, G.M., and Stern, L. (1875). Ebers Papyrus. Facsimile with a partial translation, Berlin.

Edman, P., and Begg, G. (1967). A protein sequenator. European Journal of Biochemistry ***1***, 80-91.

Emmett, M.R., and Caprioli, R.M. (1994). Micro-Electrospray Mass-Spectrometry - Ultra-High-Sensitivity Analysis of Peptides and Proteins. Journal of the American Society for Mass Spectrometry ***5***, 605-613.

Eng, J.K., McCormack, A.L., and Yates, I.I.I. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry ***5***, 976-989.

Eriksson, J., and Fenyo, D. (2002). A model of random mass-matching and its use for automated significance testing in mass spectrometric proteome analysis. Proteomics ***2***, 262-270.

Escalante-Ochoa, C., Ducatelle, R., and Haesebrouck, F. (2000). Optimal development of *Chlamydophila psittaci* in L929 fibroblast and BGM epithelial cells requires the participation of microfilaments and microtubule-motor proteins. Microbial Pathogenesis ***28***, 321-333.

Everett, K.D., Bush, R.M., and Andersen, A.A. (1999). Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standards for the identification of organisms. International Journal of Systematic Bacter_ology ***49 Pt 2***, 415-440.

Eymann, C., Homuth, G., Scharf, C., and Hecker, M. (2002). Bacillus subtilis functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. Journal of Bacteriology ***184***, 2500-2520.

Fehlner-Gardiner, C., Roshick, C., Carlson, J.H., Hughes, S., Belland, R.J., Caldwell, H.D., and McClarty, G. (2002). Molecular basis defining human *Chlamydia trachomatis* tissue tropism. A possible role for tryptophan synthase. Journal of Biological Chemistry ***277***, 26893-26903.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science ***246***, 64-71.

Ferguson, R.E., McCulloh, K.E., and Rosenstock, H.M. (1965). Observation of the Products of Ionic Collision Processes and Ion Decomposition in a Linear, Pulsed Time-of-Flight Mass Spectrometer. The Journal of chemical physics ***42***, 100-106.

Fields, K.A., Fischer, E.R., Mead, D.J., and Hackstadt, T. (2005). Analysis of putative *Chlamydia trachomatis* chaperones Scc2 and Scc3 and their use in the identification of type III secretion substrates. Journal of Bacteriology ***187***, 6466-6478.

Fields, K.A., and Hackstadt, T. (2000). Evidence for the secretion of *Chlamydia trachomatis* CopN by a type III secretion mechanism. Molecular Microbiology ***38***, 1048-1060.

Fields, K.A., Mead, D.J., Dooley, C.A., and Hackstadt, T. (2003). *Chlamydia trachomatis* type III secretion: evidence for a functional apparatus during early-cycle development. Molecular Microbiology ***48***, 671-683.

Finlay, B.B., Ruschkowski, S., and Dedhar, S. (1991). Cytoskeletal rearrangements accompanying salmonella entry into epithelial cells. Journal of Cell Science ***99 (Pt 2)***, 283-296.

Fling, S.P., Sutherland, R.A., Steele, L.N., Hess, B., D'Orazio, S.E., Maisonneuve, J., Lampe, M.F., Probst, P., and Starnbach, M.N. (2001). CD8+ T cells recognize an inclusion membrane-associated protein from the

vacuolar pathogen *Chlamydia trachomatis*. Proceedings of the National Academy of Sciences of the United States of America ***98***, 1160-1165.

Fournier, M.L., Gilmore, J.M., Martin-Brown, S.A., and Washburn, M.P. (2007). Multidimensional separations-based shotgun proteomics. Chemical Reviews ***107***, 3654-3686.

Fox, A., Rogers, J.C., Gilbart, J., Morgan, S., Davis, C.H., Knight, S., and Wyrick, P.B. (1990). Muramic acid is not detectable in *Chlamydia psittaci* or *Chlamydia trachomatis* by gas chromatography-mass spectrometry. Infection and Immunity ***58***, 835-837.

Francis, C.L., Jerse, A.E., Kaper, J.B., and Falkow, S. (1991). Characterization of interactions of enteropathogenic Escherichia coli O127:H6 with mammalian cells in vitro. Journal of Infectious Diseases ***164***, 693-703.

Francis, C.L., Ryan, T.A., Jones, B.D., Smith, S.J., and Falkow, S. (1993). Ruffles induced by Salmonella and other stimuli direct macropinocytosis of bacteria. Nature ***364***, 639-642.

Franzel, B., and Wolters, D.A. (2011). Advanced MudPIT as a next step toward high proteome coverage. Proteomics ***11***, 3651-3656.

Fritsche, T.R., Gautom, R.K., Seyedirashti, S., Bergeron, D.L., and Lindquist, T.D. (1993). Occurrence of bacterial endosymbionts in Acanthamoeba spp. isolated from corneal and environmental specimens and contact lenses. Journal of Clinical Microbiology ***31***, 1122-1126.

Fukushi, H., and Hirai, K. (1992). Proposal of *Chlamydia pecorum* sp. nov. for *Chlamydia* strains derived from ruminants. International Journal of Systematic Bacter_ology ***42***, 306-308.

Gan, C.S., Chong, P.K., Pham, T.K., and Wright, P.C. (2007). Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). Journal of Proteome Research ***6***, 821-827.

Garbis, S.D., Roumeliotis, T.I., Tyritzis, S.I., Zorpas, K.M., Pavlakis, K., and Constantinides, C.A. (2011). A novel multidimensional protein identification technology approach combining protein size exclusion prefractionation, peptide zwitterion-ion hydrophilic interaction chromatography, and nano-ultraperformance RP chromatography/nESI-MS2 for the in-depth analysis of the serum proteome and phosphoproteome: application to clinical sera derived from humans with benign prostate hyperplasia. Analytical Chemistry ***83***, 708-718.

Geiger, T., Cox, J., and Mann, M. (2010b). Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. Molecular & Cellular proteomics ***9***, 2252-2261.

Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R., and Mann, M. (2010a). Super-SILAC mix for quantitative proteomics of human tumor tissue. Nature Methods ***7***, 383-385.

Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proceedings of the National Academy of Sciences of the United States of America ***100***, 6940-6945.

Geromanos, S.J., Vissers, J.P., Silva, J.C., Dorschel, C.A., Li, G.Z., Gorenstein, M.V., Bateman, R.H., and Langridge, J.I. (2009). The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. Proteomics ***9***, 1683-1695.

Gevaert, K., Van, D.J., Goethals, M., Thomas, G.R., Hoorelbeke, B., Demol, H., Martens, L., Puype, M., Staes, A., and Vandekerckhove, J. (2002). Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 Escherichia coli proteins. Molecular & Cellular proteomics ***1***, 896-903.

Gijsen, A.P., Land, J.A., Goossens, V.J., Leffers, P., Bruggeman, C.A., and Evers, J.L. (2001). *Chlamydia pneumoniae* and screening for tubal factor subfertility. Human Reproduction ***16***, 487-491.

Giles, K., Williams, J.P., and Campuzano, I. (2011). Enhancements in travelling wave ion mobility resolution. Rapid communications in mass spectrometry : RCM ***25***, 1559-1566.

Girardin, S.E., Boneca, I.G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., Philpott, D.J., and Sansonetti, P.J. (2003). Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. Journal of Biological Chemistry ***278***, 8869-8872.

Gomes, J.P., Nunes, A., Bruno, W.J., Borrego, M.J., Florindo, C., and Dean, D. (2006). Polymorphisms in the nine polymorphic membrane proteins of *Chlamydia trachomatis* across all serovars: evidence for serovar Da recombination and correlation with tissue tropism. Journal of Bacteriology ***188***, 275-286.

Goodlett, D.R., Bruce, J.E., Anderson, G.A., Rist, B., Pasa-Tolic, L., Fiehn, O., Smith, R.D., and Aebersold, R. (2000). Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. Analytical Chemistry ***72***, 1112-1118.

Gordon, F.B., and Quan, A.L. (1965). Occurence of glycogen in inclusions of the Psittacosis-Lymphogranuloma Venereum-Trachoma agents. Journal of Infectious Diseases ***115***, 186-196.

Gorg, A., Postel, W., and Gunther, S. (1988). The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis ***9***, 531-546.

Gras, R., Muller, M., Gasteiger, E., Gay, S., Binz, P.A., Bienvenut, W., Hoogland, C., Sanchez, J.C., Bairoch, A., Hochstrasser, D.F., *et al.* (1999). Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection. Electrophoresis ***20***, 3535-3550.

Grayston, J.T. (1992). Infections caused by *Chlamydia pneumoniae* strain TWAR. Clinical Infectious Diseases ***15***, 757-761.

Grayston, J.T., Aldous, M.B., Easton, A., Wang, S.P., Kuo, C.C., Campbell, L.A., and Altman, J. (1993). Evidence that *Chlamydia pneumoniae* causes pneumonia and bronchitis. Journal of Infectious Diseases ***168***, 1231-1235.

Grayston, J.T., Wang, S.P., Kuo, C.C., and Campbell, L.A. (1989). Current knowledge on *Chlamydia pneumoniae*, strain TWAR, an important cause of pneumonia and other acute respiratory diseases. European Journal of Clinical Microbiology and Infectious Diseases ***8***, 191-202.

Greber, U.F. (1998). Virus assembly and disassembly: the adenovirus cysteine protease as a trigger factor. Reviews in Medical Virology ***8***, 213-222.

Gregory, W.W., Gardner, M., Byrne, G.I., and Moulder, J.W. (1979). Arrays of hemispheric surface projections on *Chlamydia psittaci* and *Chlamydia trachomatis* observed by scanning electron microscopy. The Journal of Bacteriology ***138***, 241-244.

Grimwood, J., and Stephens, R.S. (1999). Computational analysis of the polymorphic membrane protein superfamily of *Chlamydia trachomatis* and *Chlamydia pneumoniae*. Microbial and Comparative Genomics ***4***, 187-201.

Guilhaus, M., Selby, D., and Mlynski, V. (2000). Orthogonal acceleration time-of-flight mass spectrometry. Mass Spectrometry Reviews ***19***, 65-107.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proceedings of the National Academy of Sciences of the United States of America ***97***, 9390-9395.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999b). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology ***17***, 994-999.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999a). Correlation between protein and mRNA abundance in yeast. Molecular and Cellular Biology ***19***, 1720-1730.

Hackstadt, T., Fischer, E.R., Scidmore, M.A., Rockey, D.D., and Heinzen, R.A. (1997). Origins and functions of the chlamydial inclusion. Trends in Microbiology ***5***, 288-293.

Hackstadt, T., Scidmore, M.A., and Rockey, D.D. (1995). Lipid metabolism in *Chlamydia trachomatis*-infected cells: directed trafficking of Golgi-derived sphingolipids to the chlamydial inclusion. Proceedings of the National Academy of Sciences of the United States of America ***92***, 4877-4881.

Hackstadt, T., Todd, W.J., and Caldwell, H.D. (1985). Disulfide-mediated interactions of the chlamydial major outer membrane protein: role in the differentiation of chlamydiae? Journal of Bacteriology ***161***, 25-31.

Hahn, D.L., Dodge, R.W., and Golubjatnikov, R. (1991). Association of *Chlamydia pneumoniae* (strain TWAR) infection with wheezing, asthmatic bronchitis, and adult-onset asthma. Journal of the American Medical Association ***266***, 225-230.

Haider, S., Wagner, M., Schmid, M.C., Sixt, B.S., Christian, J.G., Hacker, G., Pichler, P., Mechtler, K., Muller, A., Baranyi, C.*, et al.* (2010). Raman microspectroscopy reveals long-term extracellular activity of Chlamydiae. Molecular Microbiology ***77***, 687-700.

Hammerschlag, M.R., Chirgwin, K., Roblin, P.M., Gelling, M., Dumornay, W., Mandel, L., Smith, P., and Schachter, J. (1992). Persistent infection with *Chlamydia pneumoniae* following acute respiratory illness. Clinical Infectious Diseases ***14***, 178-182.

Hartley, J.C., Stevenson, S., Robinson, A.J., Littlewood, J.D., Carder, C., Cartledge, J., Clark, C., and Ridgway, G.L. (2001). Conjunctivitis due to *Chlamydophila felis* (*Chlamydia psittaci* feline pneumonitis agent) acquired from a cat: case report with molecular characterization of isolates from the patient and cat. Journal of Infection ***43***, 7-11.

Hatch, T. (1998). *Chlamydia*: old ideas crushed, new mysteries bared. Science ***282***, 638-639.

Hatch, T.P. (1996). Disulfide cross-linked envelope proteins: the functional equivalent of peptidoglycan in chlamydiae? Journal of Bacteriology ***178***, 1-5.

Hatch, T.P., Al-Hossainy, E., and Silverman, J.A. (1982). Adenine nucleotide and lysine transport in *Chlamydia psittaci*. The Journal of Bacteriology ***150***, 662-670.

Hatch, T.P., Allan, I., and Pearce, J.H. (1984). Structural and polypeptide differences between envelopes of infective and reproductive life cycle forms of *Chlamydia* spp. The Journal of Bacteriology ***157***, 13-20.

Hatch, T.P., Miceli, M., and Silverman, J.A. (1985). Synthesis of protein in host-free reticulate bodies of *Chlamydia psittaci* and *Chlamydia trachomatis*. Journal of Bacteriology ***162***, 938-942.

Hatch, T.P., Miceli, M., and Sublett, J.E. (1986). Synthesis of disulfide-bonded outer membrane proteins during the developmental cycle of *Chlamydia psittaci* and *Chlamydia trachomatis*. The Journal of Bacteriology ***165***, 379-385.

Hatch, T.P., Vance, D.W., Jr., and Al-Hossainy, E. (1981). Identification of a major envelope protein in *Chlamydia* spp. The Journal of Bacteriology ***146***, 426-429.

Havlis, J., and Shevchenko, A. (2004). Absolute quantification of proteins in solutions and in polyacrylamide gels by mass spectrometry. Analytical Chemistry ***76***, 3029-3036.

Hayes, R.N., and Gross, M.L. (1990). Collision-induced dissociation. Methods in Enzymology ***193***, 237-263.

Health, A.P. (2005). HIV and other Sexually Transmitted Infections in the UK: 2005.

Health, A.P. (2010). Total number of STI diagnoses in genitourinary medicine clinics and community-based settings by gender, England, 2008-2010. [WWW document]. URL: http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/STIs/STIsAnnualDataTables/#4._STI_data_for_the_UK (accessed 27 January 2012).

Hefty, P.S., and Stephens, R.S. (2007). Chlamydial type III secretion system is encoded on ten operons preceded by sigma 70-like promoter elements. Journal of Bacteriology ***189***, 198-206.

Henderson, I.R., and Lam, A.C. (2001). Polymorphic proteins of *Chlamydia* spp.--autotransporters beyond the Proteobacteria. Trends in Microbiology ***9***, 573-578.

Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C., and Watanabe, C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proceedings of the National Academy of Sciences of the United States of America ***90***, 5011-5015.

Hilger, M., and Mann, M. (2012). Triple SILAC to determine stimulus specific interactions in the Wnt pathway. Journal of Proteome Research ***11***, 982-994.

Ho, T.D., and Starnbach, M.N. (2005). The *Salmonella enterica* serovar typhimurium-encoded type III secretion systems can translocate Chlamydia trachomatis proteins into the cytosol of host cells. Infection and Immunity ***73***, 905-911.

Hobson, D., Lee, N., Quayle, E., and Beckett, E.E. (1982). Growth of *Chlamydia trachomatis* in Buffalo green monkey cells. Lancet ***2***, 872-873.

Hoffmann, E.d., Charette, J., and Stroobant, V. (1996). Mass spectrometry : principles and applications (Chichester ; Paris : Wiley : Masson).

Hogan, R.J., Mathews, S.A., Mukhopadhyay, S., Summersgill, J.T., and Timms, P. (2004). Chlamydial persistence: beyond the biphasic paradigm. Infection and Immunity ***72***, 1843-1855.

Hoke, I.I., Morand, K.L., Greis, K.D., Baker, T.R., Harbol, K.L., and Dobson, R.L.M. (2001). Transformations in pharmaceutical research and development, driven by innovations in multidimensional mass spectrometry-based technologies. International Journal of Mass Spectrometry ***212***, 135-196.

Hommais, F., Krin, E., Laurent-Winter, C., Soutourina, O., Malpertuy, A., Le Caer, J.P., Danchin, A., and Bertin, P. (2001). Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. Molecular Microbiology ***40***, 20-36.

Horn, M., Collingro, A., Schmitz-Esser, S., Beier, C.L., Purkhold, U., Fartmann, B., Brandt, P., Nyakatura, G.J., Droege, M., Frishman, D*., et al.* (2004). Illuminating the evolutionary history of chlamydiae. Science ***304***, 728-730.

Horner, P. (2005). *Chlamydia trachomatis* and non-gonococcal urethritis. Medicine ***33***, 40-42.

Hsia, R.C., Pannekoek, Y., Ingerowski, E., and Bavoil, P.M. (1997). Type III secretion genes identify a putative virulence locus of *Chlamydia*. Molecular Microbiology ***25***, 351-359.

Iliffe-Lee, E.R., and McClarty, G. (1999). Glucose metabolism in *Chlamydia trachomatis*: the 'energy parasite' hypothesis revisited. Molecular Microbiology ***33***, 177-187.

Iliffe-Lee, E.R., and McClarty, G. (2000). Regulation of carbon metabolism in *Chlamydia trachomatis*. Molecular Microbiology ***38***, 20-30.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Molecular & Cellular proteomics ***4***, 1265-1272.

Iwasaki, M., Miwa, S., Ikegami, T., Tomita, M., Tanaka, N., and Ishihama, Y. (2010). One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the *Escherichia coli* proteome on a microarray scale. Analytical Chemistry ***82***, 2616-2620.

Jaffe, J.D., Berg, H.C., and Church, G.M. (2004). Proteogenomic mapping as a complementary method to perform genome annotation. Proteomics ***4***, 59-77.

James, P., Quadroni, M., Carafoli, E., and Gonnet, G. (1993). Protein identification by mass profile fingerprinting. Biochemical and Biophysical Research Communications ***195***, 58-64.

Jamison, W.P., and Hackstadt, T. (2008). Induction of type III secretion by cell-free *Chlamydia trachomatis* elementary bodies. Microbial Pathogenesis ***45***, 435-440.

Jeffrey, B.M., Suchland, R.J., Quinn, K.L., Davidson, J.R., Stamm, W.E., and Rockey, D.D. (2010). Genome sequencing of recent clinical *Chlamydia trachomatis* strains identifies loci associated with tissue tropism and regions of apparent recombination. Infection and Immunity ***78***, 2544-2553.

Jewett, T.J., Fischer, E.R., Mead, D.J., and Hackstadt, T. (2006). Chlamydial TARP is a bacterial nucleator of actin. Proceedings of the National Academy of Sciences of the United States of America ***103***, 15599-15604.

Jones, B.M., Edwards, R.J., Skipp, P.J., O'Connor, C.D., and Iglesias-Rodriguez, M.D. (2011). Shotgun proteomic analysis of Emiliania huxleyi, a marine phytoplankton species of major biogeochemical importance. Marine biotechnology ***13***, 496-504.

Jones, B.R. (1975). The prevention of blindness from trachoma. TransOphthalmolSocUK ***95***, 16-33.

Kahane, S., Gonen, R., Sayada, C., Elion, J., and Friedman, M.G. (1993). Description and partial characterization of a new Chlamydia-like microorganism. FEMS MicrobiolLett ***109***, 329-333.

Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W., Olinger, L., Grimwood, J., Davis, R.W., and Stephens, R.S. (1999). Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nature Genetics ***21***, 385-389.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Analytical Chemistry ***60***, 2299-2301.

Karlin, S., Mrazek, J., Campbell, A., and Kaiser, D. (2001). Characterizations of highly expressed genes of four fast-growing bacteria. The Journal of Bacteriology ***183***, 5025-5040.

Karp, N.A., Huber, W., Sadowski, P.G., Charles, P.D., Hester, S.V., and Lilley, K.S. (2010). Addressing accuracy and precision issues in iTRAQ quantitation. Molecular & Cellular proteomics ***9***, 1885-1897.

Kelleher, N.L., Lin, H.Y., Valaskovic, G.A., Aaserud, D.J., Fridriksson, E.K., and McLafferty, F.W. (1999). Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. Journal of the American Chemical Society ***121***, 806-812.

Kiselev, A.O., Stamm, W.E., Yates, J.R., and Lampe, M.F. (2007). Expression, processing, and localization of PmpD of *Chlamydia trachomatis* Serovar L2 during the chlamydial developmental cycle. PLoS One ***2***, e568.

Kitteringham, N.R., Jenkins, R.E., Lane, C.S., Elliott, V.L., and Park, B.K. (2009). Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. Journal of chromatography B, Analytical technologies in the biomedical and life sciences ***877***, 1229-1239.

Kleba, B., and Stephens, R.S. (2008). Chlamydial effector proteins localized to the host cell cytoplasmic compartment. Infection and Immunity ***76***, 4842-4850.

Klevens, M.R., Edwards, J.R., Richards, C.L., Horan, T.C., Gaynes, R.P., Pollock, D.A., and Cardo, D.M. (2002). Estimating Health Care-Associated Infections and Deaths in U.S. Hospitals. Public Health Reports ***122***, 160-165.

Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik ***26***, 231-243.

Kocan, K.M., Crawford, T.B., Dilbeck, P.M., Evermann, J.F., and McGuire, T.C. (1990). Development of a rickettsia isolated from an aborted bovine fetus. Journal of Bacteriology ***172***, 5949-5955.

Kosseim, M., and Brunham, R.C. (1986). Fallopian tube obstruction as a sequela to *Chlamydia trachomatis* infection. European Journal of Clinical Microbiology ***5***, 584-590.

Kostryukova, E., Lazarev, V., and Govorum, V. (2008). Inclusion membrane proteins of Chlamydiaceae. Biochemistry (Moscow) Supplemental Series B: Biomedical Chemistry ***2***, 148-159.

Kraaipoel, R.J., and van Duin, A.M. (1979). Isoelectric focusing of *Chlamydia trachomatis*. Infection and Immunity ***26***, 775-778.

Krijgsveld, J., Ketting, R.F., Mahmoudi, T., Johansen, J., rtal-Sanz, M., Verrijzer, C.P., Plasterk, R.H., and Heck, A.J. (2003). Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. Nature Biotechnology ***21***, 927-931.

Kruger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C.A., Forner, F., Schmidt, S., Zanivan, S., Fassler, R., and Mann, M. (2008). SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell ***134***, 353-364.

Kubo, A., and Stephens, R.S. (2000). Characterization and functional analysis of PorB, a *Chlamydia* porin and neutralizing target. Molecular Microbiology ***38***, 772-780.

Kuo, C., Takahashi, N., Swanson, A.F., Ozeki, Y., and Hakomori, S. (1996). An N-linked high-mannose type oligosaccharide, expressed at the major outer membrane protein of *Chlamydia trachomatis*, mediates attachment and infectivity of the microorganism to HeLa cells. Journal of Clinical Investigation ***98***, 2813-2818.

Kuo, C.C., Gown, A.M., Benditt, E.P., and Grayston, J.T. (1993a). Detection of *Chlamydia pneumoniae* in aortic lesions of atherosclerosis by immunocytochemical stain. Arteriosclerosis, Thrombosis, and Vascular Biology ***13***, 1501-1504.

Kuo, C.C., Grayston, J.T., Campbell, L.A., Goo, Y.A., Wissler, R.W., and Benditt, E.P. (1995b). *Chlamydia pneumoniae* (TWAR) in coronary arteries of young adults (15-34 years old). Proceedings of the National Academy of Sciences of the United States of America ***92***, 6911-6914.

Kuo, C.C., and Grayston, T. (1976). Interaction of *Chlamydia trachomatis* organisms and HeLa 229 cells. Infection and Immunity ***13***, 1103-1109.

Kuo, C.C., Jackson, L.A., Campbell, L.A., and Grayston, J.T. (1995a). *Chlamydia pneumoniae* (TWAR). Clinical Microbiology Reviews ***8***, 451-461.

Kuo, C.C., Shor, A., Campbell, L.A., Fukushi, H., Patton, D.L., and Grayston, J.T. (1993b). Demonstration of *Chlamydia pneumoniae* in atherosclerotic lesions of coronary arteries. Journal of Infectious Diseases ***167***, 841-849.

Kuo, C.C., Wang, S.P., and Grayston, J.T. (1973). Effect of polycations, polyanions and neuraminidase on the infectivity of trachoma-inclusin conjunctivitis and lymphogranuloma venereum organisms HeLa cells: sialic acid residues as possible receptors for trachoma-inclusion conjunction. Infection and Immunity ***8***, 74-79.

Kuzyk, M.A., Smith, D., Yang, J., Cross, T.J., Jackson, A.M., Hardie, D.B., Anderson, N.L., and Borchers, C.H. (2009). Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Molecular & Cellular proteomics ***8***, 1860-1877.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature ***227***, 680-685.

Lander, E.S. (1999). Array of hope. Nature Genetics ***21***, 3-4.

Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G.*, et al.* (2002). Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. Nature ***419***, 537-542.

Lehtinen, M., Ault, K.A., Lyytikainen, E., Dillner, J., Garland, S.M., Ferris, D.G., Koutsky, L.A., Sings, H.L., Lu, S., Haupt, R.M.*, et al.* (2011). *Chlamydia trachomatis* infection and risk of cervical intraepithelial neoplasia. Sexually Transmitted Infections ***87***, 372-376.

Leinonen, M., and Saikku, P. (2002). Evidence for infectious agents in cardiovascular disease and atherosclerosis. The Lancet infectious diseases ***2***, 11-17.

Lerat, E., and Ochman, H. (2005). Recognizing the pseudogenes in bacterial genomes. Nucleic Acids Research ***33***, 3125-3132.

Leuthold, L.A., Grivet, C., Allen, M., Baumert, M., and Hopfgartner, G. (2004). Simultaneous selected reaction monitoring, MS/MS and MS3 quantitation for the analysis of pharmaceutical compounds in human plasma using chip-based infusion. Rapid Communications in Mass Spectrometry ***18***, 1995-2000.

Li, G.Z., Vissers, J.P., Silva, J.C., Golick, D., Gorenstein, M.V., and Geromanos, S.J. (2009). Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. Proteomics ***9***, 1696-1719.

Lindquist., A.E., and Stephens, R.S. (1998). Transcriptional activity of a sequence variable protein family in *Chlamydia trachomatis*. in Chlamydial infections. In Proceedings of the Ninth International Symposium on Human Chlamydial Infections, S.R. S., B. G.I., G. Christiansen, C.I. N., G.J. T., R.G. Rank, G.L. Ridgway, P. Saikku, J. Schachter, and W.E. Stamm, eds., pp. 259-262.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., III (1999). Direct analysis of protein complexes using mass spectrometry. Nature Biotechnology ***17***, 676-682.

Lipton, M.S., Pasa-Tolic, L., Anderson, G.A., Anderson, D.J., Auberry, D.L., Battista, J.R., Daly, M.J., Fredrickson, J., Hixson, K.K., Kostandarithes, H.*, et al.* (2002). Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. Proceedings of the National Academy of Sciences of the United States of America ***99***, 11049-11054.

Lisby, G. (1999). Application of nucleic acid amplification in clinical microbiology. Molecular Biotechnology ***12***, 75-99.

Liu, H., Sadygov, R.G., and Yates, J.R., III (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Analytical Chemistry ***76***, 4193-4201.

Liu, X., Afrane, M., Clemmer, D.E., Zhong, G., and Nelson, D.E. (2010). Identification of *Chlamydia trachomatis* outer membrane complex proteins by differential proteomics. Journal of Bacteriology ***192***, 2852-2860.

Longbottom, D., Russell, M., Dunbar, S.M., Jones, G.E., and Herring, A.J. (1998). Molecular cloning and characterization of the genes coding for the highly immunogenic cluster of 90-kilodalton envelope proteins from the *Chlamydia psittaci* subtype that causes abortion in sheep. Infection and Immunity ***66***, 1317-1324.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E.M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nature Biotechnology ***25***, 117-124.

Lugert, R., Kuhns, M., Polch, T., and Gross, U. (2004). Expression and localization of type III secretion-related proteins of *Chlamydia pneumoniae*. Medical Microbiology and Immunology ***193***, 163-171.

Lundemose, A.G., Birkelund, S., Larsen, P.M., Fey, S.J., and Christiansen, G. (1990). Characterization and identification of early proteins in *Chlamydia trachomatis* serovar L2 by two-dimensional gel electrophoresis. Infection and Immunity ***58***, 2478-2486.

Mabey, D., and Peeling, R.W. (2002). Lymphogranuloma venereum. Sexually Transmitted Infections ***78***, 90-92.

Maillet, I., Berndt, P., Malo, C., Rodriguez, S., Brunisholz, R.A., Pragai, Z., Arnold, S., Langen, H., and Wyss, M. (2007). From the genome sequence to the proteome and back: evaluation of *E. coli* genome annotation with a 2-D gel-based proteomics approach. Proteomics *7*, 1097-1106.

Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Analytical Chemistry ***72***, 1156-1162.

Malmstrom, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E.W., and Aebersold, R. (2009). Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. Nature ***460***, 762-765.

Mamyrin, B.A. (2001). Time-of-flight mass spectrometry (concepts, achievements, and prospects). International Journal of Mass Spectrometry ***206***, 251-266.

Manadas, B., English, J.A., Wynne, K.J., Cotter, D.R., and Dunn, M.J. (2009). Comparative analysis of OFFGel, strong cation exchange with pH gradient, and RP at high pH for first-dimensional separation of peptides from a membrane-enriched protein fraction. Proteomics ***9***, 5194-5198.

Mann, M. (2006). Functional and quantitative proteomics using SILAC. Nature reviews Molecular cell biology ***7***, 952-958.

Mare, C.J. (1994). Mammalian chlamydiosis (Boca Raton, Fl., U.S.A, CRC Press).

Marshall, A.G., Hendrickson, C.L., and Jackson, G.S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrometry Reviews ***17***, 1-35.

Matsumoto, A. (1982). Electron microscopic observations of surface projections on *Chlamydia psittaci* reticulate bodies. Journal of Bacteriology ***150***, 358-364.

Matsumoto, A., Fujiwara, E., and Higashi, N. (1976). Observations of the surface projections of infectious small cell of *Chlamydia psittaci* in thin sections. Journal of Electron Microscopy ***25***, 169-170.

McClarty, G., and Stephens, R.S. (1999). Chlamydia: Intracellular Biology, Pathogenesis and Immunity, American society for Microbiology (Washington DC), pp. 69-100.

McCoy, A.J., and Maurelli, A.T. (2006). Building the invisible wall: updating the chlamydial peptidoglycan anomaly. Trends in Microbiology ***14***, 70-77.

McCoy, A.J., Sandlin, R.C., and Maurelli, A.T. (2003). In vitro and in vivo functional activity of *Chlamydia* MurA, a UDP-N-acetylglucosamine enolpyruvyl transferase involved in peptidoglycan synthesis and fosfomycin resistance. The Journal of Bacteriology ***185***, 1218-1228.

McLuckey, S.A., Goeringer, D.E., and Glish, G.L. (1992). Collisional activation with random noise in ion trap mass spectrometry. Analytical Chemistry ***64***, 1455-1460.

Meyer, K.F., and Eddie, B. (1951). Human carrier of the psittacosis virus. Journal of Infectious Diseases ***88***, 109-125.

Michalski, A., Cox, J., and Mann, M. (2011). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. Journal of Proteome Research ***10***, 1785-1793.

Michel, R., Haurîder, P., MÅller, K.D., and Weishaar, I. (1994). Acanthamoeba from human nasal mucosa infected with an obligate intracellular parasite. European Journal of Parasitology ***30***, 104-110.

Miyashita, N. (2003). Problem of *Chlamydia pneumoniae* serology today. Internal Medicine ***42***, 919.

Miyashita, N., Matsumoto, A., Kubota, Y., Nakajima, M., Niki, Y., and Matsushima, T. (1996). Continuous isolation and characterization of *Chlamydia pneumoniae* from a patient with diffuse panbronchiolitis. Microbiology and Immunology ***40***, 547-552.

Moelleken, K., and Hegemann, J.H. (2008). The *Chlamydia* outer membrane protein OmcB is required for adhesion and exhibits biovar-specific differences in glycosaminoglycan binding. Molecular Microbiology ***67***, 403-419.

Molloy, M.P., Donohoe, S., Brzezinski, E.E., Kilby, G.W., Stevenson, T.I., Baker, J.D., Goodlett, D.R., and Gage, D.A. (2005). Large-scale evaluation of quantitative reproducibility and proteome coverage using acid cleavable isotope coded affinity tag mass spectrometry for proteomic profiling. Proteomics ***5***, 1204-1208.

Montigiani, S., Falugi, F., Scarselli, M., Finco, O., Petracca, R., Galli, G., Mariani, M., Manetti, R., Agnusdei, M., Cevenini, R*., et al.* (2002). Genomic approach for analysis of surface proteins in *Chlamydia pneumoniae*. Infection and Immunity ***70***, 368-379.

Moran, N.A., and Wernegreen, J.J. (2000). Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends in Ecology & Evolution ***15***, 321-326.

Moroni, A., Pavan, G., Donati, M., and Cevenini, R. (1996). Differences in the envelope proteins of *Chlamydia pneumoniae*, *Chlamydia trachomatis*, and *Chlamydia psittaci* shown by two-dimensional gel electrophoresis. Archives of Microbiology ***165***, 164-168.

Morris, H.R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R.S., Hoyes, J., and Bateman, R.H. (1996). High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. Rapid Communications in Mass Spectrometry ***10***, 889-896.

Morrison, R.P., Belland, R.J., Lyng, K., and Caldwell, H.D. (1989). Chlamydial disease pathogenesis. The 57-kD chlamydial hypersensitivity antigen is a stress response protein. Journal of Experimental Medicine ***170***, 1271-1283.

Moulder, J.W. (1962). Some basic properties of the psittacosis-lymphogranuloma venereum group of agents. Structure and chemical composition of isolated particles. Annals of the New York Academy of Sciences ***98***, 92-99.

Moulder, J.W. (1966). The relation of the psittacosis group (Chlamydiae) to bacteria and viruses. Annual Review of Microbiology ***20***, 107-130.

Moulder, J.W. (1991). Interaction of chlamydiae and host cells in vitro. Microbiological Reviews ***55***, 143-190.

Moulder, J.W., Hatch, T.P., Kuo, C.C., Schachter, J., and torz, J. (1984). Genus *Chlamydia* . In Bergey ' s Manual of Systematic Bacteriology , vol. 1, pp. 729Ò739. Edited by N. R. Krieg. Baltimore: Williams & Wilkins, Vol 1 (Baltimore, Williams & Wilkins).

Mu, F.T., Callaghan, J.M., Steele-Mortimer, O., Stenmark, H., Parton, R.G., Campbell, P.L., McCluskey, J., Yeo, J.P., Tock, E.P., and Toh, B.H. (1995). EEA1, an early endosome-associated protein. EEA1 is a conserved alpha-helical peripheral membrane protein flanked by cysteine "fingers" and contains a calmodulin-binding IQ motif. Journal of Biological Chemistry ***270***, 13503-13511.

Mukhopadhyay, S., Good, D., Miller, R.D., Graham, J.E., Mathews, S.A., Timms, P., and Summersgill, J.T. (2006). Identification of *Chlamydia pneumoniae* proteins in the transition from reticulate to elementary body formation. Molecular & Cellular proteomics ***5***, 2311-2318.

Munday, P.E., and Taylor-Robinson, D. (1983). Chlamydial infection in proctitis and Crohn's disease. British Medical Bulletin ***39***, 155-158.

Muth, T., Keller, D., Puetz, S.M., Martens, L., Sickmann, A., and Boehm, A.M. (2010). jTraqX: a free, platform independent tool for isobaric tag quantitation at the protein level. Proteomics ***10***, 1223-1225.

Mygind, P.H., Christiansen, G., Roepstorff, P., and Birkelund, S. (2000). Membrane proteins PmpG and PmpH are major constituents of *Chlamydia trachomatis* L2 outer membrane complex. FEMS Microbiology Letters ***186***, 163-169.

Nagaraj, N., Alexander Kulak, N., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012). System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. Molecular & Cellular proteomics ***11***, M111 013722.

Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. Molecular systems biology *7*, 548.

Natsume, T., Yamauchi, Y., Nakayama, H., Shinkawa, T., Yanagida, M., Takahashi, N., and Isobe, T. (2002). A direct nanoflow liquid chromatography - Tandem system for interaction proteomics. Analytical Chemistry ***74***, 4725-4733.

Neilson, K.A., Ali, N.A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S.C., and Haynes, P.A. (2011). Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics ***11***, 535-553.

Newhall, W.J. (1987). Biosynthesis and disulfide cross-linking of outer membrane components during the growth cycle of *Chlamydia trachomatis*. Infection and Immunity ***55***, 162-168.

Nicholas, B., Skipp, P., Mould, R., Rennard, S., Davies, D.E., O'Connor, C.D., and Djukanovic, R. (2006). Shotgun proteomic analysis of human-induced sputum. Proteomics ***6***, 4390-4401.

Nicholson, T.L., Olinger, L., Chong, K., Schoolnik, G., and Stephens, R.S. (2003). Global stage-specific gene regulation during the developmental cycle of *Chlamydia trachomatis*. The Journal of Bacteriology ***185***, 3179-3189.

Nieuwenhuis, R.F., Ossewaarde, J.M., van der Meijden, W.I., and Neumann, H.A. (2003). Unusual presentation of early lymphogranuloma venereum in an HIV-1 infected patient: effective treatment with 1 g azithromycin. Sexually Transmitted Infections ***79***, 453-455.

Nokihara, K. (1998). Procedures leading to primary structure determination of proteins in complex mixtures by gel electrophoresis and modern micro-scale analyses. Analytica Chimica Acta ***372***, 21-32.

Nunes, A., Gomes, J.P., Mead, S., Florindo, C., Correia, H., Borrego, M.J., and Dean, D. (2007). Comparative expression profiling of the *Chlamydia trachomatis* pmp gene family for clinical and reference strains. PLoS One ***2***, e878.

O'Connor, C.D., Adams, P., Alefounder, P., Farris, M., Kinsella, N., Li, Y., Payot, S., and Skipp, P. (2000). The analysis of microbial proteomes: strategies and data exploitation. Electrophoresis ***21***, 1178-1186.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. Journal of Biological Chemistry ***250***, 4007-4021.

Ogawa, M., Renesto, P., Azza, S., Moinier, D., Fourquet, P., Gorvel, J.P., and Raoult, D. (2007). Proteome analysis of *Rickettsia felis* highlights the expression profile of intracellular bacteria. Proteomics *7*, 1232-1248.

Olsen, J.V., de Godoy, L.M.F., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005). Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. Molecular & Cellular proteomics ***4***, 2010-2021.

Olsen, J.V., and Mann, M. (2004). Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. Proceedings of the National Academy of Sciences of the United States of America ***101***, 13417-13422.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Molecular & Cellular proteomics ***1***, 376-386.

Ong, S.E., Foster, L.J., and Mann, M. (2003). Mass spectrometric-based approaches in quantitative proteomics. Methods ***29***, 124-130.

Opiteck, G.J., Lewis, K.C., Jorgenson, J.W., and Anderegg, R.J. (1997). Comprehensive on-line LC/LC/MS of proteins. Analytical Chemistry ***69***, 1518-1524.

Oriel, J.D. (1992). Male genital *Chlamydia trachomatis* infections. Journal of Infection ***25 Suppl 1***, 35-37.

Oriel, J.D., and Ridgway, G.L. (1982). Genital infections by *Chlamydia trachomatis* (London, Edward Arnold Ltd).

Ostergaard, L., Birkelund, S., and Christiansen, G. (1990). Use of polymerase chain reaction for detection of *Chlamydia trachomatis*. Journal of Clinical Microbiology ***28***, 1254-1260.

Ouchi, K., Fujii, B., Kanamoto, Y., Karita, M., Shirai, M., and Nakazawa, T. (1998). *Chlamydia pneumoniae* in coronary and iliac arteries of Japanese patients with atherosclerotic cardiovascular diseases. Journal of Medical Microbiology ***47***, 907-913.

Ow, S.Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P.C. (2009). iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". Journal of Proteome Research ***8***, 5347-5355.

Paavonen, J. (1998). Pelvic inflammatory disease. From diagnosis to prevention. Dermatologic Clinics ***16***, 747-756, xii.

Page, J.S., Masselon, C.D., and Smith, R.D. (2004). FTICR mass spectrometry for qualitative and quantitative bioanalyses. Current Opinion in Biotechnology ***15***, 3-11.

Page, L.A. (1966). Interspecies transfer of psittacosis-LGV-trachoma agents: pathogenicity of two avian and two mammalian strains for eight species of birds and mammals. American Journal of Veterinary Research ***27***, 397-407.

Page, L.A. (1968). Proposal for the recognition of two species in the genus *Chlamydia* Jones, Rake, and Stearns, 1945. International Journal of Systematic Bacteriology ***18***, 51-66.

Pan, S., and Aebersold, R. (2007). Quantitative proteomics by stable isotope labeling and mass spectrometry. Methods in Molecular Biology ***367***, 209-218.

Pan, S., and Aebersold, R. (2007). Quantitative proteomics by stable isotope labeling and mass spectrometry. Methods in Molecular Biology ***367***, 209-218.

Pantoja, L.G., Miller, R.D., Ramirez, J.A., Molestina, R.E., and Summersgill, J.T. (2000). Inhibition of *Chlamydia pneumoniae* replication in human aortic smooth muscle cells by gamma interferon-induced indoleamine 2, 3-dioxygenase activity. Infection and Immunity ***68***, 6478-6481.

Park, S.K., Venable, J.D., Xu, T., and Yates, J.R., 3rd (2008). A quantitative analysis software tool for mass spectrometry-based proteomics. Nature Methods ***5***, 319-322.

Patton, W.F. (2002). Detection technologies in proteome analysis. J ChromatogrB, AnalytTechnolBiomedLife Sci ***771***, 3-31.

Paul, W., and Steinwedel, H.S. (1953). Ein neues Massenspektrometer ohne Magnetfeld. RZeitschrift für Naturforschung ***8a***, 448.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis ***20***, 3551-3567.

Peters, J., Wilson, D.P., Myers, G., Timms, P., and Bavoil, P.M. (2007). Type III secretion a la *Chlamydia*. Trends in Microbiology ***15***, 241-251.

Pickett, M.A., Everson, J.S., Pead, P.J., and Clarke, I.N. (2005). The plasmids of *Chlamydia trachomatis* and *Chlamydophila pneumoniae* (N16): accurate determination of copy number and the paradoxical effect of plasmid-curing agents. Microbiology ***151***, 893-903.

Plaunt, M.R., and Hatch, T.P. (1988). Protein synthesis early in the developmental cycle of *Chlamydia psittaci*. Infection and Immunity ***56***, 3021-3025.

Polack, S., Brooker, S., Kuper, H., Mariotti, S., Mabey, D., and Foster, A. (2005). Mapping the global distribution of trachoma. Bulletin of the World Health Organization ***83***, 913-919.

Pratt, J.M., Simpson, D.M., Doherty, M.K., Rivers, J., Gaskell, S.J., and Beynon, R.J. (2006). Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nature protocols ***1***, 1029-1043.

Quirk, J.T., and Kupinski, J.M. (2001). Chronic infection, inflammation, andepithelial ovarian cancer. Medical Hypotheses ***57***, 426-428.

Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Research ***12***, 1231-1245.

Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Utterback, T., Berry, K*., et al.* (2000). Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Research ***28***, 1397-1406.

Read, T.D., Myers, G.S.A., Brunham, R.C., Nelson, W.C., Paulsen, I.T., Heidelberg, J., Holtzapple, E., Khouri, H., Federova, N.B., Carty, H.A*., et al.* (2003). Genome sequence of *Chlamydophila caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. Nucleic Acids Research ***31***, 2134-2147.

Rehm, H. (2006). Protein Biochemistry and Proteomics. Elsevier Academic Press.

Rivers, J., Simpson, D.M., Robertson, D.H., Gaskell, S.J., and Beynon, R.J. (2007). Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. Molecular & Cellular proteomics ***6***, 1416-1427.

Rocha, E.P., Pradillon, O., Bui, H., Sayada, C., and Denamur, E. (2002). A new family of highly variable proteins in the *Chlamydophila pneumoniae* genome. Nucleic Acids Research ***30***, 4351-4360.

Rockey, D.D., Grosenbach, D., Hruby, D.E., Peacock, M.G., Heinzen, R.A., and Hackstadt, T. (1997). *Chlamydia psittaci* IncA is phosphorylated by the host cell and is exposed on the cytoplasmic face of the developing inclusion. Molecular Microbiology ***24***, 217-228.

Rockey, D.D., Heinzen, R.A., and Hackstadt, T. (1995). Cloning and characterization of a *Chlamydia psittaci* gene coding for a protein localized in the inclusion membrane of infected cells. Molecular Microbiology ***15***, 617-626.

Rockey, D.D., Lenart, J., and Stephens, R.S. (2000). Genome sequencing and our understanding of chlamydiae. Infection and Immunity ***68***, 5473-5479.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., *et al.* (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Molecular & Cellular proteomics ***3***, 1154-1169.

Rouchka, E.C., and Cha, E.I. (2009). Current Trends in Pseudogene Detection and Characterization. Current Bioinformatics ***4***, 112-119.

Rurangirwa, F.R., Dilbeck, P.M., Crawford, T.B., McGuire, T.C., and McElwain, T.F. (1999). Analysis of the 16S rRNA gene of micro-organism WSU 86-1044 from an aborted bovine foetus reveals that it is a member of the order Chlamydiales: proposal of Waddliaceae fam. nov., Waddlia chondrophila gen. nov., sp. nov. International Journal of Systematic Bacter_ology ***49 Pt 2***, 577-581.

Sadowski, P.G., Groen, A.J., Dupree, P., and Lilley, K.S. (2008). Sub-cellular localization of membrane proteins. Proteomics ***8***, 3991-4011.

Saikku, P., Leinonen, M., Mattila, K., Ekman, M.R., Nieminen, M.S., Makela, P.H., Huttunen, J.K., and Valtonen, V. (1988). Serological evidence of an association of a novel *Chlamydia*, TWAR, with chronic coronary heart disease and acute myocardial infarction. Lancet ***2***, 983-986.

Saka, H.A., Thompson, J.W., Chen, Y.S., Kumar, Y., Dubois, L.G., Moseley, M.A., and Valdivia, R.H. (2011). Quantitative proteomics reveals metabolic and pathogenic properties of *Chlamydia trachomatis* developmental forms. Molecular Microbiology ***82***, 1185-1203.

Salari, S.H., and Ward, M.E. (1981). Polypeptide composition of *Chlamydia trachomatis*. Journal of General Microbiology ***123***, 197-207.

Schachter, J. (1988). Overview of human diseases. In Microbiology of *Chlamydia* (CRC press, Boca Raton, Florida), pp. 153-163.

Schachter, J., and Grossman, M. (1981). Chlamydial infections. Annual Review of Medicine ***32***, 45-61.

Schachter, J., Stamm, W.E., Quinn, T.C., Andrews, W.W., Burczak, J.D., and Lee, H.H. (1994). Ligase chain reaction to detect *Chlamydia trachomatis* infection of the cervix. Journal of Clinical Microbiology ***32***, 2540-2543.

Schirle, M., Heurtier, M.-A., and Kuster, B. (2003). Profiling Core Proteomes of Human Cell Lines by One-dimensional PAGE and Liquid Chromatography-Tandem Mass Spectrometry. Molecular & Cellular proteomics ***2***, 1297-1305.

Schmidt, A., Karas, M., and Dulcks, T. (2003). Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? Journal of the American Society for Mass Spectrometry ***14***, 492-500.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature ***473***, 337-342.

Schwartz, R., Ting, C.S., and King, J. (2001). Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. Genome Research ***11***, 703-709.

Scidmore, M.A., Fischer, E.R., and Hackstadt, T. (1996). Sphingolipids and glycoproteins are differentially trafficked to the *Chlamydia trachomatis* inclusion. The Journal of Cell Biology ***134***, 363-374.

Scidmore-Carlson, M.A., Shaw, E.I., Dooley, C.A., Fischer, E.R., and Hackstadt, T. (1999). Identification and characterization of a *Chlamydia trachomatis* early operon encoding four novel inclusion membrane proteins. Molecular Microbiology ***33***, 753-765.

Scigelova, M., and Makarov, A. (2006). Orbitrap mass analyzer--overview and applications in proteomics. Proteomics ***6 Suppl 2***, 16-21.

Shadforth, I.P., Dunkley, T.P., Lilley, K.S., and Bessant, C. (2005). i-Tracker: for quantitative proteomics using iTRAQ. BMC Genomics ***6***, 145.

Shaw, A.C., Christiansen, G., and Birkelund, S. (1999). Effects of interferon gamma on *Chlamydia trachomatis* serovar A and L2 protein expression investigated by two-dimensional gel electrophoresis. Electrophoresis ***20***, 775-780.

Shaw, A.C., Christiansen, G., Roepstorff, P., and Birkelund, S. (2000b). Genetic differences in the *Chlamydia trachomatis* tryptophan synthase alpha-subunit can explain variations in serovar pathogenesis. Microbes and Infection ***2***, 581-592.

Shaw, A.C., Gevaert, K., Demol, H., Hoorelbeke, B., Vandekerckhove, J., Larsen, M.R., Roepstorff, P., Holm, A., Christiansen, G., and Birkelund, S. (2002a). Comparative proteome analysis of *Chlamydia trachomatis* serovar A, D and L2. Proteomics ***2***, 164-186.

Shaw, A.C., Vandahl, B.B., Larsen, M.R., Roepstorff, P., Gevaert, K., Vandekerckhove, J., Christiansen, G., and Birkelund, S. (2002b). Characterization of a secreted *Chlamydia* protease. Cellular Microbiology ***4***, 411-424.

Shaw, E.I., Dooley, C.A., Fischer, E.R., Scidmore, M.A., Fields, K.A., and Hackstadt, T. (2000a). Three temporal classes of gene expression during the *Chlamydia trachomatis* developmental cycle. Molecular Microbiology ***37***, 913-925.

Shevchenko, A., Loboda, A., Ens, W., and Standing, K.G. (2000). MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. Analytical Chemistry ***72***, 2132-2141.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Analytical Chemistry ***68***, 850-858.

Shimada, T., Park, B.G., Wolf, A.J., Brikos, C., Goodridge, H.S., Becker, C.A., Reyes, C.N., Miao, E.A., Aderem, A., Gotz, F.*, et al.* (2010). *Staphylococcus aureus* evades lysozyme-based peptidoglycan digestion that links phagocytosis, inflammasome activation, and IL-1beta secretion. Cell Host Microbe ***7***, 38-49.

Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S.*, et al.* (2000). Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA. Nucleic Acids Research ***28***, 2311-2314.

Silva, J.C., Denny, R., Dorschel, C., Gorenstein, M.V., Li, G.Z., Richardson, K., Wall, D., and Geromanos, S.J. (2006a). Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. Molecular & Cellular proteomics ***5***, 589-607.

Silva, J.C., Denny, R., Dorschel, C.A., Gorenstein, M., Kass, I.J., Li, G.Z., McKenna, T., Nold, M.J., Richardson, K., Young, P.*, et al.* (2005). Quantitative proteomic analysis by accurate mass retention time pairs. Analytical Chemistry ***77***, 2187-2200.

Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P.C., and Geromanos, S.J. (2006b). Absolute quantification of proteins by LCMS^{E} - A virtue of parallel MS acquisition. Molecular & Cellular proteomics ***5***, 144-156.

Simonsen, A., Lippe, R., Christoforidis, S., Gaullier, J.M., Brech, A., Callaghan, J., Toh, B.H., Murphy, C., Zerial, M., and Stenmark, H. (1998). EEA1 links PI(3)K function to Rab5 regulation of endosome fusion. Nature ***394***, 494-498.

Sixt, B.S., Heinz, C., Pichler, P., Heinz, E., Montanaro, J., Op den Camp, H.J., Ammerer, G., Mechtler, K., Wagner, M., and Horn, M. (2011). Proteomic analysis reveals a virtually complete set of proteins for translation and energy generation in elementary bodies of the amoeba symbiont Protochlamydia amoebophila. Proteomics ***11***, 1868-1892.

Skipp, P., Robinson, J., O'Connor, C.D., and Clarke, I.N. (2005). Shotgun proteomic analysis of *Chlamydia trachomatis*. Proteomics ***5***, 1558-1573.

Skipp, P.J., and O'Connor, D. (2011). Shotgun proteomics: future perspective in bioanalysis. Bioanalysis ***3***, 2159-2160.

Slepenkin, A., Motin, V., de la Maza, L.M., and Peterson, E.M. (2003). Temporal expression of type III secretion genes of *Chlamydia pneumoniae*. Infection and Immunity ***71***, 2555-2562.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985). Measurement of protein using bicinchoninic acid. Analytical Biochemistry ***150***, 76-85.

Somboonna, N., Wan, R., Ojcius, D.M., Pettengill, M.A., Joseph, S.J., Chang, A., Hsu, R., Read, T.D., and Dean, D. (2011). Hypervirulent *Chlamydia trachomatis* clinical strain is a recombinant between lymphogranuloma venereum (L(2)) and D lineages. mBio ***2***, e00045-00011.

Spaargaren, J., Fennema, H.S., Morre, S.A., de Vries, H.J., and Coutinho, R.A. (2005a). New lymphogranuloma venereum *Chlamydia trachomatis* variant, Amsterdam. Emerging Infectious Diseases ***11***, 1090-1092.

Spaargaren, J., Schachter, J., Moncada, J., de Vries, H.J.C., Fennema, H.S.A., Pena, A.S., Coutinho, R.A., and Morre, S.A. (2005b). Slow epidemic of lymphogranuloma venereum L2b strain. Emerging Infectious Diseases ***11***, 1787-1788.

Spahr, C.S., Davis, M.T., McGinley, M.D., Robinson, J.H., Bures, E.J., Beierle, J., Mort, J., Courchesne, P.L., Chen, K., Wahl, R.C.*, et al.* (2001). Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. Proteomics ***1***, 93-107.

Stary, A., Schuh, E., Kerschbaumer, M., Gotz, B., and Lee, H. (1998). Performance of transcription-mediated amplification and ligase chain reaction assays for detection of chlamydial infection in urogenital samples obtained by invasive and noninvasive methods. Journal of Clinical Microbiology ***36***, 2666-2670.

Stenner-Liewen, F., Liewen, H., Zapata, J.M., Pawlowski, K., Godzik, A., and Reed, J.C. (2002). CADD, a *Chlamydia* protein that interacts with death receptors. The Journal of biological chemistry ***277***, 9633-9636.

Stephens, R.S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q.*, et al.* (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science ***282***, 754-759.

Stephens, R.S., and Lammel, C.J. (2001). *Chlamydia* outer membrane protein discovery using genomics. Current Opinion in Microbiology ***4***, 16-20.

Stephens, W. (1946). Proceedings of the American Physica Society: Minutes of the Spring Meeting at Cambridge, pp. 691.

Storz, J. (1988). Overview of animal diseases induced by chlamydial infections, A.L. Barron, ed. (CRC press), pp. 167-192.

Storz, J., and Page, L.A. (1971). Taxonomy of the Chlamydiae : reasons for classifying organisms of the genus *Chlamydia* , family Chlamydiaceae , in a separate order, Chlamydiales ord. nov, pp. 332-334.

Stothard, D.R., Toth, G.A., and Batteiger, B.E. (2003). Polymorphic membrane protein H has evolved in parallel with the three disease-causing groups of *Chlamydia trachomatis*. Infection and Immunity ***71***, 1200-1208.

Strohman, R. (1994). Epigenesis: the missing beat in biotechnology? Biotechnology (NY) ***12***, 156-164.

Struyve, M., Moons, M., and Tommassen, J. (1991). Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. Journal of Molecular Biology ***218***, 141-148.

Su, H., Watkins, N.G., Zhang, Y.X., and Caldwell, H.D. (1990). *Chlamydia trachomatis*-host cell interactions: role of the chlamydial major outer membrane protein as an adhesin. Infection and Immunity ***58***, 1017-1025.

Subtil, A., Blocker, A., and utry-Varsat, A. (2000). Type III secretion system in *Chlamydia* species: identified members and candidates. Microbes and Infection ***2***, 367-369.

Subtil, A., Delevoye, C., Balana, M.E., Tastevin, L., Perrinet, S., and Dautry-Varsat, A. (2005). A directed screen for chlamydial proteins secreted by a type III mechanism identifies a translocated protein and numerous other new candidates. Molecular Microbiology ***56***, 1636-1647.

Subtil, A., Parsot, C., and Dautry-Varsat, A. (2001). Secretion of predicted Inc proteins of *Chlamydia pneumoniae* by a heterologous type III machinery. Molecular Microbiology ***39***, 792-800.

Subtil, A., Parsot, C., and utry-Varsat, A. (2001). Secretion of predicted Inc proteins of *Chlamydia pneumoniae* by a heterologous type III machinery. Molecular Microbiology ***39***, 792-800.

Suchland, R.J., Rockey, D.D., Weeks, S.K., Alzhanov, D.T., and Stamm, W.E. (2005). Development of Secondary Inclusions in Cells Infected by *Chlamydia trachomatis*. Infection and Immunity ***73***, 3954-3962.

Swanson, A.F., and Kuo, C.C. (1990). Identification of lectin-binding proteins in *Chlamydia* species. Infection and Immunity ***58***, 502-507.

Swanson, A.F., and Kuo, C.C. (1991a). The characterization of lectin-binding proteins of *Chlamydia trachomatis* as glycoproteins. Microbial Pathogenesis ***10***, 465-473.

Swanson, A.F., and Kuo, C.C. (1991b). Evidence that the major outer membrane protein of *Chlamydia trachomatis* is glycosylated. Infection and Immunity ***59***, 2120-2125.

Szaflarski, W., and Nierhaus, K.H. (2007). Question 7: optimized energy consumption for protein synthesis. Origins of Life and Evolution of the Biosphere ***37***, 423-428.

Takeda, K., Kaisho, T., and Akira, S. (2003). Toll-like receptors. Annual Review of Immunology ***21***, 335-376.

Tannu, N.S., and Hemby, S.E. (2006). Two-dimensional fluorescence difference gel electrophoresis for comparative proteomics profiling. Nature protocols ***1***, 1732-1742.

Tanzer, R.J., and Hatch, T.P. (2001). Characterization of outer membrane proteins in *Chlamydia trachomatis* LGV serovar L2. Journal of Bacteriology ***183***, 2686-2690.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., and Conway, T. (1999). Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. Journal of Bacteriology ***181***, 6425-6440.

Taylor-Robinson, D. (1992). The value of non-culture techniques for diagnosis of *Chlamydia trachomatis* infections: making the best of a bad job. European Journal of Clinical Microbiology and Infectious Diseases ***11***, 499-503.

Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., and Hamon, C. (2003). Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS (vol 15, pg 1895, 2003). Analytical Chemistry ***75***, 4942-4942.

Thomson, N.R., Holden, M.T., Carder, C., Lennard, N., Lockey, S.J., Marsh, P., Skipp, P., O'Connor, C.D., Goodhead, I., Norbertzcak, H.*, et al.* (2008). Chlamydia trachomatis: genome sequence analysis of lymphogranuloma venereum isolates. Genome Research ***18***, 161-171.

Thomson, N.R., Yeats, C., Bell, K., Holden, M.T., Bentley, S.D., Livingstone, M., Cerdeno-Tarraga, A.M., Harris, B., Doggett, J., Ormond, D.*, et al.* (2005). The *Chlamydophila abortus* genome sequence reveals an array of variable proteins that contribute to interspecies variation. Genome Research ***15***, 629-640.

Tipples, G., and McClarty, G. (1993). The obligate intracellular bacterium *Chlamydia trachomatis* is auxotrophic for three of the four ribonucleoside triphosphates. Molecular Microbiology ***8***, 1105-1114.

Tjaden, J., Winkler, H.H., Schwoppe, C., Van Der, L.M., Mohlmann, T., and Neuhaus, H.E. (1999). Two nucleotide transport proteins in *Chlamydia trachomatis*, one for net nucleoside triphosphate uptake and the other for transport of energy. The Journal of Bacteriology ***181***, 1196-1202.

Tran, J.C., Zamdborg, L., Ahlf, D.R., Lee, J.E., Catherman, A.D., Durbin, K.R., Tipton, J.D., Vellaichamy, A., Kellie, J.F., Li, M*., et al.* (2011). Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature ***480***, 254-258.

Tucker, A.M., Driskell, L.O., Pannell, L.K., and Wood, D.O. (2011). Differential proteomic analysis of *Rickettsia prowazekii* propagated in diverse host backgrounds. Applied and Environmental Microbiology ***77***, 4712-4718.

Unwin, R.D., Griffiths, J.R., and Whetton, A.D. (2009). A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). Nature protocols ***4***, 870-877.

Utleg, A.G., Yi, E.C., Xie, T., Shannon, P., White, J.T., Goodlett, D.R., Hood, L., and Lin, B. (2003). Proteomic analysis of human prostasomes. Prostate ***56***, 150-161.

Uyeda, C.T., Welborn, P., Ellison-Birang, N., Shunk, K., and Tsaouse, B. (1984). Rapid diagnosis of chlamydial infections with the MicroTrak direct test. Journal of Clinical Microbiology ***20***, 948-950.

Van de Velde, S., Delaive, E., Dieu, M., Carryn, S., Van Bambeke, F., Devreese, B., Raes, M., and Tulkens, P.M. (2009). Isolation and 2-D-DIGE proteomic analysis of intracellular and extracellular forms of *Listeria monocytogenes*. Proteomics ***9***, 5484-5496.

Vandahl, B.B., Birkelund, S., and Christiansen, G. (2002). Proteome analysis of *Chlamydia pneumoniae*. Methods in Enzymology ***358***, 277-288.

Vandahl, B.B., Birkelund, S., and Christiansen, G. (2004). Genome and proteome analysis of *Chlamydia*. Proteomics ***4***, 2831-2842.

Vandahl, B.B., Birkelund, S., Demol, H., Hoorelbeke, B., Christiansen, G., Vandekerckhove, J., and Gevaert, K. (2001). Proteome analysis of the *Chlamydia pneumoniae* elementary body. Electrophoresis ***22***, 1204-1223.

Vaughan, T.E., Skipp, P.J., O'Connor, C.D., Hudson, M.J., Vipond, R., Elmore, M.J., and Gorringe, A.R. (2006). Proteomic analysis of *Neisseria lactamica* and *Neisseria meningitidis* outer membrane vesicle vaccine antigens. Vaccine ***24***, 5277-5293.

Veenstra, T.D., Conrads, T.P., and Issaq, H.J. (2004). What to do with "one-hit wonders"? Electrophoresis ***25***, 1278-1279.

Vissers, J.P.C., Langridge, J.I., and Aerts, J.M.F.G. (2007). Analysis and quantification of diagnostic serum markers and protein signatures for Gaucher disease. Molecular & Cellular proteomics ***6***, 755-766.

Wallin, K.L., Wiklund, F., Luostarinen, T., Angstrom, T., Anttila, T., Bergman, F., Hallmans, G., Ikaheimo, I., Koskela, P., Lehtinen, M.*, et al.* (2002). A population-based prospective study of *Chlamydia trachomatis* infection and cervical carcinoma. International Journal of Cancer ***101***, 371-374.

Wang, H., Alvarez, S., and Hicks, L.M. (2012). Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two *Chlamydomonas reinhardtii* strains of interest for biofuels engineering. Journal of Proteome Research ***11***, 487-501.

Wang, H., Chang-Wong, T., Tang, H.Y., and Speicher, D.W. (2010). Comparison of extensive protein fractionation and repetitive LC-MS/MS analyses on depth of analysis for complex proteomes. Journal of Proteome Research ***9***, 1032-1040.

Wang, S.P., and Grayston, J.T. (1991a). Serotyping of *Chlamydia trachomatis* by indirect fluorescent-antibody staining of inclusions in cell culture with monoclonal antibodies. Journal of Clinical Microbiology ***29***, 1295-1298.

Wang, S.P., and Grayston, J.T. (1991b). Three new serovars of *Chlamydia trachomatis*: Da, Ia, and L2a. Journal of Infectious Diseases ***163***, 403-405.

Wang, S.P., Kuo, C.C., Barnes, R.C., Stephens, R.S., and Grayston, J.T. (1985). Immunotyping of *Chlamydia trachomatis* with monoclonal antibodies. Journal of Infectious Diseases ***152***, 791-800.

Wang, Y., Kahane, S., Cutcliffe, L.T., Skilton, R.J., Lambden, P.R., and Clarke, I.N. (2011). Development of a transformation system for *Chlamydia trachomatis*: restoration of glycogen biosynthesis by acquisition of a plasmid shuttle vector. PLoS pathogens ***7***, e1002258.

Ward, A.M., Rogers, J.H., and Estcourt, C.S. (1999). *Chlamydia trachomatis* infection mimicking testicular malignancy in a young man. Sexually Transmitted Infections ***75***, 270.

Ward, M.E. (1988). The chlamydial developmental cycle. Chapter 4 In, A.L. Barron, ed. (Boca Raton Florida, CRC Press), pp. 71-95.

Washburn, M.P., Wolters, D., and Yates, J.R., III (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature Biotechnology ***19***, 242-247.

Wehrl, W., Brinkmann, V., Jungblut, P.R., Meyer, T.F., and Szczepek, A.J. (2004). From the inside out--processing of the Chlamydial autotransporter PmpD and its role in bacterial adhesion and activation of human host cells. Molecular Microbiology ***51***, 319-334.

Wenman, W.M., and Meuser, R.U. (1986). *Chlamydia trachomatis* elementary bodies possess proteins which bind to eucaryotic cell membranes. Journal of Bacteriology ***165***, 602-607.

Whitcher, J.P., Srinivasan, M., and Upadhyay, M.P. (2001). Corneal blindness: a global perspective. Bulletin of the World Health Organization ***79***, 214-221.

WHO (2005). Prevalence and incidence of selected sexually transmitted infections: Centers for Disease Control and Prevention, STD Surveillance Report [WWW document]. URL: whqlibdoc.who.int/publications/2011/9789241502450_eng.pdf (accessed January 2011)
.

Wichlan, D.G., and Hatch, T.P. (1993). Identification of an early-stage gene of *Chlamydia psittaci* 6BC. The Journal of Bacteriology ***175***, 2936-2942.

Wiley, W.C., and McLaren, I.H. (1955). Time-of-Flight Mass Spectrometer with Improved Resolution. Review of Scientific Instruments ***26***, 1150-1157.

Williams, J.N., Skipp, P.J., Humphries, H.E., Christodoulides, M., O'Connor, C.D., and Heckels, J.E. (2007). Proteomic analysis of outer membranes and vesicles from wild-type serogroup B *Neisseria meningitidis* and a lipopolysaccharide-deficient mutant. Infection and Immunity ***75***, 1364-1372.

Wilm, M., and Mann, M. (1996). Analytical properties of the nanoelectrospray ion source. Analytical Chemistry ***68***, 1-8.

Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D.A., and White, F.M. (2007). Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. Proceedings of the National Academy of Sciences of the United States of America ***104***, 5860-5865.

Wolters, D.A., Washburn, M.P., and Yates, J.R., III (2001). An automated multidimensional protein identification technology for shotgun proteomics. Analytical Chemistry ***73***, 5683-5690.

Wylie, J.L., Hatch, G.M., and McClarty, G. (1997). Host cell phospholipids are trafficked to and then modified by *Chlamydia trachomatis*. Journal of Bacteriology ***179***, 7233-7242.

Wyrick, P.B. (2000). Intracellular survival by *Chlamydia* . Microreview. Cellular microbiology ***2***, 275-282.

Wyrick, P.B. (2010). *Chlamydia trachomatis* persistence in vitro: an overview. The Journal of infectious diseases ***201 Suppl 2***, S88-95.

Xiang, Y., and Koomen, J.M. (2012). Evaluation of direct infusion-multiple reaction monitoring mass spectrometry for quantification of heat shock proteins. Analytical Chemistry ***84***, 1981-1986.

Xu, G., and West, T.P. (1992). Protein synthesis during germination of heterothallic yeast ascospores. Experientia ***48***, 786-788.

Yang, J.L., Schachter, J., Moncada, J., Habte, D., Zerihun, M., House, J.I., Zhou, Z., Hong, K.C., Maxey, K., Gaynor, B.D.*, et al.* (2007). Comparison of an rRNA-based and DNA-based nucleic acid amplification test for the detection of *Chlamydia trachomatis* in trachoma. The British journal of ophthalmology ***91***, 293-295.

Yates, J.R., III, Speicher, S., Griffin, P.R., and Hunkapiller, T. (1993). Peptide mass maps: a highly informative approach to protein identification. Analytical Biochemistry ***214***, 397-408.

Yuan, Y., Zhang, Y.X., Watkins, N.G., and Caldwell, H.D. (1989). Nucleotide and deduced amino acid sequences for the four variable domains of the major outer membrane proteins of the 15 *Chlamydia trachomatis* serovars. Infection and Immunity ***57***, 1040-1049.

Zenobi, R., and Knochenmuss, R. (1998). Ion formation in MALDI mass spectrometry. Mass Spectrometry Reviews ***17***, 337-366.

Zhang, W., and Chait, B.T. (2000). ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. Analytical Chemistry ***72***, 2482-2489.

Zhong, G., Fan, P., Ji, H., Dong, F., and Huang, Y. (2001a). Identification of a Chlamydial Protease-like Activity Factor Responsible for the Degradation of Host Transcription Factors. The Journal of Experimental Medicine ***193***, 935-942.

Zhong, J., Douglas, A.L., and Hatch, T.P. (2001b). Characterization of integration host factor (IHF) binding upstream of the cysteine-rich protein operon (omcAB) promoter of *Chlamydia trachomatis* LGV serovar L2. Molecular Microbiology ***41***, 451-462.

APPENDIX 1

Appendix I, Table 5.1. Proteins assigned and quantified by iTRAQ analysis of EBs and RBs from *C. trachomatis* L2\434\Bu.

Category/protein name	Gene name	Primary locus	GenInfo identifier	UniProt accession number	Serovar D orthologue	pI[a)]	Mass (kDa)[a)]	Total number of peptide matches	Number of unique peptide matches	Reporter ion ratios[b)] 114/115	Reporter ion ratios[b)] 116/117	mean protein ratio	Standard deviation	Fold-change[c)]
Amino acid biosynthesis														
Serine Hydroxymethyltransferase	*glyA*	CTL0691	166154645	B0B804	CT432	6.04	54.2	13	9	1.32	0.98	1.15	0.24	1.1
Aromatic AA Aminotransferase	*aspC*	CTL0005	166153978	B0B8L2	CT637	5.42	44.7	6	5	ND	ND	ND	ND	ND
Dehyroquinate Synthase	*aroB*	CTL0623	166154580	B0B7T8	CT369	7.67	41.2	8	7	ND	ND	ND	ND	ND
Dihydrodipicolinate Synthase	*dapA*	CTL0615	166154572	B0B7T0	CT361	5.52	31.2	3	3	ND	ND	ND	ND	ND
Biosynthesis of cofactors														
Geranylgeranyl pyrophosphate synthase	*ispA*	CTL0892	166154843	B0B8K3	CT628	4.88	32.5	6	4	1.10	0.95	1.02	0.10	1.0
GTP Cyclohydratase and DHBP Synthase	*ribA*	CTL0100	166154073	B0B8V7	CT731	5.11	46.8	7	3	ND	ND	ND	ND	ND
Porphobilinogen Synthase	*hemB*	CTL0001	166153974	B0B8K8	CT633	5.94	37.7	5	4	ND	ND	ND	ND	ND
Phenylacrylate Decarboxylase		CTL0472	166154430	B0B9X0	CT220	6.64	19.2	4	4	ND	ND	ND	ND	ND
Cell Envelope														
Major Outer Membrane Protein	*ompA*	CTL0050	166154023	B0B8Q7	CT681	5.06	42.5	449	20	1.10	1.02	1.06	0.05	1.1
60kDa Cysteine-Rich OMP	*omcB*	CTL0702	166154656	B0B815	CT443	8.08	56.4	170	28	0.48	0.46	0.47	0.02	-2.1
Putative outer membrane protein B	*pmpB*	CTL0670	166154624	B0B7Y3	CT413	8.21	183.0	64	37	0.72	0.72	0.72	0.01	-1.4
Putative Outer Membrane Protein D	*pmpD*	CTL0183	166154155	B0B940	CT812	4.75	156.7	69	30	0.80	0.78	0.79	0.02	1.3
OmpH-Like Outer Membrane Protein	*ompH*	CTL0494	166154453	B0B7F8	CT242	4.65	17.3	18	8	1.38	1.33	1.35	0.04	+1.4
Candidate inclusion membrane protein		CTL0476	166154434	B0B9X4	CT223	6.75	29.6	39	13	0.67	0.68	0.67	0.00	-1.5
Peptidoglycan-Associated Lipoprotein	*pal*	CTL0863	166154814	B0B8H4	CT600	7.91	19.0	14	3	0.85	0.8	0.82	0.04	1.2
Putative Outer Membrane Protein G	*pmpG*	CTL0250	166154217	B0B9A3	CT871	5.66	104.2	22	13	1.13	1.14	1.14	0.01	1.1

Putative Outer Membrane Protein I	*pmpI*	CTL0254	166154220	B0B9A6	CT874	6.07	92.9	36	12	0.87	0.92	0.89	0.03	1.1
Outer Membrane Protein Analog	*ompB*	CTL0082	166154055	B0B8T9	CT713	5.18	34.6	11	4	0.91	0.91	0.91	0.00	1.1
UDP-N-Acetylglucosamine Transferase	*murA*	CTL0715	166154669	B0B828	CT455	6.15	48.4	15	9	ND	ND	ND	ND	ND
Putative Outer Membrane Protein H	*pmpH*	CTL0251	166154218	B0B9A4	CT872	6.6	104.6	20	15	1.01	0.99	1.00	0.02	1.0
Glucosamine-Fructose-6-P Aminotransferase	*glmS*	CTL0188	166154160	B0B945	CT816	5.38	67.4	20	14	1.52	1.56	1.54	0.03	+1.5
Putative outer membrane protein C	*pmpC*	CTL0671	166154625	B0B7Y4	CT414	4.58	184.9	24	20	0.86	1.04	0.95	0.13	1.1
Muramidase (invasin repeat family)	*nlpD*	CTL0128	166154101	B0B8Y5	CT759	9.08	24.1	12	6	ND	ND	ND	ND	ND
60kDa Inner Membrane Protein	*oxaA*	CTL0503	166154462	B0B7G7	CT251	8.35	85.0	32	14	1.35	1.26	1.30	0.06	1.3
Possible Transmembrane Protein		CTL0386	166154352	B0B9N9	CT131	6.59	123.0	34	18	ND	ND	ND	ND	ND
Inclusion Membrane Protein A	*incA*	CTL0374	166154340	B0B9M7	CT119	6.15	24.7	13	8	ND	ND	ND	ND	ND
Omp85 Analog		CTL0493	166154452	B0B7F7	CT241	8.96	85.6	21	14	ND	ND	ND	ND	ND
PBP2-transglycolase/transpeptidase	*pbpB*	CTL0051	166154024	B0B8Q8	CT682	6.36	124.0	34	17	ND	ND	ND	ND	ND
D-alanyl-D-alanine carboxypeptidase	*dacC*	CTL0813	166154766	B0B8C5	CT551	8.98	46.0	6	5	ND	ND	ND	ND	ND
UDP-N-acetylmuramoylalanine-D-glutamate ligase	*murD*	CTL0127	166154100	B0B8Y4	CT758	5.48	46.2	13	9	ND	ND	ND	ND	ND
Membrane Thiol Protease (predicted)		CTL0247	166154214	B0B9A0	CT868	8.24	44.9	3	3	ND	ND	ND	ND	ND
Polymorphic outer membrane protein		CTL0255	166154221	B0B9A7	CT875	5.39	65.8	34	18	0.64	0.61	0.62	0.02	-1.6
Cellular processes														
Thio-specific Antioxidant (TSA) Peroxidase	*ahpC*	CTL0866	166154817	B0B8H7	CT603	4.78	21.7	66	8	0.99	0.94	0.96	0.04	1.0
FHA domain; homology to adenylate cyclase		CTL0033	166154006	B0B8P0	CT664	4.53	89.7	49	16	0.75	0.75	0.75	0.00	1.3
Yop proteins translocation protein L	*sctL*	CTL0824	166154776	B0B8D5	CT561	5.71	24.8	17	6	0.96	0.88	0.92	0.05	1.1
Yop proteins translocation lipoprotein J	*sctJ*	CTL0822	166154774	B0B8D3	CT559	5.24	33.0	23	5	1.24	1.13	1.18	0.07	1.2
Flagellar Motor Switch Domain/YscQ family	*sctQ*	CTL0041	166154014	B0B8P8	CT672	4.53	41.2	19	10	1.11	1.09	1.10	0.01	1.1
Trigger Factor-peptidyl prolyl isomerase	*tig*	CTL0076	166154049	B0B8T3	CT707	5.02	50.1	28	15	1.14	1.18	1.16	0.03	1.2
Superoxide Dismutase (Mn)	*sodM*	CTL0546	166154505	B0B7L0	CT294	6.04	23.4	7	4	1.04	1.06	1.05	0.02	1.0
probable Yop proteins translocation protein C/general secretion pathway protein		CTL0043	166154016	B0B8Q0	CT674	5.43	95.7	30	16	0.87	0.78	0.83	0.06	1.2

Protein Translocase	*secA*	CTL0070	166154043	B0B8S7	CT701	5.65	110.3	33	24	1.11	1.23	1.17	0.08	1.2
GTP Binding Protein		CTL0347	166154313	B0B9K0	CT092	5.23	40.0	29	10	1.13	1.24	1.18	0.08	1.2
Low Calcium Response D	*lcrD*	CTL0345	166154311	B0B9J8	CT090	7.74	77.9	28	17	0.84	0.87	0.86	0.02	1.2
Signal Recognition Particle GTPase	*ffh*	CTL0280	166154246	B0B9D2	CT025	8.56	49.7	14	11	ND	ND	ND	ND	ND
Low Calcium Response E	*copN*	CTL0344	166154310	B0B9J7	CT089	4.9	45.2	17	7	0.75	0.74	0.74	0.01	1.3
Yops secretion ATPase	*sctN*	CTL0038	166154011	B0B8P5	CT669	5.59	48.2	13	9	ND	ND	ND	ND	ND
GTP Binding Protein		CTL0634	166154589	B0B7U8	CT379	7.09	50.8	12	8	ND	ND	ND	ND	ND
Low Calcium Response Protein H	*scc2*	CTL0839	166154791	B0B8F0	CT576	9.1	26.0	16	5	ND	ND	ND	ND	ND
Gen. Secretion Protein D	*gspD*	CTL0835	166154787	B0B8E6	CT572	5.24	81.2	42	21	ND	ND	ND	ND	ND
Protein Translocase		CTL0396	166154362	B0B9P9	CT141	8.61	16.8	6	5	ND	ND	ND	ND	ND
Signal Peptidase I	*lepB*	CTL0275	166154241	B0B9C7	CT020	8.42	71.5	13	6	ND	ND	ND	ND	ND
Flagellum-specific ATP Synthase	*fliI*	CTL0086	166154059	B0B8U3	CT717	6.55	47.6	7	7	ND	ND	ND	ND	ND
Central intermediary metabolism														
Inorganic Pyrophosphatase	*ppa*	CTL0141	166154114	B0B8Z8	CT772	4.75	23.4	7	5	0.78	0.81	0.80	0.02	1.3
Glycogen Synthase	*glgA*	CTL0167	166154140	B0B925	CT798	5.6	53.4	4	4	ND	ND	ND	ND	ND
Glycogen Phosphorylase	*glgP*	CTL0500	166154459	B0B7G4	CT248	5.67	92.7	21	16	ND	ND	ND	ND	ND
Replication														
Molecular chaperone DnaK	*DnaK*	CTL0653	5858788	B0B7W6	CT652	5.03	70.8	290	32	0.86	0.84	0.85	0.01	1.2
RECA Protein	*recA*	CTL0018	166153991	B0B8M5	CT650	7.02	37.8	26	10	1.45	1.48	1.46	0.02	+1.5
Integration Host Factor Alpha	*ihfA*	CTL0519	166154478	B0B7I3	CT267	11.07	11.4	60	12	1.89	1.87	1.88	0.02	+1.9
SWIB (YM74) complex protein		CTL0720	166154674	B0B833	CT460	9.35	9.7	33	12	1.26	1.24	1.25	0.02	1.3
Histone-Like Developmental Protein	*hctA*	CTL0112	166154085	B0B8W9	CT743	10.69	13.7	9	5	0.42	0.40	0.41	0.01	-2.4
DNA Gyrase Subunit B	*gyrB2*	CTL0442	166154400	B0B9U0	CT190	5.49	89.7	25	17	0.92	0.94	0.93	0.01	1.1
DNA Pol III (beta chain)	*dnaN*	CTL0331	166154296	B0B9I3	CT075	5.92	46.5	7	3	ND	ND	ND	ND	ND
DNA Mismatch Repair	*mutS*	CTL0160	166154134	B0B918	CT792	6.69	92.1	32	17	ND	ND	ND	ND	ND

Exinuclease ABC Subunit B	*uvrB*	CTL0849	166154801	B0B8G0	CT586	5.43	75.8	16	14	ND	ND	ND	ND	ND
Holliday Junction Helicase	*ruvB*	CTL0296	166154261	B0B9E8	CT040	7.03	37.3	16	5	0.86	0.84	0.85	0.02	1.2
Crossover Junction Endonuclease	*ruvC*	CTL0764	166154717	B0B876	CT502	9.17	18.7	4	4	ND	ND	ND	ND	ND
DNA Pol III Alpha	*dnaE*	CTL0807	166154760	B0B8B9	CT545	5.71	139.4	38	25	ND	ND	ND	ND	ND
DNA Helicase	*uvrD*	CTL0872	166154823	B0B8I3	CT608	6.08	72.7	23	10	ND	ND	ND	ND	ND
DNA Gyrase Subunit A	*gyrA*	CTL0441	166154399	B0B9T9	CT189	6.27	94.2	38	26	0.90	0.88	0.89	0.02	1.1
SWF/SNF family helicase		CTL0077	166154050	B0B8T4	CT708	5.4	133.1	17	14	ND	ND	ND	ND	ND
DNA Pol III Epsilon Chain	*dnaQ*	CTL0513	166154472	B0B7H7	CT261	5.55	26.5	8	6	1.07	1.02	1.04	0.04	1.0
DNA Ligase	*dnlJ*	CTL0401	166154367	B0B9Q4	CT146	6.61	73.5	20	11	ND	ND	ND	ND	ND
SWI/SNF family helicase		CTL0818	166154770	B0B8C9	CT555	5.67	136.2	22	13	ND	ND	ND	ND	ND
DNA Pol III Epsilon Chain	*dnaQ2*	CTL0798	166154751	B0B8B0	CT536	8.9	28.9	25	13	ND	ND	ND	ND	ND
Primosomal Protein N	*priA*	CTL0147	166154120	B0B904	CT778	9.1	84.8	34	19	ND	ND	ND	ND	ND
Replication Initiation Factor	*dnaA2*	CTL0527	166154486	B0B7J1	CT275	8.67	51.3	21	9	ND	ND	ND	ND	ND
DNA Gyrase Subunit B	*gyrB*	CTL0030	166154003	B0B8N7	CT661	8.62	68.2	18	14	ND	ND	ND	ND	ND
DNA Gyrase Subunit A	*gyrA2*	CTL0029	166154002	B0B8N6	CT660	6.8	55.2	11	9	ND	ND	ND	ND	ND
Exodoxyribonuclease VII	*xseA*	CTL0583	166154539	B0B7P6	CT329	9.58	58.7	12	7	ND	ND	ND	ND	ND
Holliday Junction Helicase	*ruvA*	CTL0763	166154716	B0B875	CT501	6.12	22.2	6	5	ND	ND	ND	ND	ND
DNA Pol III Gamma and Tau	*dnaX*	CTL0588	166154545	B0B7Q2	CT334	6.35	51.6	13	6	ND	ND	ND	ND	ND
DNA repair protein	*radA*	CTL0550	166154509	B0B7L4	CT298	7.18	49.8	12	9	ND	ND	ND	ND	ND
Energy metabolism														
ATP Synthase Subunit E	*atpE*	CTL0562	166154520	B0B7M6	CT310	5.44	22.9	44	12	0.60	0.59	0.60	0.01	-1.7
Transaldolase	*tal*	CTL0565	166154523	B0B7M9	CT313	4.94	36.1	37	11	1.21	1.32	1.27	0.07	1.3
Triosephosphate Isomerase	*tpiS*	CTL0582	166154538	B0B7P5	CT328	5.42	29.8	28	7	0.90	0.92	0.91	0.02	1.1
ATP Synthase Subunit A	*atpA*	CTL0560	166154518	B0B7M4	CT308	5.12	65.5	18	12	0.87	0.87	0.87	0.00	1.1
ADP/ATP Translocase		CTL0321	166154286	B0B9H3	CT065	9.31	58.1	23	13	1.06	1.01	1.04	0.04	1.0
Enolase	*eno*	CTL0850	166154802	B0B8G1	CT587	4.65	45.4	10	7	1.10	0.87	0.99	0.16	1.0

ATP Synthase Subunit B	*atpB*	CTL0559	166154517	B0B7M3	CT307	5.45	48.6	14	10	0.97	0.94	0.96	0.02	1.0
Predicted 1,6-Fructose biphosphate aldolase (dehydrin family)	*dhnA*	CTL0467	166154425	B0B9W5	CT215	6.31	38.0	16	8	1.37	1.33	1.35	0.03	+1.4
Lipoamide Dehydrogenase	*lpdA*	CTL0820	166154772	B0B8D1	CT557	6.33	49.4	11	5	0.91	0.98	0.94	0.05	1.1
Malate Dehydrogenase	*mdhC*	CTL0630	166154586	B0B7U5	CT376	6.08	35.5	8	4	ND	ND	ND	ND	ND
NADH-ubiquinone oxidoreductase, alpha chain	*nqrA*	CTL0002	166153975	B0B8K9	CT634	8.91	51.7	13	9	0.91	0.94	0.93	0.02	1.1
Glucose-6-P Isomerase	*pgi*	CTL0633	166154588	B0B7U7	CT378	5.49	57.7	3	3	ND	ND	ND	ND	ND
Phosphoglucomutase		CTL0187	166154159	B0B944	CT815	6.04	49.3	9	7	ND	ND	ND	ND	ND
Cytochrome Oxidase Subunit I	*cydA*	CTL0268	166154234	B0B9C0	CT013	9.29	50.2	8	6	ND	ND	ND	ND	ND
Phosphoglycerate mutase	*pgmA*	CTL0091	166154064	B0B8U8	CT722	6.66	25.8	13	7	ND	ND	ND	ND	ND
Dihydrolipoamide Acetyltransferase	*pdhC*	CTL0499	166154458	B0B7G3	CT247	5.69	46.4	19	13	ND	ND	ND	ND	ND
Glycerol-3-P Dehydrogenase	*gpdA*	CTL0083	166154056	B0B8U0	CT714	8.19	36.2	12	7	ND	ND	ND	ND	ND
ATP Synthase Subunit D	*atpD*	CTL0558	166154516	B0B7M2	CT306	9.01	23.2	43	10	0.98	1.01	1.00	0.02	1.0
Succinyl-CoA Synthetase, Alpha	*sucD*	CTL0194	166154166	B0B951	CT822	5.34	30.2	6	4	ND	ND	ND	ND	ND
Pyruvate Kinase	*pykF*	CTL0586	166154543	B0B7Q0	CT332	6	53.6	24	10	ND	ND	ND	ND	ND
Fructose-6-P Phosphotransferase	*pfkA*	CTL0457	166154415	B0B9V5	CT205	6.09	61.9	13	8	ND	ND	ND	ND	ND
6-Phosphogluconate Dehydrogenase	*gnd*	CTL0319	166154284	B0B9H1	CT063	5.43	52.6	46	14	ND	ND	ND	ND	ND
Ribulose-P Epimerase	*araD*	CTL0376	166154342	B0B9M9	CT121	4.85	23.0	9	4	ND	ND	ND	ND	ND
Succinate Dehydrogenase	*sdhB*	CTL0854	166154806	B0B8G5	CT591	6.26	25.7	5	3	ND	ND	ND	ND	ND
(pyruvate) Oxoisovalerate Dehydrogenase Alpha and Beta Fusion		CTL0594	166154552	B0B7Q9	CT340	5.47	74.4	5	5	ND	ND	ND	ND	ND
Lipoate Protein Ligase	*lplA*	CTL0761	166154714	B0B873	CT499	7.09	26.9	8	4	ND	ND	ND	ND	ND
Succinate Dehydrogenase	*sdhA*	CTL0855	166154807	B0B8G6	CT592	6.76	67.7	12	12	ND	ND	ND	ND	ND
Phosphoglycerate Kinase	*pgk*	CTL0062	166154035	B0B8R9	CT693	5.65	43.0	10	5	ND	ND	ND	ND	ND
Fatty acid and phospholipid metabolism														
Acyl Carrier Protein	*acpP*	CTL0488	166154447	B0B7F2	CT236	3.86	8.7	13	2	0.67	0.71	0.69	0.03	-1.5
Enoyl-Acyl-Carrier Protein Reductase	*fabI*	CTL0359	166154325	B0B9L2	CT104	5.22	32.0	7	6	1.00	1.12	1.06	0.08	1.1

Biotin Carboxyl Carrier Protein	*accB*	CTL0378	166154344	B0B9N1	CT123	5.07	18.2	8	4	ND	ND	ND	ND	ND
Oxoacyl (Carrier Protein) Reductase	*fabG*	CTL0489	166154448	B0B7F3	CT237	7.7	26.0	14	6	ND	ND	ND	ND	ND
UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	*lpxD*	CTL0495	166154454	B0B7F9	CT243	7.35	38.4	7	6	0.96	0.94	0.95	0.01	1.0
Acylglycerophosphoethanolamine Acyltransferase	*aas*	CTL0145	166154118	B0B902	CT776	7.15	59.4	9	6	1.22	1.06	1.14	0.11	1.1
Acyl Carrier Protein Synthase	*fabF*	CTL0139	166154112	B0B8Z6	CT770	5.47	44.8	21	9	1.17	1.25	1.21	0.06	1.2
Biotin Carboxylase 4	*accC*	CTL0379	166154345	B0B9N2	CT124	6.4	50.1	12	9	ND	ND	ND	ND	ND
AcCoA Carboxylase/Transferase Alpha	*accA*	CTL0517	166154476	B0B7I1	CT265	5.91	36.4	17	10	ND	ND	ND	ND	ND
Lipid A Disaccharide Synthase	*lpxB*	CTL0668	166154622	B0B7Y1	CT411	9.23	69.1	10	6	ND	ND	ND	ND	ND
Acyl-Carrier UDP-GlcNAc O-Acyltransferase	*lpxA*	CTL0793	166154746	B0B8A5	CT531	6.08	30.7	9	7	ND	ND	ND	ND	ND
AcCoA Carboxylase/Transferase Beta	*accD*	CTL0545	166154504	B0B7K9	CT293	7.52	33.7	32	9	0.76	0.81	0.78	0.04	1.3
Hydroxymyristoyl-(acyl carrier protein) dehydratase	*fabZ*	CTL0794	166154747	B0B8A6	CT532	9.19	16.6	5	3	ND	ND	ND	ND	ND
Hypothetical proteins														
hypothetical protein		CTL0847	166154799	B0B8F8	CT584	5.61	21.1	82	7	1.00	1.03	1.02	0.02	1.0
hypothetical protein		CTL0874	166154825	B0B8I5	CT610	4.94	26.8	62	13	1.02	0.93	0.98	0.06	1.0
hypothetical protein		CTL0040	166154013	B0B8P7	CT671	4.79	31.0	40	10	0.93	0.94	0.94	0.00	1.1
hypothetical protein		CTL0028	166154001	B0B8N5	CT659	7.99	8.8	26	5	0.98	0.97	0.97	0.01	1.0
hypothetical protein		CTL0512	166154471	B0B7H6	CT260	4.76	18.9	44	9	0.85	0.87	0.86	0.01	1.2
hypothetical protein	*copD*	CTL0842	166154794	B0B8F3	CT579	9.47	44.0	15	11	0.89	0.91	0.90	0.02	1.1
hypothetical protein		CTL0322	166154287	B0B9H4	CT066	9.95	17.9	39	9	1.65	1.79	1.72	0.10	+1.7
hypothetical protein		CTL0034	166154007	B0B8P1	CT665	9.1	9.3	12	4	1.30	1.24	1.27	0.04	1.3
hypothetical protein		CTL0137	166154110	B0B8Z4	CT768	5.41	64.1	40	18	0.90	0.91	0.90	0.00	1.1
hypothetical protein		CTL0800	166154753	B0B8B2	CT538	5.33	27.5	12	8	1.03	1.04	1.04	0.01	1.0
hypothetical protein		CTL0655	166154609	B0B7W8	CT398	7.07	29.5	20	11	1.06	1.09	1.08	0.03	1.1
hypothetical protein		CTL0688	166154642	B0B801	CT429	5.1	39.2	16	9	0.95	0.93	0.94	0.02	1.1
hypothetical protein		CTL0299	166154264	B0B9F1	CT043	5.04	18.3	15	7	1.12	1.14	1.13	0.02	1.1

hypothetical protein	CTL0840	166154792	B0B8F1	CT577	6.51	13.3	5	3	0.24	0.19	0.21	0.03	-4.7
hypothetical protein	CTL0036	166154009	B0B8P3	CT667	4.87	16.5	8	3	1.55	1.45	1.50	0.07	+1.5
hypothetical protein	CTL0110	166154083	B0B8W7	CT741	9.45	10.4	12	4	1.00	1.16	1.08	0.11	1.1
hypothetical protein	CTL0402	166154368	B0B9Q5	CT147	8.75	162.2	89	53	1.03	0.98	1.01	0.03	1.0
hypothetical protein	CTL0103	166154076	B0B8W0	CT734	8.47	20.6	11	6	0.91	0.91	0.91	0.00	1.1
hypothetical protein	CTL0222	166154194	B0B979	CT849.1	4.01	6.8	12	4	0.75	0.75	0.75	0.00	1.3
hypothetical protein	CTL0272	166154238	B0B9C4	CT017	6.64	47.7	22	14	0.60	0.63	0.61	0.02	-1.6
hypothetical protein	CTL0626	166154583	B0B7U1	CT372	9.04	49.4	11	5	0.91	0.93	0.92	0.01	1.1
hypothetical protein	CTL0897	166154847	B0B8K7	CT632	5.83	60.9	18	13	0.90	0.88	0.89	0.01	1.1
hypothetical protein	CTL0045	166154018	B0B8Q2	CT676	5.54	19.8	6	2	1.05	1.10	1.08	0.03	1.1
hypothetical protein	CTL0589	166154546	B0B7Q3	CT335	5.16	10.5	5	2	0.78	0.79	0.79	0.01	1.3
hypothetical protein	CTL0060	166154033	B0B8R7	CT691	4.95	25.2	8	6	1.54	1.38	1.46	0.11	+1.5
hypothetical protein	CTL0271	166154237	B0B9C3	CT016	4.74	26.7	5	4	ND	ND	ND	ND	ND
hypothetical protein	CTL0541	166154500	B0B7K5	CT289	9.82	41.8	24	10	1.00	0.97	0.98	0.02	1.0
hypothetical protein	CTL0037	166154010	B0B8P4	CT668	4.6	24.4	10	8	0.99	0.87	0.93	0.08	1.1
hypothetical protein	CTL0238	166154209	B0B995	CT863	5.27	53.6	4	4	0.63	0.67	0.65	0.03	-1.5
hypothetical protein	CTL0885	166154836	B0B8J6	CT621	4.88	92.6	27	12	ND	ND	ND	ND	ND
hypothetical protein	CTL0540	166154499	B0B7K4	CT288	8.38	63.5	21	13	0.49	0.49	0.49	0.00	-2.0
hypothetical protein	CTL0097	166154070	B0B8V4	CT728	6.24	27.9	4	3	ND	ND	ND	ND	ND
hypothetical protein	CTL0063	166154036	B0B8S0	CT694	5.11	34.7	3	2	ND	ND	ND	ND	ND
hypothetical protein	CTL0463	166154421	B0B9W1	CT211	4.61	20.8	4	4	ND	ND	ND	ND	ND
hypothetical protein	CTL0305	166154270	B0B9F7	CT049	5.62	50.8	43	8	ND	ND	ND	ND	ND
hypothetical protein	CTL0087	166154060	B0B8U4	CT718	5.89	19.6	3	2	ND	ND	ND	ND	ND
hypothetical protein	CTL0791	166154744	B0B8A3	CT529	9.41	31.2	5	5	0.78	0.76	0.77	0.02	1.3
hypothetical protein	CTL0297	166154262	B0B9E9	CT041	4.84	27.6	6	4	ND	ND	ND	ND	ND
hypothetical protein	CTL0266	166154232	B0B9B8	CT011	6.36	44.7	10	10	0.88	0.82	0.85	0.04	1.2
hypothetical protein	CTL0507	166154466	B0B7H1	CT255	4.79	14.4	4	2	ND	ND	ND	ND	ND

hypothetical protein		CTL0643	166154598	B0B7V7	CT387	5.81	77.1	36	13	ND	ND	ND	ND	ND
hypothetical protein		CTL0102	166154075	B0B8V9	CT733	8.93	47.1	7	7	ND	ND	ND	ND	ND
hypothetical protein		CTL0293	166154258	B0B9E5	CT038	9.54	13.5	3	3	ND	ND	ND	ND	ND
hypothetical protein		CTL0605	166154563	B0B7S0	CT351	9.17	78.3	31	11	ND	ND	ND	ND	ND
hypothetical protein		CTL0563	166154521	B0B7M7	CT311	9.07	23.9	5	5	ND	ND	ND	ND	ND
hypothetical protein		CTL0609	166154567	B0B7S4	CT355	5.44	37.4	20	10	ND	ND	ND	ND	ND
hypothetical protein		CTL0314	166154279	B0B9G6	CT058	9.37	39.7	11	4	1.17	1.51	1.34	0.24	1.3
hypothetical protein		CTL0039	166154012	B0B8P6	CT670	7.83	20.2	21	11	ND	ND	ND	ND	ND
hypothetical protein		CTL0309	166154274	B0B9G1	CT053	4.41	17.2	10	5	ND	ND	ND	ND	ND
hypothetical protein		CTL0742	166154695	B0B854	CT481	8.46	27.9	5	3	ND	ND	ND	ND	ND
hypothetical protein	*rmuC*	CTL0197	166154169	B0B954	CT825	6.55	48.5	13	9	ND	ND	ND	ND	ND
hypothetical protein		CTL0891	166154842	B0B8K2	CT627	6.04	37.7	15	6	ND	ND	ND	ND	ND
hypothetical protein		CTL0477	166154435	B0B9X5	CT224	5.22	15.9	13	5	ND	ND	ND	ND	ND
hypothetical protein		CTL0853	166154805	B0B8G4	CT590	5.52	109.0	23	14	ND	ND	ND	ND	ND
hypothetical protein		CTL0065	166154038	B0B8S2	CT696	6.34	45.7	8	6	ND	ND	ND	ND	ND
hypothetical protein		CTL0433	166154391	B0B9T1	CT181	5.22	24.2	27	10	ND	ND	ND	ND	ND
hypothetical protein		CTL0010	166153983	B0B8L7	CT642	9.16	32.1	2	2	ND	ND	ND	ND	ND
hypothetical protein		CTL0577	5858294	B0B7P1	CT326	8.76	16.9	21	5	ND	ND	ND	ND	ND
hypothetical protein		CTL0408	166154374	B0B9R1	CT153	6.17	90.8	16	11	ND	ND	ND	ND	ND
hypothetical protein		CTL0883	166154834	B0B8J4	CT619	4.75	96.7	10	10	ND	ND	ND	ND	ND
hypothetical protein		CTL0003	166153976	B0B8L0	CT635	6.17	16.7	8	5	ND	ND	ND	ND	ND
hypothetical protein		CTL0369	166154335	B0B9M2	CT114	7.91	55.2	16	13	ND	ND	ND	ND	ND
hypothetical protein		CTL0561	166154519	B0B7M5	CT309	5.36	32.1	9	6	ND	ND	ND	ND	ND
hypothetical protein		CTL0496	166154455	B0B7G0	CT244	8.32	45.6	7	6	ND	ND	ND	ND	ND
hypothetical protein		CTL0640	166154595	B0B7V4	CT384	5.84	59.8	9	4	ND	ND	ND	ND	ND

Other, categories														
FKBP-type peptidyl-prolyl cis-trans isomerase	*mip*	CTL0803	166154756	B0B8B5	CT541	4.83	24.5	46	13	0.87	0.82	0.85	0.04	1.2
hydrolase/phosphatase homolog	*hsp60_1*	CTL0140	166154113	B0B8Z7	CT771	5.04	17.4	21	6	0.89	0.86	0.87	0.02	1.1
CHLPN 76kDa Homolog		CTL0887	166154838	B0B8J8	CT623	6.43	45.8	42	18	0.91	0.91	0.91	0.00	1.1
CHLPN 76kDa Homolog		CTL0886	166154837	B0B8J7	CT622	4.93	68.9	32	14	0.52	0.47	0.49	0.03	-2.0
Hit Family Hydrolase		CTL0641	166154596	B0B7V5	CT385	5.43	12.3	16	6	1.28	1.19	1.23	0.07	1.2
predicted metal dependent hydrolase (histidinic triad)	*phnP*	CTL0635	166154590	B0B7U9	CT380	5.79	30.1	6	4	0.81	0.89	0.85	0.06	1.2
hypothetical protein	*ltuB*	CTL0336	166154301	B0B9I8	CT080	9.84	11.3	2	2	ND	ND	ND	ND	ND
phosphohydrolase		CTL0721	166154675	B0B834	CT461	9.27	37.0	7	6	ND	ND	ND	ND	ND
ACR family		CTL0363	166154329	B0B9L6	CT108	5.68	27.5	2	2	ND	ND	ND	ND	ND
SurE-like Acid Phosphatase	*surE*	CTL0470	166154428	B0B9W8	CT218	4.84	31.5	9	6	1.14	1.23	1.19	0.06	1.2
Leucine Dehydrogenase	*ldh*	CTL0142	166154115	B0B8Z9	CT773	5.21	37.3	9	6	ND	ND	ND	ND	ND
Dpredicted metal dependent hydrolase		CTL0642	166154597	B0B7V6	CT386	5.14	33.1	3	2	ND	ND	ND	ND	ND
utQ/KpsF Family Sugar-P Isomerase		CTL0656	166154610	B0B7W9	CT399	5.45	36.0	12	4	ND	ND	ND	ND	ND
Hypothetical protein containing CBS domains		CTL0508	166154467	B0B7H2	CT256	6.44	41.5	6	4	ND	ND	ND	ND	ND
hypothetical protein		CTL0648	166154603	B0B7W2	CT392	5.26	41.3	9	8	ND	ND	ND	ND	ND
hypothetical protein	*gcpE*	CTL0313	166154278	B0B9G5	CT057	5.69	66.5	13	9	ND	ND	ND	ND	ND
PTS IIA Protein	*ptsN_2*	CTL0543	166154502	B0B7K7	CT291	4.86	17.1	31	3	0.92	1.27	1.10	0.25	1.1
predicted polysaccharide hydrolase-invasin repeat family		CTL0382	166154348	B0B9N5	CT127	8.75	31.7	5	4	ND	ND	ND	ND	ND
Methylase		CTL0748	166154701	B0B860	CT487	8.56	20.9	2	2	ND	ND	ND	ND	ND
Monooxygenase	*mhpA*	CTL0403	166154369	B0B9Q6	CT148	8.33	57.8	35	14	ND	ND	ND	ND	ND
YhgN family		CTL0225	166154197	B0B982	CT852	9.87	23.7	2	2	ND	ND	ND	ND	ND
PTS PEP Phosphotransferase	*ptsI*	CTL0590	166154547	B0B7Q4	CT336	5.78	63.7	11	7	ND	ND	ND	ND	ND
Arginine Kinase	*aspC*	CTL0044	166154017	B0B8Q1	CT675	5.96	40.1	18	5	ND	ND	ND	ND	ND

Purines, pyrimidines, nucleosides and nucleotides														
UMP Kinase	*pyrH*	CTL0047	166154020	B0B8Q4	CT678	5.29	26.1	10	5	0.98	0.89	0.94	0.06	1.1
Thioredoxin	*trxA*	CTL0801	166154754	B0B8B3	CT539	5.02	11.2	3	3	ND	ND	ND	ND	ND
CMP Kinase	*cmk*	CTL0712	166154666	B0B825	CT452	5.13	24.0	16	7	0.88	0.87	0.87	0.01	1.1
adenylate kinase	*adk*	CTL0383	166154349	B0B9N6	CT128	4.85	27.7	15	5	0.94	0.90	0.92	0.02	1.1
Ribonucleoside Reductase, Large Chain	*nrdA*	CTL0199	166154171	B0B956	CT827	5.74	119.3	37	20	1.17	1.15	1.16	0.01	1.2
Thymidylate Kinase	*tdk*	CTL0440	166154398	B0B9T8	CT188	6.53	22.4	4	3	ND	ND	ND	ND	ND
CTP Synthetase	*pyrG*	CTL0435	166154393	B0B9T3	CT183	5.68	58.0	12	11	ND	ND	ND	ND	ND
dUTP Nucleotidohydrolase	*dut*	CTL0544	166154503	B0B7K8	CT292	5.18	15.3	3	2	ND	ND	ND	ND	ND
Ribonucleoside Reductase, Small Chain	*nrdB*	CTL0200	166154172	B0B957	CT828	5.33	40.5	10	8	ND	ND	ND	ND	ND
Regulatory functions														
General Stress Protein	*rplY*	CTL0168	166154141	B0B926	CT799	8.99	20.4	4	3	ND	ND	ND	ND	ND
Transcription														
RNA Polymerase Alpha	*rpoA*	CTL0769	166154722	B0B881	CT507	5.34	41.8	44	15	0.99	0.98	0.99	0.01	1.0
RNA Polymerase Beta	*rpoB*	CTL0567	166154525	B0B7N1	CT315	5.63	140.0	127	52	1.02	0.98	1.00	0.03	1.0
RNA Polymerase Beta	*rpoC*	CTL0566	166154524	B0B7N0	CT314	7.17	154.7	88	52	0.94	0.93	0.93	0.00	1.1
Transcription antitermination factor	*nusA*	CTL0352	166154318	B0B9K5	CT097	5.19	48.9	67	15	1.44	1.69	1.56	0.18	+1.6
RNA Polymerase Sigma-66	*rpoD*	CTL0879	166154830	B0B8J0	CT615	8.09	66.1	86	26	0.94	0.95	0.95	0.00	1.1
Sigma Regulatory Factor	*rsbV*	CTL0683	166154637	B0B7Z6	CT424	5.23	12.5	14	3	0.96	0.96	0.96	0.00	1.0
Transcription Termination Factor	*rho*	CTL0752	166154705	B0B864	CT491	6.84	51.6	32	18	0.84	0.82	0.83	0.02	1.2
Transcription Elongation Factor G	*greA*	CTL0004	166153977	B0B8L1	CT636	5.28	80.9	38	27	1.18	1.18	1.18	0.00	1.2
Transcriptional termination protein	*nusG*	CTL0572	166154530	B0B7N6	CT320	5.26	20.7	9	9	0.89	1.24	1.06	0.25	1.1
Methionyl tRNA Formyltransferase	*fmt*	CTL0792	166154745	B0B8A4	CT530	8.87	33.9	20	9	ND	ND	ND	ND	ND

Polyribonucleotide Nucleotidyltransferase	*pnp*	CTL0214	166154186	B0B971	CT842	5.66	75.4	19	11	1.25	1.26	1.25	0.01	1.3
tRNA (guanine N-1) Methyltransferase	*trmD*	CTL0282	166154248	B0B9D4	CT027	5.6	39.8	7	4	ND	ND	ND	ND	ND
rRNA Methyltransferse		CTL0111	166154084	B0B8W8	CT742	6.71	44.6	37	15	ND	ND	ND	ND	ND
Poly A Polymerase	*pcnB*	CTL0073	166154046	B0B8T0	CT704	7.76	46.5	7	7	ND	ND	ND	ND	ND
Ribonuclease III	*rnc*	CTL0549	166154508	B0B7L3	CT297	6.07	25.6	6	4	ND	ND	ND	ND	ND
Ribonuclease HII	*rnhC*	CTL0263	166154229	B0B9B5	CT008	9.07	33.1	2	2	ND	ND	ND	ND	ND
predicted pseudouridine synthetase family		CTL0361	166154327	B0B9L4	CT106	9.41	31.4	3	2	ND	ND	ND	ND	ND
Sigma Regulatory Factor	*rbsV*	CTL0134	166154107	B0B8Z1	CT765	7.69	12.4	6	6	ND	ND	ND	ND	ND
2-component regulatory system-ATPase	*atoC*	CTL0728	166154682	B0B841	CT468	6.24	43.1	6	5	ND	ND	ND	ND	ND
Pseudouridine Synthase	*sfhB*	CTL0027	166154000	B0B8N4	CT658	6.77	37.8	7	7	ND	ND	ND	ND	ND
Translation														
HSP-60	*hsp60_1*	CTL0365	166154331	B0B9L8	CT110	5.26	58.1	239	29	0.70	0.69	0.69	0.01	-1.4
10KDa Chaperonin	*groES*	CTL0366	166154332	B0B9L9	CT111	4.85	11.2	273	6	0.79	0.69	0.74	0.07	1.3
Elongation Factor Tu	*tufA*	CTL0574	166154532	B0B7N8	CT322	5.44	43.3	170	21	1.08	1.05	1.06	0.02	1.1
DO Serine Protease	*htrA*	CTL0195	166154167	B0B952	CT823	5.89	51.4	71	21	0.81	0.79	0.80	0.02	1.2
Molecular chaperone DnaK	*dnaK*	CTL0652	166154607	B0B7W6	CT396	5.03	70.8	64	11	0.94	0.86	0.90	0.05	1.1
S15 Ribosomal Protein	*rpsO*	CTL0215	166154187	B0B972	CT843	9.74	10.4	51	7	1.11	0.99	1.05	0.08	1.1
ClpC Protease ATPase	*clpC*	CTL0538	166154497	B0B7K2	CT286	6.03	95.2	76	35	1.06	1.04	1.05	0.02	1.0
Heat Shock Protein J	*dnaJ*	CTL0595	166154553	B0B7R0	CT341	7.54	41.9	50	24	0.92	0.88	0.90	0.03	1.1
Elongation Factor TS	*tsf*	CTL0048	166154021	B0B8Q5	CT679	5.65	30.9	56	17	1.41	1.51	1.46	0.07	+1.5
Thioredoxin Disulfide Isomerase		CTL0149	166154122	B0B906	CT780	7.93	16.2	21	5	1.45	1.35	1.40	0.07	+1.4
Alanyl tRNA Synthetase	*alaS*	CTL0118	166154091	B0B8X5	CT749	5.53	97.6	41	19	1.05	1.07	1.06	0.01	1.1
HSP-70 Cofactor	*grpE*	CTL0651	166154606	B0B7W5	CT395	4.62	21.7	17	5	0.66	0.71	0.69	0.04	-1.5
Clp Protease ATPase	*clpB*	CTL0368	166154334	B0B9M1	CT113	5.36	96.6	63	34	1.13	1.14	1.14	0.01	1.1
L21 Ribosomal Protein	*rplU*	CTL0677	166154631	B0B7Z0	CT420	9.27	12.2	9	3	0.96	0.90	0.93	0.04	1.1

S1 Ribosomal Protein	*rpsA*	CTL0353	166154319	B0B9K6	CT098	5.2	63.4	22	15	0.79	0.89	0.84	0.07	1.2
ATP-dependent zinc protease	*ftsH*	CTL0213	166154185	B0B970	CT841	5.78	98.1	37	25	0.85	0.87	0.86	0.01	1.2
DnaK Suppressor	*dksA*	CTL0664	166154618	B0B7X7	CT407	5.1	13.9	13	4	1.04	1.11	1.08	0.05	1.1
L7/L12 Ribosomal Protein	*rplL*	CTL0568	166154526	B0B7N2	CT316	4.9	13.6	88	9	1.03	0.98	1.00	0.04	1.0
S2 Ribosomal Protein	*rpsB*	CTL0049	166154022	B0B8Q6	CT680	6.55	31.2	42	16	1.11	1.07	1.09	0.03	1.1
L5 Ribosomal Protein	*rplE*	CTL0778	166154731	B0B890	CT516	9.44	20.5	18	8	0.82	0.74	0.78	0.06	1.3
L9 Ribosomal Protein	*rplI*	CTL0172	166154145	B0B930	CT803	6.11	18.4	15	5	0.91	1.00	0.95	0.07	1.0
Insulinase family/Protease III	*ptr*	CTL0175	166154148	B0B933	CT806	5.05	104.4	35	22	1.23	1.16	1.20	0.04	1.2
Glu-tRNA Gln Amidotransferase (B Subunit)	*gatB*	CTL0259	166154225	B0B9B1	CT004	5.92	54.9	17	14	1.16	1.37	1.26	0.15	1.3
L10 Ribosomal Protein	*rplJ*	CTL0569	166154527	B0B7N3	CT317	5.98	18.9	11	5	0.99	0.95	0.97	0.03	1.0
L4 Ribosomal Protein	*rplD*	CTL0789	166154742	B0B8A1	CT527	9.79	24.6	11	5	1.33	1.28	1.31	0.04	1.3
Protease	*sohB*	CTL0755	166154708	B0B867	CT494	6.51	31.8	8	4	1.04	0.96	1.00	0.06	1.0
CLP Protease	*clpP*	CTL0690	166154644	B0B803	CT431	5.21	20.9	10	5	0.69	0.70	0.69	0.01	-1.4
L13 Ribosomal Protein	*rplM*	CTL0380	166154346	B0B9N3	CT125	10.19	16.8	8	3	1.59	1.32	1.45	0.19	+1.5
L16 Ribosomal Protein	*rplP*	CTL0783	166154736	B0B895	CT521	11.3	15.8	10	5	0.92	0.95	0.93	0.02	1.1
L1 Ribosomal Protein	*rplA*	CTL0570	166154528	B0B7N4	CT318	9.03	24.7	25	10	0.81	0.83	0.82	0.01	1.2
L11 Ribosomal Protein	*rplK*	CTL0571	166154529	B0B7N5	CT319	9.68	15.0	7	6	1.10	1.12	1.11	0.01	1.1
S5 Ribosomal Protein	*rpsE*	CTL0774	166154727	B0B886	CT512	9.87	17.8	11	6	1.18	1.21	1.19	0.02	1.2
S3 Ribosomal Protein	*rpsC*	CTL0784	166154737	B0B896	CT522	10.03	24.3	10	5	1.14	1.12	1.13	0.02	1.1
Tryptophanyl tRNA Synthetase	*trpS*	CTL0848	166154800	B0B8F9	CT585	6.51	39.4	10	7	1.02	0.87	0.94	0.10	1.1
Lon ATP-dependent protease	*lon*	CTL0598	166154556	B0B7R3	CT344	6.9	91.9	59	31	0.82	0.80	0.81	0.02	1.2
Initiation Factor-2	*infA*	CTL0351	166154317	B0B9K4	CT096	8.21	97.3	41	25	1.31	1.23	1.27	0.06	1.3
Leucyl Aminopeptidase A		CTL0301	166154266	B0B9F3	CT045	5.73	54.2	43	23	0.99	0.97	0.98	0.02	1.0
L3 Ribosomal Protein	*rplC*	CTL0790	166154743	B0B8A2	CT528	9.72	23.5	16	10	1.37	1.23	1.30	0.10	1.3
Glu-tRNA Gln Amidotransferase (A subunit)	*gatA*	CTL0258	166154224	B0B9B0	CT003	5.87	53.6	18	14	1.23	1.18	1.20	0.03	1.2
S9 Ribosomal Protein	*rpsI*	CTL0381	166154347	B0B9N4	CT126	11.02	14.5	7	4	ND	ND	ND	ND	ND
Aspartyl tRNA Synthetase	*aspS*	CTL0804	166154757	B0B8B6	CT542	5.31	66.2	43	21	1.05	1.06	1.06	0.01	1.1

Glycyl tRNA Synthetase	*glyQ*	CTL0165	166154138	B0B923	CT796	5.66	112.4	23	14	0.88	0.95	0.91	0.05	1.1
Glu-tRNA Gln Amidotransferase (C subunit)	*gatC*	CTL0257	166154223	B0B9A9	CT002	4.14	11.1	3	2	ND	ND	ND	ND	ND
tyrosyl tRNA Synthetase	*tyrS*	CTL0318	166154283	B0B9H0	CT062	6.62	45.4	25	11	1.51	1.52	1.52	0.00	+1.5
S8 Ribosomal Protein	*rpsH*	CTL0777	166154730	B0B889	CT515	10.27	15.1	9	7	ND	ND	ND	ND	ND
Initiation Factor 3	*infC*	CTL0205	166154177	B0B962	CT833	9.5	21.1	11	8	ND	ND	ND	ND	ND
S13 Ribosomal Protein	*rpsM*	CTL0771	166154724	B0B883	CT509	11.01	13.9	12	4	0.86	0.91	0.88	0.04	1.1
L19 Ribosomal Protein	*rplS*	CTL0283	166154249	B0B9D5	CT028	9.95	13.1	9	6	ND	ND	ND	ND	ND
Ribosome Releasing Factor	*rrf*	CTL0046	166154019	B0B8Q3	CT677	8.59	20.0	9	7	ND	ND	ND	ND	ND
Peptide Chain Releasing Factor (RF-1)	*prfA*	CTL0278	166154244	B0B9D0	CT023	5.29	40.0	7	5	ND	ND	ND	ND	ND
Elongation Factor P	*efp*	CTL0121	166154094	B0B8X8	CT752	4.74	20.2	12	6	ND	ND	ND	ND	ND
L6 Ribosomal Protein	*rplF*	CTL0776	166154729	B0B888	CT514	9.96	19.8	10	9	0.94	0.87	0.91	0.05	1.1
S10 Ribosomal Protein	*rpsJ*	CTL0695	166154649	B0B808	CT436	10.47	11.9	13	6	1.31	1.41	1.36	0.07	+1.4
L24 Ribosomal Protein	*rplX*	CTL0779	166154732	B0B891	CT517	10.44	12.6	11	8	ND	ND	ND	ND	ND
Metalloprotease	*ispH*	CTL0234	166154205	B0B991	CT859	6.1	34.2	9	7	0.86	0.79	0.83	0.05	1.2
S6 Ribosomal Protein	*rpsF*	CTL0170	166154143	B0B928	CT801	8.75	12.9	4	2	ND	ND	ND	ND	ND
L18 Ribosomal Protein	*rplR*	CTL0775	166154728	B0B887	CT513	10.32	13.4	3	2	1.13	1.03	1.08	0.07	1.1
L22 Ribosomal Protein	*rplV*	CTL0785	166154738	B0B897	CT523	11.34	12.4	3	3	0.98	1.22	1.10	0.17	1.1
S4 Ribosomal Protein	*rpsD*	CTL0890	166154841	B0B8K1	CT626	10.01	23.6	13	9	1.00	1.01	1.01	0.01	1.0
L17 Ribosomal Protein	*rplQ*	CTL0768	166154721	B0B880	CT506	11.27	16.1	19	9	ND	ND	ND	ND	ND
Metalloprotease		CTL0328	166154293	B0B9I0	CT072	6.34	66.3	28	14	1.07	1.17	1.12	0.07	1.1
Histidyl tRNA Synthetase	*hisS*	CTL0805	166154758	B0B8B7	CT543	6.65	49.1	15	7	ND	ND	ND	ND	ND
S11 Ribosomal Protein	*rpsK*	CTL0770	166154723	B0B882	CT508	11.26	13.8	9	5	ND	ND	ND	ND	ND
Threonyl tRNA Synthetase	*thrS*	CTL0844	166154796	B0B8F5	CT581	5.68	72.6	10	7	ND	ND	ND	ND	ND
Leucyl tRNA Synthetase	*leuS*	CTL0461	166154419	B0B9V9	CT209	5.41	92.8	22	14	ND	ND	ND	ND	ND
Oligoendopeptidase	*pepF*	CTL0367	166154333	B0B9M0	CT112	5.6	69.0	25	10	ND	ND	ND	ND	ND
HSP-60	*groEL 2*	CTL0867	166154818	B0B8H8	CT604	5.05	58.8	22	8	ND	ND	ND	ND	ND

Lysyl tRNA Synthetase	*lysS*	CTL0150	166154123	B0B907	CT781	5.33	60.0	19	10	ND	ND	ND	ND	ND
L28 Ribosomal Protein	*rpmB*	CTL0341	166154307	B0B9J4	CT086	11.65	10.1	6	4	ND	ND	ND	ND	ND
Isoleucyl-tRNA Synthetase	*ileS*	CTL0274	166154240	B0B9C6	CT019	5.36	118.7	9	9	ND	ND	ND	ND	ND
Transport and binding proteins														
Solute Protein Binding Family		CTL0323	166154288	B0B9H5	CT067	5.01	33.3	48	11	0.73	0.75	0.74	0.02	-1.4
ABC Transporter ATPase		CTL0054	166154027	B0B8R1	CT685	5.14	28.4	14	9	0.96	0.99	0.97	0.02	1.0
oligopeptide Binding Lipoprotein	*oppA4*	CTL0741	166154694	B0B853	CT480	5	77.5	26	14	1.48	1.44	1.46	0.03	+1.5
Arginine Binding Protein	*artJ*	CTL0636	166154591	B0B7V0	CT381	4.89	25.6	12	8	1.17	1.20	1.19	0.02	1.2
ABC Transporter Membrane Protein		CTL0055	166154028	B0B8R2	CT686	6.03	44.6	8	4	0.83	0.86	0.84	0.03	1.2
Metal Transport P-type ATPase		CTL0096	166154069	B0B8V3	CT727	6.83	70.4	10	7	ND	ND	ND	ND	ND
Solute-binding protein		CTL0672	166154626	B0B7Y5	CT415	6.86	29.4	2	2	ND	ND	ND	ND	ND
ABC Amino Acid Transporter ATPase		CTL0385	166154351	B0B9N8	CT130	5.75	25.5	12	3	ND	ND	ND	ND	ND
polysaccharide transporter	*exbB*	CTL0860	166154810	B0B8H0	CT596	9.21	25.9	2	2	ND	ND	ND	ND	ND
ABC Amino Acid Transporter Permease		CTL0384	166154350	B0B9N7	CT129	9.03	23.6	6	3	ND	ND	ND	ND	ND
Mg++ Transporter (CBS Domain)	*mgtE*	CTL0446	166154404	B0B9U4	CT194	4.94	51.4	20	8	ND	ND	ND	ND	ND
Transport ATP Binding Protein		CTL0516	166154475	B0B7I0	CT264	8.09	68.3	17	10	ND	ND	ND	ND	ND
Oligopeptide Permease	*oppC2*	CTL0739	166154692	B0B851	CT478	8.83	61.6	26	11	ND	ND	ND	ND	ND
ABC Transport ATPase (Nitrate/Fe)		CTL0432	166154390	B0B9T0	CT180	6.6	25.6	28	7	ND	ND	ND	ND	ND
Oligopeptide Transport ATPase	*oppD*	CTL0453	166154411	B0B9V1	CT201	6.75	30.7	12	5	ND	ND	ND	ND	ND
Translocated actin-recruiting phosphoprotein	*tarp*	CTL0716	166154670	Q6GX35	CT456	4.1	103.2	22	15	0.88	0.98	0.93	0.07	1.1
Plasmid														
Putative uncharacterized protein		pL2-05	5857580	B0BCM1		4.78	27.9	38	7	0.92	0.90	0.91	0.01	1.1
Putative uncharacterized protein		pL2-07a	5857577	B0BCL9		6.25	14.3	7	4	ND	ND	ND	ND	ND
Putative uncharacterized protein		pL2-07	5857574	B0BCM7		5.63	15.2	7	2	ND	ND	ND	ND	ND

Virulence plasmid integrase pGP8-D	pL2-02	5857576	B0BCM4	10.05	37.8	24	15	ND	ND	ND	ND	ND
Putative uncharacterized protein	pL2-06	5857575	B0BCM0	9.33	11.8	10	6	ND	ND	ND	ND	ND

a) pI & molecular mass were obtained using 'Compute MW/p*I* ' (ExPASy bioinformatics resource portal).

b) Reporter ion tags correspond to the following samples: 114 = EB, 115= RB, 116=EB, 117=RB.

c) Values with (-) are down-regulated below the 1.4 fold threshold and values with (+) are up-regulated above the 1.4 fold threshold.

Appendix I, Table 5.2. Proteins assigned and quantified in EBs and RBs from *C. trachomatis* L2\434\Bu using the label-free approach.

UniProt accession/ Category	Locus	Gene name	Protein description	MW kDa[a]	p*I*[b]	Total peptides[c]	Total unique peptides[d]	Mean RB (mol/cell)[e]	RB (N)	STDev[f]	Mean EB (mol/cell)[e]	EB (N)	STDev[f]	EB/RB ratio
Amino Acid Biosynthesis														
B0B8L2	CTL0005	*aspC*	Aspartate aminotransferase	44737	5.4	167	22	21	3	11	15	1	ND	0.7
B0B9S1	CTL0423	*trpB*	tryptophan synthase beta chain	42618	6.8	71	11	25	1	ND	13	2	2	0.5
B0B9S2	CTL0424	*trpA*	tryptophan synthase alpha chain	28101	4.8	112	14	67	4	46	53	4	22	0.8
B0B7J8	CTL0534	*gcsH*	glycine cleavage system H protein	13147	4.5	19	4	38	1	ND	17	3	4	0.4
B0B7T0	CTL0615	*dapA*	dihydrodipicolinate synthase	31262	5.5	85	13	61	4	32	31	4	7	0.5
B0B7T1	CTL0616	*lysC*	aspartokinase	47553	5.1	88	13	26	3	12	21	2	15	0.8
B0B7T3	CTL0618	*dapB*	dihydrodipicolinate reductase	27789	5.8	97	12	49	4	22	28	4	8	0.6
B0B7T8	CTL0623	*aroB*	3-dehydroquinate synthase	41183	7.7	31	3	0	0	ND	18	1	ND	EB
B0B7T9	CTL0624	*aroE*	shikimate 5-dehydrogenase	53252	8.8	84	14	35	1	ND	20	1	ND	0.6
B0B804	CTL0691	*glyA*	serine hydroxymethyltransferase	54249	6.0	218	15	65	4	4	48	2	4	0.7
B0B9Q3	CTL0400	*pkn1*	serine-threonine-protein kinase	69638	5.1	198	32	70	4	37	34	4	11	0.5
Biosynthesis of Cofactors														
B0B8K8	CTL0001	*hemB*	delta-aminolevulinic acid dehydratase	37726	5.9	39	4	19	1	ND	0	0	ND	RB
B0B8M4	CTL0017		conserved hypothetical protein	20692	6.2	15	5	8	1	ND	0	0	ND	RB
B0B8V6	CTL0099	*ribD*	Riboflavin biosynthesis protein (diaminohydroxyphosphoribosylaminopyrimidine deaminase)	41104	7.5	69	6	132	1	ND	7	1	ND	0.1
B0B8V8	CTL0101	*ribH*	Riboflavin synthase beta chain (6_7-dimethyl-8-ribityllumazine synthase)	16412	6.3	48	5	80	4	14	76	1	ND	1.0

B0B8X1	CTL0114	*hemG*	protoporphyrinogen oxidase	47237	9.7	55	5	33	1	ND	27	1	ND	0.8
B0B8X2	CTL0115	*hemN*	coproporphyrinogen oxidase (NAD)	52989	7.0	37	12	50	1	ND	0	0	ND	RB
B0B8X3	CTL0116	*hemE*	uroporphyrinogen decarboxylase [	37703	5.6	36	3	7	1	ND	0	0	ND	RB
B0B9W0	CTL0462	*hemL*	glutamate-1-semialdehyde 2_1-aminomutase	45940	6.0	126	16	44	3	16	24	3	6	0.6
B0B7X5	CTL0662	*ribC*	riboflavin synthase alpha chain	22123	5.9	24	2	9	1	ND	0	0	ND	RB
B0B8D2	CTL0821	*lipA*	lipoic acid synthetase	34695	7.2	126	13	20	3	14	19	3	3	1.0
B0B8I8	CTL0877	*folP*	2-amino-4-hydroxy-6-hydroxymethyldihydropteridi ne pyrophosphokinase	50263	5.4	183	21	61	4	21	19	3	4	0.3
B0B8I9	CTL0878	*folX*	dihydroneopterin aldolase	13941	5.2	7	4	0	0	ND	6	1	ND	EB
B0B8K3	CTL0892	*ispA*	dimethylallyltransferase	32524	4.9	77	16	37	2	15	20	3	6	0.5
Base & Nucleotide Metabolism														
B0B8Q4	CTL0047	*pyrH*	uridylate kinase	26163	5.3	145	9	95	4	33	63	4	15	0.7
B0B8X7	CTL0120	*amn*	AMP nucleosidase	32034	6.6	56	12	20	2	7	20	2	11	1.0
B0B956	CTL0199	*nrdA*	ribonucleoside-diphosphate reductase alpha chain	119414	5.7	590	65	117	4	78	70	3	32	0.6
B0B957	CTL0200	*nrdB*	ribonucleoside-diphosphate reductase beta chain	40504	5.3	104	14	28	2	6	23	2	2	0.8
B0B973	CTL0216		Cytosine deaminase	18476	6.2	9	2	0	0	ND	3	1	ND	EB
B0B9D7	CTL0285	*gmk*	guanylate kinase	23109	5.6	68	9	26	2	1	25	1	ND	0.9
B0B9E6	CTL0294	*dcd*	deoxycytidine triphosphate deaminase	21385	4.8	84	8	34	4	13	21	4	4	0.6
B0B9K7	CTL0354	*trxB*	thioredoxin reductase	37868	7.5	128	15	65	3	28	44	3	22	0.7
B0B9N6	CTL0383	*adk*	adenylate kinase	27727	4.9	141	15	142	4	28	89	4	25	0.6
B0B9T3	CTL0435	*pyrG*	CTP synthase	60189	5.7	165	23	54	3	14	32	4	5	0.6
B0B9T8	CTL0440	*tdk*	thymidylate kinase	22457	6.5	64	7	31	3	11	19	3	4	0.6

B0B7K6	CTL0542		PTS-family membrane transport protein IIA component	25758	5.3	86	12	59	4	26	42	4	12	0.7
B0B7K8	CTL0544	*dut*	deoxyuridine 5'-triphosphate nucleotidohydrolase	15337	5.2	49	6	61	4	22	74	4	2	1.2
B0B7U8	CTL0634		putative nucleotide-binding protein	50819	7.1	136	19	119	4	102	70	4	27	0.6
B0B825	CTL0712	*cmk*	cytidylate kinase	24044	5.1	141	17	62	4	13	55	4	12	0.9
B0B874	CTL0762	*ndk*	nucleoside diphosphate kinase	15265	5.4	104	10	59	4	8	28	4	10	0.5
B0B8B3	CTL0801	*trxA*	thioredoxin	11197	5.0	5	2	0	0	ND	3	1	ND	EB
B0B8H7	CTL0866	*ahpC*	thioredoxin peroxidase	21717	4.8	143	16	549	4	61	445	4	169	0.8
Cell Envelope														
B0B9A6	CTL0254	*pmpI*	polymorphic outer membrane protein	95518	6.1	442	37	241	4	125	82	4	18	0.3
B0B9A4	CTL0251	*pmpH*	polymorphic outer membrane protein	107331	6.6	507	41	341	4	191	94	4	18	0.3
B0B9A3	CTL0250	*pmpG*	polymorphic outer membrane protein	107244	5.7	537	38	637	4	334	154	4	26	0.2
B0B9A2	CTL0249	*pmpF*	polymorphic outer membrane protein	112539	9.1	529	36	123	4	63	30	3	4	0.2
B0B9A1	CTL0248	*pmpE*	polymorphic outer membrane protein	104711	6.7	440	37	292	4	164	62	4	20	0.2
B0B940	CTL0183	*pmpD*	polymorphic outer membrane protein	160521	4.8	831	77	685	4	291	361	4	80	0.5
B0B7Y4	CTL0671	*pmpC*	polymorphic outer membrane protein	187044	4.6	649	66	265	4	167	76	4	4	0.3
B0B7Y3	CTL0670	*pmpB*	polymorphic outer membrane protein	183115	8.2	959	77	376	4	48	237	4	14	0.6
B0B7Y2	CTL0669	*pmpA*	polymorphic outer membrane protein	105625	8.6	457	40	112	4	123	33	2	1	0.3
B0B7I6	CTL0522	*pbp*	penicillin-binding protein	73428	9.3	62	6		0	ND	19	1	ND	EB

B0B8H4	CTL0863	*pal*	peptidoglycan-associated lipoprotein	21532	7.9	112	10	281	4	163	100	4	21	0.4
B0B7G7	CTL0503	*oxaA*	inner membrane protein	88024	8.4	417	36	237	4	108	66	4	4	0.3
B0B7F8	CTL0494	*ompH*	outer membrane protein	19445	4.7	153	9	475	4	123	328	4	96	0.7
B0B8T9	CTL0082	*ompB*	outer membrane protein B	37406	5.2	174	9	459	4	194	215	4	50	0.5
B0B8Q7	CTL0050	*ompA*	major outer membrane protein	42550	5.1	414	21	2041	4	1117	2728	4	559	1.3
B0B815	CTL0702	*omcB*	60kD cysteine-rich outer membrane protein	59452	8.1	384	34	1704	4	361	518	4	91	0.3
B0B8Y5	CTL0128	*nlpD*	muramidase	27234	9.1	94	10	44	4	10	63	2	2	1.4
B0B8Y4	CTL0127	*murD*	UDP-N-acetylmuramoylalanine--D-glutamate ligase	46183	5.5	109	12	19	3	10	0	0	ND	RB
B0B8Y8	CTL0131	*murC*	UDP-N-acetylmuramate--alanine ligase	89204	5.2	189	24	42	4	28	23	1	ND	0.5
B0B7I8	CTL0524	*mraW*	S-adenosyl-L-methionine-dependent methyltransferase	34017	8.9	104	16	55	3	38	27	3	11	0.5
B0B9M6	CTL0373	*incG*	inclusion membrane protein G	17540	8.3	63	6	161	4	111	19	4	4	0.1
B0B9M4	CTL0371	*incE*	inclusion membrane protein E	13594	8.1	59	3	188	4	87	50	4	12	0.3
B0B7E9	CTL0485	*incC*	inclusion membrane protein C	18512	8.8	6	2	7	1	ND	0	0	ND	RB
B0B9M7	CTL0374	*incA*	inclusion membrane protein A	27489	6.2	160	27	84	4	27	20	2	1	0.2
B0B945	CTL0188	*glmS*	glucosamine--fructose-6-phosphate aminotransferase	67471	5.4	250	32	121	4	52	62	4	42	0.5
B0B814	CTL0701	*crpA*	cysteine-rich membrane protein	15832	6.2	88	4	174	4	97	33	4	16	0.2
B0B7I4	CTL0520	*amiA*	N-acetylmuramoyl-L-alanine amidase	28690	9.8	78	6	24	2	16	12	2	6	0.5
B0B8L7	CTL0010		putative membrane protein	32122	9.2	17	7	59	1	ND	0	0	ND	RB
B0B941	CTL0184		candidate inclusion membrane protein	29429	5.4	31	5	94	1	ND	0	0	ND	RB
B0B942	CTL0185		putative membrane protein	11421	10.7	54	4	242	4	112	32	4	23	0.1
B0B980	CTL0223		putative integral membrane protein	45891	5.2	92	7	44	2	2	0	0	ND	RB
B0B999	CTL0246		putative membrane protein	38368	4.6	14	4	17	1	ND	0	0	ND	RB
B0B9A0	CTL0247		putative membrane protein	44978	8.2	222	17	112	4	75	29	4	9	0.3

B0B9B2	CTL0260		putative membrane protein	39562	9.2	17	4	30	1	ND	0	0	ND	RB
B0B9B9	CTL0267		putative integral membrane protein	29973	9.5	56	7	34	4	20	13	2	3	0.4
B0B9C5	CTL0273		putative membrane protein	17817	9.2	71	9	25	4	9	7	3	3	0.3
B0B9L0	CTL0357		putative membrane protein	17289	9.6	21	5	44	1	ND	0	0	ND	RB
B0B9Q2	CTL0399		putative membrane protein	31355	8.9	160	15	96	4	30	31	2	7	0.3
B0B9Q5	CTL0402		putative integral membrane protein	162274	8.8	813	80	132	4	61	42	4	9	0.3
B0B9Q9	CTL0406		lipoprotein releasing system_ inner membrane component	56294	5.2	52	3	30	1	ND	0	0	ND	RB
B0B9R1	CTL0408		MAC/perforin family protein	90871	6.2	300	44	96	4	46	85	4	43	0.9
B0B9W4	CTL0466		candidate inclusion membrane protein	59776	9.2	72	12	41	2	21	0	0	ND	RB
B0B9X4	CTL0476		candidate inclusion membrane protein	29592	6.8	212	17	968	4	432	186	4	68	0.2
B0B9X7	CTL0478		candidate inclusion membrane protein	18263	5.5	9	4	3	1	ND	0	0	ND	RB
B0B9X9	CTL0480		candidate inclusion membrane protein	20777	5.1	104	8	46	4	34	13	3	7	0.3
B0B9Y0	CTL0481		candidate inclusion membrane protein	23534	9.2	88	6	75	4	52	18	4	7	0.2
B0B7F7	CTL0493		outer membrane protein (variable surface antigen)	88771	9.0	389	38	309	4	161	91	4	11	0.3
B0B7H0	CTL0506		inner membrane protein	28293	9.2	21	1	13	1	ND	0	0	ND	RB
B0B7K4	CTL0540	*omp85*	candidate inclusion membrane protein	63512	8.4	214	26	85	4	11	23	2	2	0.3
B0B7K5	CTL0541		putative membrane protein [Chlamydia trachomatis 434/Bu]	41811	9.8	122	20	115	3	26	109	4	29	0.9
B0B7T4	CTL0619		putative integral membrane protein	61073	9.4	90	14	49	3	11	0	0	ND	RB
B0B7Z1	CTL0678		putative inner membrane protein	25762	9.1	127	5	89	4	43	47	4	8	0.5
B0B812	CTL0699		putative membrane protein	12667	9.6	5	1	26	1	ND	0	0	ND	RB
B0B8A3	CTL0791		putative membrane protein	31250	9.4	133	11	329	4	166	45	2	16	0.1

B0B8E2	CTL0831		putative membrane protein	17110	7.5	20	3	14	1	ND	0	0	ND	RB
B0B8H5	CTL0864		putative soluble transglycosylase	22467	8.8	104	7	23	3	2	0	0	ND	RB
B0B8J1	CTL0880		putative integral membrane protein	49923	8.6	157	17	11	2	18	22	1	ND	0.5
B0B8J3	CTL0882		putative membrane protein	27913	9.6	109	10	340	4	172	48	4	36	0.1
B0B9U5	CTL0447		putative integral membrane protein	40800	9.5	26	3	0	0	ND	22	1	ND	EB
B0B983	CTL0226		putative integral membrane protein	21180	9.0	15	1	0	0	ND	10	1	ND	EB
Cellular Processes														
B0B8T3	CTL0076	*tig*	trigger factor (chaperone)	50098	5.0	304	39	279	4	101	283	4	80	1.0
B0B7L0	CTL0546	*sodM*	superoxide dismutase	23396	6.0	129	8	85	4	10	58	4	33	0.7
B0B884	CTL0772	*secY*	protein translocase subunit	50227	10.1	32	3	31	2	16	0	0	ND	RB
B0B821	CTL0708	*secF*	protein translocase (secFG fusion protein)	156503	7.4	#N/A	64	47	4	19	49	2	24	1.0
B0B7N7	CTL0573	*secE*	protein translocase subunit	9542	10.1	25	3	56	3	18	47	4	28	0.8
B0B8S7	CTL0070	*secA*	preprotein translocase subunit	110403	5.7	491	68	136	4	49	51	4	4	0.4
B0B9J9	CTL0346	*sctU*	type III secretion system inner membrane protein	40453	9.3	91	8	33	3	10	0	0	ND	RB
B0B8D6	CTL0825	*sctR*	type III secretion system_ membrane protein	33799	7.0	77	4	32	3	16	0	0	ND	RB
B0B8P8	CTL0041	*sctQ*	Type III secretion component_ basal body	41198	4.5	157	16	162	4	68	54	4	10	0.3
B0B8P5	CTL0038	*sctN*	ATP synthase (predicted TTSS protein)	48217	5.6	245	28	91	4	11	45	3	18	0.5
B0B8D5	CTL0824	*sctL*	type III secretion system protein	24767	5.7	139	14	159	4	35	115	4	27	0.7
B0B8D3	CTL0822	*sctJ*	type III secretion system protein_ membrane component	35512	5.2	242	15	566	4	249	228	4	24	0.4
B0B8F0	CTL0839	*scc2*	type III secretion chaperone (low calcium response protein H)	26053	9.1	89	10	60	4	25	27	2	8	0.5
B0B8F3	CTL0842	*copD*	putative type III secretion system protein	44066	9.5	305	14	134	4	21	76	4	12	0.6

B0B8F2	CTL0841	*copB*	putative type III secretion system membrane protein]	50321	9.2	300	23	73	4	8	31	4	5	0.4
B0B9J7	CTL0344	*copN*	low calcium response protein E (TTSS effector protein)	45195	4.9	250	24	54	4	11	42	4	18	0.8
B0B8Q0	CTL0043	*SctC*	Type III secretion structural protein (outer membrane ring)	100237	5.4	377	35	260	4	125	75	4	12	0.3
B0B9F1	CTL0299		putative type III secretion system chaperone	18356	5.0	149	11	192	4	49	98	4	47	0.5
B0B8P9	CTL0042	*pkn5*	putative serine/threonine-protein kinase (TTSS effector protein)	55968	6.3	30	7	8	1	ND	0	0	ND	RB
B0B8R4	CTL0057	*parB*	putative chromosome partitioning protein	31524	8.7	160	19	27	4	7	27	3	9	1.0
B0B8T5	CTL0078	*mreB*	Cell shape determining protein	39510	8.7	119	13	46	4	33	27	4	6	0.6
B0B8F6	CTL0845	*minD*	Chromosome partitioning ATPase (ParA family)	28208	6.7	67	8	44	4	28	16	3	2	0.4
B0B9C7	CTL0275	*lepB*	signal peptidase I	71535	8.4	281	27	137	4	95	47	4	28	0.3
B0B9H2	CTL0320	*lepA*	GTP-binding protein	67408	6.0	218	25	54	4	32	32	4	11	0.6
B0B9J8	CTL0345	*SctV*	low calcium response protein D (predicted to be part of the TTSS apparatus) [	77974	7.7	407	33	252	4	132	65	4	20	0.3
B0B8E3	CTL0832	*gspG*	general secretion pathway protein G	12169	7.7	83	8	69	4	35	35	4	8	0.5
B0B8E5	CTL0834	*gspE*	general secretion pathway protein E	55754	6.6	93	10	34	2	5	0	0	ND	RB
B0B8E6	CTL0835	*gspD*	general secretion pathway protein D	83768	5.2	334	45	46	4	21	17	3	2	0.4
B0B949	CTL0192	*ftsY*	Cell Division Protein	31455	7.7	42	5	22	3	10	15	1	ND	0.7
B0B8Y6	CTL0129	*ftsW*	Cell division protein	41994	9.5	15	3	64	1	ND	0	0	ND	RB
B0B9D2	CTL0280	*ffh*	signal recognition particle_ subunit FFH/SRP54	49705	8.6	268	28	41	3	14	39	4	12	1.0
B0B8S9	CTL0072	*engA*	GTP-binding protein	55606	8.6	96	14	28	2	8	10	1		0.3
B0B935	CTL0177	*cafE*	ribonuclease E	59365	6.5	279	34	112	4	43	65	4	26	0.6
B0B8P0	CTL0033		phosphopeptide binding protein	89739	4.5	496	44	924	4	441	341	4	7	0.4

B0B8Q9	CTL0052		tetratricopeptide repeat protein	38333	4.9	93	13	14	4	4	10	1	ND	0.7
B0B9K0	CTL0347		GTP-binding protein	39990	5.2	189	18	88	4	54	67	4	33	0.8
B0B9P9	CTL0396		protein translocase	16791	8.6	38	5	18	2	1	17	3	10	1.0
B0B7H5	CTL0511		protein phosphatase 2C	28051	5.9	20	2	0	0	ND	0	1	ND	EB
Central Intermediary Metabolism														
B0B8T6	CTL0079	*pckA*	phosphoenolpyruvate carboxykinase (GTP)	66117	5.6	212	24	70	4	16	54	2	0	0.8
B0B8Z8	CTL0141	*ppa*	inorganic pyrophosphatase	23391	4.8	64	11	108	3	22	81	3	36	0.8
B0B925	CTL0167	*glgA*	glycogen synthase	53394	5.6	182	20	100	4	61	50	4	14	0.5
B0B7G4	CTL0500	*glgP*	glycogen phosphorylase	92767	5.7	375	47	39	4	2	41	2	10	1.1
B0B7R5	CTL0600		Ribonuclease Z	34662	6.6	94	12	48	1	ND	21	2	13	0.4
B0B807	CTL0694		Putative oxidoreductase	39884	7.6	38	11	28	1	ND	0	0	ND	RB
B0B862	CTL0750	*glgC*	glucose-1-phosphate adenylyltransferase	49935	6.2	61	12	17	1	ND	0	0	ND	RB
DNA Replication														
B0B8L8	CTL0011	*topA*	DNA topoisomerase I	96705	8.8	173	24	32	1	ND	18	1	ND	0.6
B0B8M5	CTL0018	*recA*	recombinase A	37818	7.0	213	22	311	4	159	255	4	34	0.8
B0B8N6	CTL0029	*gyrA2*	DNA gyrase subunit A	55250	6.8	66	16	26	2	6	18	2	6	0.7
B0B8T4	CTL0077		putative helicase	133190	5.4	513	66	113	4	71	50	4	14	0.4
B0B8W9	CTL0112	*hctA*	histone H1--like developmental protein	13699	10.7	28	1	162	2	49	61	1	ND	0.4
B0B8X4	CTL0117	*mfd*	transcription-repair coupling factor	121163	6.1	412	56	45	4	5	76	2	74	1.7
B0B918	CTL0160	*mutS*	DNA mismatch repair protein	92131	6.7	429	41	67	4	21	40	4	17	0.6
B0B954	CTL0197	*rmuC*	DNA recombination protein	48522	6.6	189	20	48	4	17	34	3	10	0.7
B0B9D1	CTL0279		peptide release factor-glutamine N5-methyltransferase	32772	6.3	23	4	12	1	ND	0	0	ND	RB
B0B9E8	CTL0296	*ruvB*	Holliday junction ATP-dependent DNA helicase	37289	7.0	63	10	22	2	11	16	1	ND	0.7

B0B9F2	CTL0300	*ssb*	single-strand DNA binding protein	17147	4.8	44	9	36	4	19	26	2	17	0.7
B0B9I3	CTL0331	*dnaN*	DNA polymerase III_ beta chain	46530	5.9	195	19	85	4	53	49	4	4	0.6
B0B9Q4	CTL0401	*dnlJ*	NAD-dependent DNA ligase	73498	6.6	59	8	34	1	ND	0	0	ND	RB
B0B9T9	CTL0441	*gyrA*	DNA gyrase subunit A	94235	6.3	478	52	176	4	66	82	4	7	0.5
B0B9U0	CTL0442	*gyrB2*	DNA gyrase subunit B	89761	5.5	458	53	143	4	53	103	4	2	0.7
B0B7G6	CTL0501	*dnaA*	Chromosomal replication initiation protein	51929	6.5	179	17	30	3	13	28	2	11	0.9
B0B7H7	CTL0513	*dnaQ*	DNA polymerase III_ epsilon chain	26559	5.6	126	13	37	4	16	55	4	12	1.5
B0B7I3	CTL0519	*ihfA*	integration host factor alpha-subunit	11411	11.1	101	7	204	4	62	472	4	86	2.3
B0B7J1	CTL0527	*dnaA2*	Chromosomal replication initiation protein	51348	8.7	205	22	46	4	20	31	4	9	0.7
B0B7L4	CTL0550	*radA*	DNA repair protein	49811	7.2	27	3	47	1	ND	0	0	ND	RB
B0B7P6	CTL0583	*xseA*	exodeoxyribonuclease VII large subunit	58704	9.6	127	16	29	3	12	0	0	ND	RB
B0B7P7	CTL0583A	*xseB*	exodeoxyribonuclease VII small subunit	8358	4.8	12	3	49	1	ND	61	1	ND	1.2
B0B7Q2	CTL0588	*dnaX*	DNA polymerase III subunit gamma/tau	51622	6.4	84	11	72	2	44	11	1	ND	0.1
B0B820	CTL0707	*recJ*	single-stranded-DNA-specific exonuclease	64550	9.4	173	22	35	3	23	35	1	ND	1.0
B0B850	CTL0738		putative DNA methyltransferase	19174	7.6	29	6	8	1	ND	16	2	11	2.1
B0B866	CTL0754	*polA*	DNA polymerase I	96663	5.3	312	35	45	3	29	27	2	5	0.6
B0B871	CTL0759	*dnaB*	replicative DNA helicase	53366	6.0	193	24	36	4	20	24	2	5	0.7
B0B875	CTL0763	*ruvA*	holliday junction DNA helicase [	22222	6.1	84	8	13	3	4	9	4	2	0.7
B0B8B0	CTL0798	*dnaQ2*	DNA polymerase III_ epsilon chain	28960	8.9	85	10	50	2	55	10	2	2	0.2
B0B8B9	CTL0807	*dnaE*	DNA polymerase III alpha subunit	139474	5.7	243	31	84	1	ND	47	1	ND	0.6
B0B8C9	CTL0818		putative helicase (SWF/SNF family)	136257	5.7	76	13	25	1	ND	0	0	ND	RB
B0B8G0	CTL0849	*uvrB*	excinuclease ABC subunit B	75854	5.4	195	23	60	3	46	17	1	ND	0.3
B0B8I3	CTL0872	*uvrD*	DNA helicase	72751	6.1	318	35	135	4	88	55	4	10	0.4
B0B8K0	CTL0889	*nfo*	endonuclease IV	31648	5.7	49	4	18	1	ND	18	1	ND	1.0

B0B876	CTL0764	*ruvC*	Holliday junction resolvase	18715	9.2	6	3	0	0	ND	3	1	ND	EB
B0B7Q1	CTL0587	*urvA*	excinuclease ABC subunit A	196870	6.3	81	14	0	0	ND	11	1	ND	EB
Energy Metabolism														
B0B8K9	CTL0002	*nqrA*	Na(+)-translocating NADH-quinone reductase subunit A	51757	8.9	127	26	108	4	56	29	3	3	0.3
B0B8R9	CTL0062	*pgk*	phosphoglycerate kinase	43071	5.7	111	18	71	4	30	49	3	1	0.7
B0B8U0	CTL0083	*gpdA*	glycerol-3-phosphate dehydrogenase	36229	8.2	136	11	38	4	10	26	4	7	0.7
B0B8U8	CTL0091	*pgmA*	phosphoglycerate mutase	25832	6.7	269	14	64	4	12	43	4	11	0.7
B0B8W6	CTL0109	*dmpP*	Na(+)-translocating NADH-quinone reductase subunit F	48518	5.2	147	13	75	4	44	40	3	20	0.5
B0B8X6	CTL0119	*tktB*	transketolase	73164	5.3	142	21	72	3	35	33	2	5	0.5
B0B944	CTL0187	*glmM*	phosphoglucosamine mutase	49288	6.0	193	24	50	4	20	51	4	9	1.0
B0B950	CTL0193	*sucC*	succinyl-CoA synthetase beta chain	41762	5.4	196	15	59	4	10	46	2	10	0.8
B0B951	CTL0194	*sucD*	succinyl-CoA synthetase alpha chain	30239	5.3	95	12	53	4	12	36	3	12	0.7
B0B9C0	CTL0268	*cydA*	cytochrome d ubiquinol oxidase subunit I	50223	9.3	54	10	117	3	41	0	0	ND	RB
B0B9C1	CTL0269	*cydB*	cytochrome d ubiquinol oxidase subunit II	39567	8.8	80	2	103	1	ND	10	1	ND	0.1
B0B9G2	CTL0310	*sucA*	2-oxoglutarate dehydrogenase E1 component	102586	5.3	199	29	67	3	37	20	1	ND	0.3
B0B9G3	CTL0311	*sucB*	dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex	40344	5.1	133	14	98	4	36	43	3	11	0.4
B0B9H1	CTL0319	*gnd*	6-phosphogluconate dehydrogenase	52619	5.4	79	14	38	1	ND	30	1	ND	0.8
B0B9H3	CTL0321	*npt1*	ADP_ATP carrier protein	58100	9.3	145	19	828	4	490	252	4	45	0.3
B0B9M9	CTL0376	*araD*	ribulose-phosphate 3-epimerase	24848	4.9	21	5	0	0	ND	14	1	ND	EB
B0B9T5	CTL0437	*zwf*	glucose-6-phosphate 1-dehydrogenase	58597	5.4	253	23	42	4	24	20	2	2	0.5

B0B9T6	CTL0438	*devB*	6-phosphogluconolactonase	29005	5.2	159	16	46	4	9	29	4	6	0.6
B0B9V5	CTL0457	*pfkA*	pyrophosphate--fructose 6-phosphate 1-phosphotransferase	61986	6.1	131	18	36	4	15	30	2	13	0.8
B0B9V7	CTL0459	*pfkA_2*	pyrophosphate--fructose 6-phosphate 1-phosphotransferase	61186	6.0	117	20	37	3	18	17	1	ND	0.5
B0B9W3	CTL0465	*rpiA*	ribose 5-phosphate isomerase	26673	5.4	45	10	15	2	7	8	1	ND	0.5
B0B9W5	CTL0467	*dhnA*	fructose-bisphosphate aldolase	38020	6.3	177	17	185	4	25	104	4	66	0.6
B0B7G1	CTL0497	*pdhA*	pyruvate dehydrogenase E1 component alpha subunit	37204	5.6	113	14	35	4	17	31	4	9	0.9
B0B7G2	CTL0498	*pdhB*	pyruvate dehydrogenase E1 component beta subunit	36168	5.6	117	14	53	4	24	23	3	3	0.4
B0B7G3	CTL0499	*pdhC*	dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex [	46400	5.7	173	18	59	4	18	64	4	5	1.1
B0B7J4	CTL0530	*nqrB*	Na(+)-translocating NADH-quinone reductase subunit B	54947	8.7	79	9	49	3	18	11	1	ND	0.2
B0B7J5	CTL0531	*nqrC*	Na(+)-translocating NADH-quinone reductase subunit C	34412	5.6	113	9	53	4	23	17	1	ND	0.3
B0B7L1	CTL0547	*mrsA*	phosphoglucomutase	67411	5.1	315	38	74	4	24	36	4	25	0.5
B0B7M1	CTL0557	*atpI*	V-type sodium ATP synthase subunit I	73285	6.5	88	13	62	4	41	31	1	ND	0.5
B0B7M2	CTL0558	*atpD*	V-type ATP synthase subunit D	23182	9.0	97	11	63	4	28	31	4	8	0.5
B0B7M3	CTL0559	*atpB*	V-type sodium ATP synthase subunit B	48645	5.5	196	17	84	4	21	34	3	5	0.4
B0B7M4	CTL0560	*atpA*	V-type ATP synthase alpha chain	65495	5.1	317	36	123	4	51	56	4	11	0.5
B0B7M6	CTL0562	*atpE*	V-type ATP synthase subunit E	22946	5.4	104	17	169	4	40	85	4	32	0.5
B0B7M9	CTL0565	*tal*	transaldolase	36161	4.9	212	21	369	4	163	242	4	45	0.7
B0B7P5	CTL0582	*tpiS*	triosephosphate isomerase	29776	5.4	61	9	36	2	9	54	1	ND	1.5
B0B7Q0	CTL0586	*pykF*	pyruvate kinase	53631	6.0	297	27	111	4	49	55	4	10	0.5
B0B7Q9	CTL0594		2-oxoisovalerate dehydrogenase	74457	5.5	111	17	26	3	12	0	0	ND	RB

B0B7U5	CTL0630	*mdhC*	malate dehydrogenase	35561	6.1	175	13	120	4	53	52	3	19	0.4
B0B7U7	CTL0633	*pgi*	glucose-6-phosphate isomerase	57700	5.5	245	24	65	4	32	36	4	12	0.6
B0B7X0	CTL0657	*sucB*	lipoamide acyltransferase component (E2) of branched-chain alpha-keto acid dehydrogenase complex	42535	5.6	93	15	43	2	8	13	1	ND	0.3
B0B868	CTL0756	*npt2*	putative nucleotide transport protein	59620	9.3	96	16	70	4	16	0	0	ND	RB
B0B879	CTL0767	*gapA*	glyceraldehyde 3-phosphate dehydrogenase	36294	5.6	123	17	115	4	9	73	4	37	0.6
B0B8D1	CTL0820	*lpdA*	dihydrolipoamide dehydrogenase	49477	6.3	261	21	88	4	20	84	4	13	1.0
B0B8G1	CTL0850	*eno*	enolase	45434	4.7	208	20	525	4	250	556	4	152	1.1
B0B873	CTL0761	*lplA*	lipoate-protein ligase A	26908	7.1	27	4	0	0	ND	7	1	ND	EB
Exported protein														
B0B8L3	CTL0006		exported protein	37184	5.0	36	3	25	1	ND	0	0	ND	RB
B0B8M2	CTL0015		putative exported protein	22071	8.6	81	7	38	3	25	6	1	ND	0.2
B0B8M6	CTL0019		putative exported lipoprotein	70454	5.9	50	10	36	1	ND	0	0	ND	RB
B0B8V9	CTL0102		putative exported protein	49899	8.9	82	12	33	2	14	0	0	ND	RB
B0B914	CTL0156		putative exported protein	19063	4.8	52	6	78	4	24	27	4	2	0.3
B0B9B8	CTL0266		putative exported protein	48112	6.4	282	22	155	4	90	46	4	5	0.3
B0B9E9	CTL0297		putative exported protein	29852	4.8	134	19	201	4	112	49	4	2	0.2
B0B9I1	CTL0329		exported protein	36478	9.3	92	9	14	2	9	0	0	ND	RB
B0B9T1	CTL0433		putative exported protein	26979	5.2	80	7	50	3	18	17	2	2	0.3
B0B9X1	CTL0473		putative exported protein	33080	6.1	101	11	187	4	86	62	4	5	0.3
B0B7M7	CTL0563		putative exported protein	26290	9.1	19	4	41	2	21	0	0	ND	RB
B0B7S4	CTL0609		putative exported protein	39968	5.4	167	13	22	4	7	0	0	ND	RB
B0B7V9	CTL0645		putative exported protein	47018	5.3	56	9	72	2	7	0	0	ND	RB
B0B849	CTL0737		putative exported protein	36313	8.3	139	17	60	4	36	10	1	ND	0.2
B0B855	CTL0743		putative exported protein	24846	8.0	57	10	17	2	1	9	2	4	0.5
B0B8C1	CTL0809		putative exported protein	36307	8.8	133	18	72	4	48	16	3	4	0.2

B0B8E7	CTL0836		putative exported protein	45849	5.5	64	12	38	2	2	0	0	ND	RB
B0B8J8	CTL0887		putative exported protein	48477	6.4	389	28	1464	4	577	840	4	261	0.6
B0B909	CTL0152		putative exported protein	39630	4.6	123	13	55	4	23	0	0	ND	RB
Fatty Acid & Phospholipid Metabolism														
B0B8N1	CTL0024	*kdsA*	2-dehydro-3-deoxyphosphooctonate aldolase	29645	5.7	158	17	51	4	21	47	4	9	0.9
B0B8W2	CTL0105		conserved hyporthetical protein	16508	4.6	54	3	33	4	7	37	2	15	1.1
B0B8Z6	CTL0139	*fabF*	3-oxoacyl-[acyl-carrier-protein] synthase	44874	5.5	156	17	284	4	109	185	4	23	0.7
B0B901	CTL0144		1-acyl-sn-glycerol-3-phosphate acyltransferase	29187	9.3	83	10	56	4	18	26	2	2	0.5
B0B902	CTL0145	*aas*	2-acylglycerophosphoethanolamine acyltransferase	59393	7.2	194	22	106	4	64	25	3	5	0.2
B0B924	CTL0166	*pgsA*	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	22753	8.8	76	3	71	4	64	26	3	15	0.4
B0B934	CTL0176	*plsB*	glycerol-3-phosphate acyltransferase	38125	5.7	134	17	50	4	24	40	3	11	0.8
B0B939	CTL0182	*plsX*	Fatty acid/phospholipid synthase	34439	6.2	66	5	29	1	ND	14	2	8	0.5
B0B9L2	CTL0359	*fabI*	enoyl-acyl-carrier protein reductase	32012	5.2	152	14	121	4	39	63	4	19	0.5
B0B9N1	CTL0378	*accB*	biotin carboxyl carrier protein of acetyl-CoA carboxylase	18198	5.1	98	7	52	4	47	33	3	19	0.6
B0B9N2	CTL0379	*accC*	biotin carboxylase	50106	6.4	222	23	112	4	65	48	4	3	0.4
B0B9P4	CTL0391		putative phospholipase-carboxylesterase family protein	26823	5.3	77	5	55	3	20	36	1	ND	0.6
B0B9V6	CTL0458		conserved hypothetical protein	31610	6.0	40	3	12	2	2	0	0	ND	RB
B0B7F2	CTL0488	*acpP*	acyl carrier protein	8702	3.9	59	5	502	4	120	662	4	202	1.3
B0B7F3	CTL0489	*fabG*	3-oxoacyl-[acyl-carrier protein] reductase	26036	7.7	138	12	115	4	25	66	4	30	0.6
B0B7F4	CTL0490	*fabD*	malonyl-CoA-[acyl-carrier-protein] transacylase	33489	4.9	87	9	27	4	9	18	2	3	0.7

B0B7F5	CTL0491	*fabH*	3-oxoacyl-[acyl-carrier-protein] synthase III	35327	7.0	122	15	53	3	19	42	4	10	0.8
B0B7F9	CTL0495	*lpxD*	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	38403	7.4	168	8	105	4	77	46	3	7	0.4
B0B7I1	CTL0517	*accA*	acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	36374	5.9	156	18	91	4	57	49	4	16	0.5
B0B7K9	CTL0545	*accD*	acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	33688	7.5	130	12	81	4	27	65	4	23	0.8
B0B826	CTL0713	*plsC*	1-acyl-sn-glycerol-3-phosphate acyltransferase	23825	10.0	86	13	41	4	20	38	1	ND	0.9
B0B8A5	CTL0793	*lpxA*	acyl-[acyl-carrier-protein]--UDP-N-acetylglucos amine O-acyltransferase	30724	6.1	120	14	61	4	31	35	4	8	0.6
B0B8A6	CTL0794	*fabZ*	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase	16624	9.2	20	3	27	1	ND	20	1	ND	0.7
B0B8A7	CTL0795	*lpxC*	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	31298	5.7	56	8	26	2	25	46	3	20	1.8
B0B8A9	CTL0797		acyl-CoA hydrolase	18526	8.6	36	5	20	2	17	9	1	ND	0.5
hypothetical proteins														
B0B8L0	CTL0003		conserved hypothetical protein	16754	6.2	122	9	109	4	35	64	4	47	0.6
B0B8M1	CTL0014		conserved hypothetical protein	51478	6.7	23	5	22	1	ND	0	0	ND	RB
B0B8N5	CTL0028		conserved hypothetical protein	8783	8.0	60	7	343	4	14	268	4	53	0.8
B0B8P1	CTL0034		conserved hypothetical protein	9326	9.1	85	7	113	4	3	144	4	32	1.3
B0B8P3	CTL0036		conserved hypothetical protein	16465	4.9	61	7	103	4	39	88	4	4	0.8
B0B8P4	CTL0037		conserved hypothetical protein	24444	4.6	169	12	190	4	69	162	4	18	0.9
B0B8P7	CTL0040		conserved hypothetical protein	31030	4.8	62	12	130	4	40	150	4	21	1.1
B0B8Q2	CTL0045		conserved hypothetical protein	19858	5.5	149	10	47	4	11	42	4	6	0.9
B0B8R7	CTL0060		conserved hypothetical protein	25176	5.0	110	11	78	4	6	69	3	30	0.9

B0B8S0	CTL0063	conserved hypothetical protein	34744	5.1	229	16	97	4	18	44	4	5	0.5
B0B8S1	CTL0064	conserved hypothetical protein	44310	5.2	199	20	42	4	15	25	4	10	0.6
B0B8S8	CTL0071	conserved hypothetical protein	20483	4.6	75	5	37	3	13	19	3	7	0.5
B0B8T7	CTL0080	conserved hypothetical protein	86301	4.9	107	11	57	1	ND	0	0	ND	RB
B0B8T8	CTL0081	conserved hypothetical protein	44048	5.3	72	11	9	1	ND	25	1	ND	2.6
B0B8U4	CTL0087	conserved hypothetical protein	19602	5.9	25	5	14	2	2	0	0	ND	RB
B0B8X9	CTL0122	hypothetical protein	8617	9.1	11	2	9	1	ND	0	0	ND	RB
B0B8Y9	CTL0132	conserved hypothetical protein	15082	4.8	80	5	178	4	96	289	4	188	1.6
B0B8Z4	CTL0137	conserved hypothetical protein	64095	5.4	430	37	222	4	85	158	4	52	0.7
B0B916	CTL0158	hypothetical protein	18416	4.6	48	9	391	4	240	416	4	78	1.1
B0B937	CTL0179	conserved hypothetical protein	12470	10.2	6	3	0	0	ND	4	1	ND	EB
B0B974	CTL0217	conserved hypothetical protein	7992	3.9	24	5	7	2	2	15	3	10	2.0
B0B978	CTL0221	conserved hypothetical protein	17539	5.0	66	7	13	4	7	16	1	ND	1.3
B0B979	CTL0222	conserved hypothetical protein	6843	4.0	39	3	143	4	78	206	4	78	1.4
B0B995	CTL0238	conserved hypothetical protein	53614	5.3	261	22	70	4	14	55	4	28	0.8
B0B997	CTL0244	conserved hypothetical protein	37483	9.2	100	12	45	4	28	18	1	ND	0.4
B0B9A7	CTL0255	conserved hypothetical protein	65820	5.4	510	40	574	4	148	163	4	4	0.3
B0B9B4	CTL0262	conserved hypothetical protein	35549	6.1	138	10	45	4	28	16	1	ND	0.4
B0B9C3	CTL0271	conserved hypothetical protein	26737	4.7	115	9	127	4	22	27	4	13	0.2
B0B9C4	CTL0272	conserved hypothetical protein	47726	6.6	259	30	129	4	58	46	4	13	0.4
B0B9D8	CTL0286	conserved hypothetical protein	11680	9.8	25	8	12	2	3	10	1	ND	0.9
B0B9E7	CTL0295	conserved hypothetical protein	5675	5.1	19	3	3	1	ND	2	1	ND	0.6
B0B9F5	CTL0303	conserved hypothetical protein	35794	6.3	36	3	23	1	ND	0	0	ND	RB
B0B9F7	CTL0305	conserved hypothetical protein	50786	5.6	130	12	94	3	83	21	2	5	0.2
B0B9F8	CTL0306	conserved hypothetical protein	58928	4.9	103	13	36	1	ND	31	2	2	0.9
B0B9F9	CTL0307	conserved hypothetical protein	60546	5.3	160	18	38	3	13	35	2	11	0.9
B0B9G1	CTL0309	conserved hypothetical protein	17204	4.4	36	6	35	4	17	35	4	8	1.0
B0B9H4	CTL0322	conserved hypothetical protein	17948	10.0	137	11	186	4	74	349	4	59	1.9
B0B9J0	CTL0338	conserved hypothetical protein	59568	4.8	259	18	68	4	6	33	4	3	0.5
B0B9J1	CTL0338A	conserved hypothetical protein	18320	9.9	28	5	19	1	ND	0	0	ND	RB

B0B9Q0	CTL0397		conserved hypothetical protein	31449	8.5	117	10	80	4	57	0	0	ND	RB
B0B9Q1	CTL0398		conserved hypothetical protein	31497	5.8	160	16	88	4	48	37	4	43	0.4
B0B9V3	CTL0455		conserved hypothetical protein	28263	4.5	37	6	19	4	3	21	2	4	1.1
B0B9W1	CTL0463		conserved hypothetical protein	20814	4.6	49	5	74	4	33	53	2	0	0.7
B0B9W2	CTL0464		conserved hypothetical protein	17082	5.0	55	10	27	4	11	8	2	3	0.3
B0B7H1	CTL0507		conserved hypothetical protein	14372	4.8	9	3	8	2	3	0	0	ND	RB
B0B7H6	CTL0512		conserved hypothetical protein	18876	4.8	132	11	421	4	96	450	4	71	1.1
B0B7H8	CTL0514		conserved hypothetical protein	28754	6.4	109	11	30	4	16	24	3	7	0.8
B0B7H9	CTL0515		conserved hypothetical protein	21942	5.6	19	4	21	2	1	0	0	ND	RB
B0B7I9	CTL0525		conserved hypothetical protein	21558	9.1	68	6	25	2	4	66	1	ND	2.7
B0B7J0	CTL0526		conserved hypothetical protein	15607	4.6	59	5	48	4	15	17	4	5	0.3
B0B7J2	CTL0528		conserved hypothetical protein	21294	4.5	72	8	52	4	34	24	2	9	0.5
B0B7J9	CTL0535		conserved hypothetical protein	79885	7.3	113	14	29	2	1	0	0	ND	RB
B0B7M5	CTL0561		conserved hypothetical protein	32129	5.4	184	17	121	4	88	35	4	11	0.3
B0B7P8	CTL0584		conserved hypothetical protein	9953	4.3	26	2	36	2	12	33	3	8	0.9
B0B7Q3	CTL0589		conserved hypothetical protein	10537	5.2	68	5	50	4	5	59	4	10	1.2
B0B7U1	CTL0626	*aaxA*	conserved hypothetical protein	49447	9.0	235	15	368	4	119	159	4	17	0.4
B0B7V7	CTL0643		conserved hypothetical protein	77101	5.8	182	21	79	4	19	56	3	28	0.7
B0B7V8	CTL0644		conserved hypothetical protein	12867	9.9	39	5	39	3	20	28	4	12	0.7
B0B7W8	CTL0655		conserved hypothetical protein	29552	7.1	229	19	154	4	51	227	4	88	1.5
B0B7Z7	CTL0684		conserved hypothetical protein	70066	4.9	196	28	69	4	25	35	3	6	0.5
B0B801	CTL0688		conserved hypothetical protein	39227	5.1	171	22	92	4	19	84	3	15	0.9
B0B830	CTL0717		conserved hypothetical protein	26532	5.5	105	8	57	4	11	100	4	49	1.7
B0B839	CTL0726		conserved hypothetical protein	12782	9.4	51	8	26	4	11	19	4	4	0.7
B0B845	CTL0733		conserved hypothetical protein	30147	5.3	58	7	59	4	49	24	1	ND	0.4
B0B854	CTL0742		conserved hypothetical protein	27889	8.5	22	5	33	1	ND	0	0	ND	RB
B0B877	CTL0765		conserved hypothetical protein	17709	5.0	164	11	29	4	26	21	3	12	0.7
B0B878	CTL0766		conserved hypothetical protein	32034	4.7	99	12	217	4	25	351	4	109	1.6
B0B8B2	CTL0800		conserved hypothetical protein	27481	5.3	177	19	239	4	44	253	4	25	1.1
B0B8C0	CTL0808		conserved hypothetical protein	32658	9.2	134	9	23	3	6	9	1	ND	0.4

B0B8C4	CTL0812		conserved hypothetical protein	15804	4.3	48	7	12	4	6	7	3	4	0.6
B0B8D4	CTL0823		conserved hypothetical protein	32051	6.6	78	7	44	3	7	21	1	ND	0.5
B0B8F1	CTL0840		conserved hypothetical protein	13291	6.5	38	4	117	2	88	10	1	ND	0.3
B0B8F8	CTL0847		conserved hypothetical protein	21142	5.6	268	11	914	4	115	1145	4	208	1.3
B0B8G4	CTL0853		conserved hypothetical protein	109092	5.5	175	29	31	2	9	22	1	ND	0.7
B0B8H6	CTL0865		conserved hypothetical protein	14407	4.6	7	2	3	1	ND	0	0	ND	RB
B0B8I0	CTL0869		conserved hypothetical protein	23278	7.0	84	11	19	3	7	66	3	5	3.5
B0B8I1	CTL0870		conserved hypothetical protein	8896	4.8	60	5	23	2	8	12	4	6	0.5
B0B8I5	CTL0874		conserved hypothetical protein	26833	4.9	168	14	461	4	161	807	4	267	1.8
B0B8I6	CTL0875		conserved hypothetical protein	27141	5.5	28	4	11	1	ND	0	0	ND	RB
B0B8J4	CTL0883		conserved hypothetical protein	96809	4.8	92	13	32	1	ND	0	0	ND	RB
B0B8J5	CTL0884		conserved hypothetical protein	93181	4.8	75	5	0	0	ND	25	2	8	EB
B0B8J6	CTL0885		conserved hypothetical protein	92625	4.9	354	28	55	4	20	44	3	13	0.8
B0B8K6	CTL0895		conserved hypothetical protein	9245	5.3	45	3	48	3	25	58	3	53	1.2
B0B8K7	CTL0897		conserved hypothetical protein	60901	5.8	280	36	113	4	32	41	3	10	0.4
B0B8Y0	CTL0123		conserved hypothetical protein	32717	7.3	65	9	30	4	10	17	2	9	0.6
B0B8Z7	CTL0140		conserved hypothetical protein	17374	5.0	87	12	106	4	27	62	4	25	0.6
B0B9L1	CTL0358		conserved hypothetical protein	34188	5.1	184	14	74	4	40	45	4	5	0.6
B0B9L7	CTL0364		conserved hypothetical protein	29671	8.9	41	7	30	2	12	0	0	ND	RB
B0B9F6	CTL0304		conserved hypothetical protein	26452	7.6	36	5	20	2	6	16	1	ND	0.8
B0B7V5	CTL0641		conserved hypothetical protein	12319	5.4	81	5	111	3	24	207	4	88	1.9
B0B7W2	CTL0648		conserved hypothetical protein	41295	5.3	119	14	19	4	4	19	3	18	1.0
B0B7Z4	CTL0681		conserved hypothetical protein	18456	4.4	21	5	0	0	ND	46	2	33	EB
B0B7Z8	CTL0685		conserved hypothetical protein	41372	7.2	93	7	27	3	11	20	3	6	0.7
B0B819	CTL0706	*euo*	conserved hypothetical protein	20930	7.8	76	8	10	4	2	14	2	4	1.3
B0B8J7	CTL0886		conserved hypothetical protein	68969	4.9	530	39	87	4	24	83	4	52	1.0
B0B833	CTL0720		conserved hypothetical protein	9736	9.4	117	7	82	4	52	136	4	102	1.7
B0B7S9	CTL0614		conserved hypothetical protein	21845	4.7	34	6	0	0	ND	4	1	ND	EB
B0B9L6	CTL0363		conserved hypothetical protein	27514	5.7	46	6	0	0	ND	7	1	ND	EB
B0B9C2	CTL0270		conserved hypothetical protein	48830	8.7	33	7	0	0	ND	4	1	ND	EB

B0B8P6	CTL0039		conserved hypothetical protein	20247	7.8	16	2	0	0	ND	5	1	ND	EB
B0B966	CTL0209		conserved hypothetical protein	76407	6.0	121	17	0	0	ND	30	1	ND	EB
B0B8M3	CTL0016		conserved hypothetical protein	47796	7.5	30	7	0	0	ND	4	1	ND	EB
B0B8R0	CTL0053		conserved hypothetical protein	53946	5.7	182	23	47	4	12	39	4	25	0.8
Other Categories														
B0B8L9	CTL0012		putative oxidoreductase	37055	5.4	148	16	29	4	20	16	2	8	0.6
B0B8W0	CTL0103		putative lipoprotein	23967	8.5	89	11	79	2	19	90	4	29	1.1
B0B8Z9	CTL0142	*ldh*	leucine dehydrogenase	37349	5.2	183	17	91	4	17	43	4	21	0.5
B0B9H9	CTL0327	*dxr*	1-deoxy-D-xylulose 5-phosphate reductoisomerase	41758	5.8	120	19	31	4	12	22	2	11	0.7
B0B9Q6	CTL0403	*mhpA*	FAD-dependent monooxygenase	57859	8.3	60	11	24	1	ND	15	1	ND	0.6
B0B9W8	CTL0470	*surE*	5'-nucleotidase	31516	4.8	140	16	101	4	55	42	4	3	0.4
B0B9X3	CTL0475		candidate inclusion membrane protein	13879	9.0	37	2	47	3	26	0	0	ND	RB
B0B7G9	CTL0505		putative lipoprotein	24041	8.7	167	7	382	4	290	104	4	25	0.3
B0B7H4	CTL0510		putative cysteine desulfurase	40245	6.2	66	7	17	3	10	0	0	ND	RB
B0B7K3	CTL0539	*trmU*	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	40174	6.6	81	7	18	1	ND	16	1	ND	0.9
B0B7Q4	CTL0590	*ptsI*	phosphoenolpyruvate-protein phosphotransferase	63718	5.8	35	3	79	1	ND	0	0	ND	RB
B0B7U4	CTL0629		putative oxidoreductase	38037	5.3	80	8	25	3	5	0	0	ND	RB
B0B7U9	CTL0635	*phnP*	metal-dependent hydrolase	30143	5.8	34	5	23	2	5	0	0	ND	RB
B0B7W9	CTL0656		carbohydrate isomerase	36020	5.5	96	9	17	3	6	14	2	3	0.8
B0B7X2	CTL0659		Tetraacyldisaccharide-1-P 4'-kinase	40882	9.3	35	9	29	1	ND	0	0	ND	RB
B0B831	CTL0718		ribosomal-protein-alanine acetyltransferase	19281	8.9	27	5	31	2	1	0	0	ND	RB
B0B834	CTL0721		putative metallo-phosphoesterase	37056	9.3	100	13	51	4	12	27	1	ND	0.5
B0B835	CTL0722	*ispD*	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	24228	5.4	49	7	25	4	10	0	0	ND	RB

B0B837	CTL0724		hydrolase_ haloacid dehalogenase-like family	26188	5.0	72	10	68	3	48	0	0	ND	RB
B0B860	CTL0748		methyltransferase	20887	8.6	30	2	19	3	6	0	0	ND	RB
B0B8B5	CTL0803	*mip*	peptidyl-prolyl cis-trans isomerase	26648	4.8	301	16	1956	4	853	1292	4	295	0.7
B0B8C7	CTL0815		putative methyltransferase	42100	9.5	29	9	46	1	ND	0	0	ND	RB
B0B8G8	CTL0858		putative hydrolase	29798	5.0	22	3		0	ND	3	1	ND	EB
B0B865	CTL0753	*yacE*	dephospho-CoA kinase	22982	5.5	15	4	0	0	ND	5	1	ND	EB
B0B8F7	CTL0846	*gp6D*	virulence plasmid protein pGP6-D-related protein	30733	9.4	110	11	21	1	ND	19	3	8	0.9
B0B8U7	CTL0090		cysteine desulfurase	41793	6.2	32	3	0	0	ND	3	1	ND	EB
Plasmid														
B0BCL9	pL2-07a		Putative uncharacterized protein	14281	6.3	95	7	45	4	33	18	4	6	0.4
B0BCM1	pL2-05		Putative uncharacterized protein	27923	4.8	172	14	320	4	48	153	4	80	0.5
B0BCM2	pL2-04		Putative uncharacterized protein	41425	9.5	54	11	39	2	5	0	0	ND	RB
B0BCM3	pL2-03		Virulence plasmid helicase	51457	8.6	127	20	52	3	6	12	1	ND	0.2
B0BCM6	pL2-08		Putative uncharacterized protein	28307	8.8	60	9	26	1	ND	25	2	12	1.0
B0BCM7	pL2-07		Putative uncharacterized protein	15217	5.6	30	7	26	2	2	7	1	ND	0.3
Transcription														
B0B8L1	CTL0004	*greA*	Transcription elongation factor	80997	5.3	465	59	139	4	70	105	4	17	0.8
B0B8T0	CTL0073	*pcnB*	polyA polymerase	46553	7.8	80	13	31	2	19	19	2	1	0.6
B0B8U9	CTL0092		ribosomal large subunit pseudouridine synthase B	26425	9.6	17	5	16	1	ND	0	0	ND	RB
B0B8Z1	CTL0134	*rbsV*	anti-sigma F factor antagonist	12412	7.7	54	7	23	3	6	29	2	18	1.2
B0B959	CTL0202		putative SAM-dependent methyltransferase	21559	5.7	24	1	0	0	ND	68	1	ND	EB
B0B971	CTL0214	*pnp*	polyribonucleotide nucleotidyltransferase	75403	5.7	408	42	117	4	48	62	4	17	0.5
B0B9B6	CTL0264		conserved hypothetical protein	16226	5.7	66	8	39	4	24	52	3	18	1.3

B0B9D4	CTL0282	*trmD*	tRNA (guanine-N(1)-)-methyltransferase	39780	5.6	36	10	0	0	ND	28	2	19	EB
B0B9D6	CTL0284	*rnhB*	ribonuclease HII	23975	6.1	87	15	19	3	3	17	4	4	0.9
B0B9K5	CTL0352	*nusA*	N utilization substance protein A	48885	5.2	284	26	147	4	44	163	4	47	1.1
B0B9S0	CTL0422	*trpR*	trp operon repressor	10875	7.9	30	5	8	1	ND	20	1	ND	2.7
B0B7L3	CTL0549	*rnc*	ribonuclease III	25595	6.1	14	2	33	1	ND	0	0	ND	RB
B0B7N0	CTL0566	*rpoC*	DNA-directed RNA polymerase beta-prime chain	154814	7.2	900	103	396	4	186	287	4	82	0.7
B0B7N1	CTL0567	*rpoB*	DNA-directed RNA polymerase beta chain	140058	5.6	802	88	406	4	176	339	4	146	0.8
B0B7N6	CTL0572	*nusG*	transcription antitermination protein	20760	5.3	188	14	155	4	88	93	4	22	0.6
B0B7S3	CTL0608	*ksgA*	dimethyladenosine transferase	31462	7.7	28	3	22	1	ND	0	0	ND	RB
B0B7W4	CTL0650	*hrcA*	Putative transcriptional regulatory protein	44979	5.9	128	19	63	4	37	0	0	ND	RB
B0B7W7	CTL0654	*vacB*	exoribonuclease II	77957	9.3	188	24	76	4	50	34	2	15	0.4
B0B7X3	CTL0660		23S rRNA methyltransferase	30136	5.8	35	4	8	2	2	0	0	ND	RB
B0B7X6	CTL0663	*nrdR*	transcriptional repressor	17660	9.1	110	6	43	4	28	37	4	10	0.9
B0B7Y0	CTL0667	*pcnB*	polyA polymerase	49446	9.1	163	21	49	4	14	20	4	2	0.4
B0B7Z6	CTL0683	*rsbV*	anti-sigma F factor antagonist	12530	5.2	117	9	59	4	24	71	4	41	1.2
B0B864	CTL0752	*rho*	transcription termination factor rho	51666	6.8	280	32	180	4	85	133	4	36	0.7
B0B881	CTL0769	*rpoA*	DNA-directed RNA polymerase alpha chain	41810	5.3	226	21	486	4	254	364	4	68	0.7
B0B8B4	CTL0802		putatve rRNA methylase (SpoU family)	16747	5.8	24	3	7	1	ND	4	1	ND	0.5
B0B8C3	CTL0811	*rsbW*	sigma regulatory factor-histidine kinase	16424	5.8	37	7	32	2	32	20	2	6	0.6
B0B8J0	CTL0879	*rpoD*	RNA polymerase sigma factor	66133	8.1	330	36	107	4	46	104	4	30	1.0
B0B8K5	CTL0894		transcriptional regulatory protein	25835	8.4	76	10	39	4	13	25	1	ND	0.7
B0B9L4	CTL0361		tRNA pseudouridine synthase A	31465	9.4	16	4	0	0	ND	4	1	ND	EB

Translation														
B0B8Q3	CTL0046	*rrf*	ribosome recycling factor	20054	8.6	90	15	62	3	13	83	4	15	1.3
B0B8Q5	CTL0048	*tsf*	translation elongation factor TS	30953	5.7	299	20	453	4	189	584	4	169	1.3
B0B8Q6	CTL0049	*rpsB*	SSU ribosomal protein S2	31220	6.6	203	22	395	4	165	317	4	135	0.8
B0B8T1	CTL0074	*clpX*	ATP-dependent Clp protease_ ATP-binding component	46183	5.6	159	23	113	4	45	86	4	45	0.8
B0B8T2	CTL0075	*clpP*	ATP-dependent Clp protease proteolytic component	22050	5.1	49	6	58	3	31	0	0	ND	RB
B0B8V5	CTL0098	*serS*	seryl-tRNA synthetase	48406	5.7	169	19	109	4	28	87	4	37	0.8
B0B8X5	CTL0118	*alaS*	alanyl-tRNA synthetase	97639	5.5	514	48	180	4	84	121	4	28	0.7
B0B8X8	CTL0121	*efp*	translation elongation factor P	20190	4.7	55	6	126	4	12	109	4	4	0.9
B0B906	CTL0149		protein disulfide isomerase	18586	7.9	141	11	102	4	32	84	4	5	0.8
B0B907	CTL0150	*lysS*	lysyl-tRNA synthetase	60076	5.3	231	24	72	4	67	40	4	19	0.6
B0B908	CTL0151	*cysS*	cysteinyl-tRNA synthetase	57210	6.0	54	12	26	1	ND	11	1	ND	0.4
B0B913	CTL0155	*rpsN*	SSU ribosomal protein S14P	11716	11.4	53	6	40	4	10	66	4	13	1.6
B0B923	CTL0165	*glyQ*	glycyl-tRNA synthetase	112503	5.7	398	44	60	4	23	37	4	4	0.6
B0B926	CTL0168	*rplY*	LSU ribosomal protein L25	20425	9.0	99	10	131	4	67	88	4	24	0.7
B0B927	CTL0169	*pth*	peptidyl-tRNA hydrolase	19963	7.8	28	6	7	2	3	16	1	ND	2.5
B0B928	CTL0170	*rpsF*	SSU ribosomal protein S6P	12907	8.8	67	7	194	4	94	163	4	54	0.8
B0B929	CTL0171	*rpsR*	SSU ribosomal protein S18	9402	11.4	44	5	65	4	2	91	4	33	1.4
B0B930	CTL0172	*rplI*	LSU ribosomal protein L9	18437	6.1	125	11	159	4	59	227	4	82	1.4
B0B933	CTL0175	*ptr*	exported insulinase/protease	108365	5.1	442	39	443	4	232	176	4	25	0.4
B0B952	CTL0195	*htrA*	serine protease	53270	5.9	460	26	774	4	293	399	4	91	0.5
B0B953	CTL0196		metalloprotease-insulinase	109245	5.3	359	35	87	4	35	30	2	10	0.3
B0B962	CTL0205	*infC*	bacterial protein translation initiation factor 3 (IF-3)	21067	9.5	71	9	33	3	17	41	4	5	1.3
B0B964	CTL0207	*rplT*	LSU ribosomal protein L20P	13936	11.9	82	7	77	4	42	65	4	35	0.9
B0B965	CTL0208	*pheS*	phenylalanyl-tRNA synthetase alpha chain	38677	5.8	153	22	30	4	6	19	4	4	0.6
B0B970	CTL0213	*ftsH*	Cell division protein	101816	5.8	543	55	257	4	91	119	4	36	0.5

B0B972	CTL0215	*rpsO*	SSU ribosomal protein S15P	10400	9.7	27	3	119	4	38	129	4	69	1.1
B0B981	CTL0224	*map*	methionine aminopeptidase	32641	6.1	105	9	47	4	6	45	4	17	1.0
B0B990	CTL0233	*cpa*	putative exported protease	67281	5.5	220	20	87	4	51	40	4	18	0.5
B0B991	CTL0234	*ispH*	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	34250	6.1	129	11	187	4	60	135	4	25	0.7
B0B9A9	CTL0257	*gatC*	glutamyl-tRNA(Gln) amidotransferase subunit C	11074	4.1	48	2	434	4	219	268	4	184	0.6
B0B9B0	CTL0258	*gatA*	aspartyl-glutamyl-tRNA(Asn-Gln) amidotransferase subunit A	53602	5.9	224	31	188	4	106	145	4	32	0.8
B0B9B1	CTL0259	*gatB*	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	54964	5.9	316	26	306	4	163	220	4	45	0.7
B0B9C6	CTL0274	*ileS*	isoleucyl-tRNA synthetase_ mupirocin resistant	118804	5.4	383	48	57	4	22	33	3	4	0.6
B0B9C9	CTL0277	*rpmE*	LSU ribosomal protein L31P	12188	9.6	13	3	0	0	ND	23	2	24	EB
B0B9D0	CTL0278	*prfA*	bacterial peptide chain release factor 1	40022	5.3	226	24	101	4	38	109	4	35	1.1
B0B9D3	CTL0281	*rpsP*	SSU ribosomal protein S16P	13410	10.4	67	6	171	4	109	198	4	102	1.2
B0B9D5	CTL0283	*rplS*	LSU ribosomal protein L19P	13142	10.0	107	12	129	4	62	130	4	72	1.0
B0B9D9	CTL0287	*metG*	methionyl-tRNA synthetase	62749	5.1	238	24	35	4	4	24	3	5	0.7
B0B9F3	CTL0301	*pepA*	putative aminopeptidase	54186	5.7	295	28	193	4	39	78	4	22	0.4
B0B9H0	CTL0318	*tyrS*	tyrosyl-tRNA synthetase	45435	6.6	208	24	118	4	59	82	4	14	0.7
B0B9I0	CTL0328		putative protease	69272	6.3	284	26	284	4	149	89	4	11	0.3
B0B9J4	CTL0341	*rpmB*	LSU ribosomal protein L28P	10145	11.7	29	4	47	3	24	43	3	11	0.9
B0B9K4	CTL0351	*infA*	bacterial protein translation initiation factor 2 (IF-2)	97378	8.2	431	46	118	4	8	165	4	4	1.4
B0B9K6	CTL0353	*rpsA*	SSU ribosomal protein S1P	63474	5.2	283	36	418	4	184	363	4	182	0.9
B0B9L8	CTL0365	*hsp60_1*	chaperonin GroEL	58091	5.3	549	35	1871	4	186	1300	4	194	0.7
B0B9L9	CTL0366	*groES*	10 kDa chaperonin GroES	11169	4.9	94	7	622	4	183	475	4	197	0.8
B0B9M0	CTL0367	*pepF*	oligoendopeptidase F	69030	5.6	207	29	130	4	18	59	3	19	0.5
B0B9M1	CTL0368	*clpB*	chaperone-protease ClpB	96631	5.4	481	58	270	4	98	150	4	22	0.6

B0B9N0	CTL0377	*efp*	translation elongation factor P	20529	5.0	75	9	53	2	35	26	2	6	0.5
B0B9N3	CTL0380	*rplM*	LSU ribosomal protein L13P	16850	10.2	78	6	118	4	60	146	4	60	1.2
B0B9N4	CTL0381	*rpsI*	SSU ribosomal protein S9P	14542	11.0	91	10	176	4	83	203	4	86	1.2
B0B9Q8	CTL0405	*rpmG*	LSU ribosomal protein L33P	6285	10.3	5	1	58	3	35	56	3	12	1.0
B0B9S7	CTL0429	*dsbG*	disulfide bond chaperone	26822	7.7	122	8	38	4	17	17	3	4	0.5
B0B9U7	CTL0449	*gcp*	O-sialoglycoprotein endopeptidase	35892	6.1	34	7	20	2	1	0	0	ND	RB
B0B9V9	CTL0461	*leuS*	leucyl-tRNA synthetase	92846	5.4	253	36	59	3	31	33	4	8	0.6
B0B7K2	CTL0538	*clpC*	ATP-dependent Clp protease	95239	6.0	476	50	307	4	139	206	4	29	0.7
B0B7L8	CTL0554	*valS*	valyl-tRNA synthetase	107103	5.6	126	18	21	1	ND	15	1	ND	0.7
B0B7N2	CTL0568	*rplL*	LSU ribosomal protein L12P (L7/L12)	13571	4.9	71	10	626	4	342	1006	4	284	1.6
B0B7N3	CTL0569	*rplJ*	LSU ribosomal protein L10P	18875	6.0	112	11	152	4	131	129	4	68	0.8
B0B7N4	CTL0570	*rplA*	LSU ribosomal protein L1P	24742	9.0	158	16	157	4	30	384	4	62	2.4
B0B7N5	CTL0571	*rplK*	LSU ribosomal protein L11P	15054	9.7	102	8	94	4	10	113	4	27	1.2
B0B7N8	CTL0574	*tufA*	translation elongation factor Tu	43309	5.4	404	32	2619	4	1474	2156	4	701	0.8
B0B7N9	CTL0575	*infA2*	bacterial protein translation initiation factor 1 (IF-1)	8413	9.4	51	5	29	4	15	17	2	14	0.6
B0B7R0	CTL0595	*dnaJ*	heat shock chaperone protein	41916	7.5	226	28	389	4	99	316	4	79	0.8
B0B7R3	CTL0598	*lon*	ATP-dependent protease La	91949	6.9	459	56	134	4	59	89	4	28	0.7
B0B7S2	CTL0607	*def*	peptide deformylase	20523	5.7	92	11	108	4	25	52	4	4	0.5
B0B7W3	CTL0649	*proS*	prolyl-tRNA synthetase	65676	5.6	118	17	39	4	15	18	1	ND	0.4
B0B7W5	CTL0651	*grpE*	HSP-70 Cofactor	21668	4.6	151	9	211	4	48	190	4	142	0.9
B0B7W6	CTL0652	*dnaK*	chaperone protein	70843	5.0	718	54	1424	4	274	1660	4	654	1.2
B0B7X7	CTL0664	*dksA*	dnaK suppressor protein	13938	5.1	91	7	140	4	14	184	4	22	1.3
B0B7Y9	CTL0676	*rpmA*	LSU ribosomal protein L27P	8916	11.0	37	4	34	4	28	32	3	22	0.9
B0B7Z0	CTL0677	*rplU*	LSU ribosomal protein L21P	12163	9.3	91	5	112	4	23	118	4	30	1.1
B0B803	CTL0690	*clpP*	ATP-dependent Clp protease proteolytic subunit	21073	5.2	94	6	284	4	119	245	4	70	0.9
B0B808	CTL0695	*rpsJ*	SSU ribosomal protein S10P	11869	10.5	59	7	50	4	22	75	4	29	1.5
B0B809	CTL0696	*fusA*	translation elongation factor G	76493	5.3	536	46	623	4	257	451	4	37	0.7

B0B810	CTL0697	*rpsG*	SSU ribosomal protein S7P	17799	9.9	100	5	239	4	71	252	4	92	1.1
B0B811	CTL0698	*rpsL*	SSU ribosomal protein S12P	14524	11.1	47	2	24	2	13	27	1	ND	1.2
B0B818	CTL0705	*gltX*	glutamyl-tRNA synthetase	58563	6.6	242	33	81	4	32	54	2	7	0.7
B0B827	CTL0714	*argS*	arginyl-tRNA synthetase	63029	6.1	256	25	81	4	28	42	4	12	0.5
B0B832	CTL0719	*prfB*	bacterial peptide chain release factor 2	42402	5.4	254	29	28	4	18	19	4	3	0.7
B0B848	CTL0736	*pheT*	phenylalanyl-tRNA synthetase beta chain	87115	6.1	70	20	46	2	5	0	0	ND	RB
B0B867	CTL0755	*sohB*	exported protease IV	35771	6.5	204	14	105	4	35	48	4	3	0.5
B0B880	CTL0768	*rplQ*	LSU ribosomal protein L17P	16152	11.3	49	5	0	0	ND	57	4	26	EB
B0B882	CTL0770	*rpsK*	SSU ribosomal protein S11P	13822	11.3	75	10	53	4	21	35	4	9	0.7
B0B883	CTL0771	*rpsM*	SSU ribosomal protein S13P	13895	11.0	66	8	142	4	42	174	4	86	1.2
B0B885	CTL0773	*rplO*	LSU ribosomal protein L15P	16114	10.2	60	8	100	4	47	97	4	34	1.0
B0B886	CTL0774	*rpsE*	SSU ribosomal protein S5P	17762	9.9	123	7	175	4	48	150	4	36	0.9
B0B887	CTL0775	*rplR*	LSU ribosomal protein L18P	13379	10.3	48	4	37	4	7	19	3	19	0.5
B0B888	CTL0776	*rplF*	LSU ribosomal protein L6P	19839	10.0	121	14	151	4	75	170	4	75	1.1
B0B889	CTL0777	*rpsH*	SSU ribosomal protein S8P	15070	10.3	42	4	35	3	34	7	2	2	0.2
B0B890	CTL0778	*rplE*	LSU ribosomal protein L5P	20489	9.4	103	12	152	4	69	118	4	49	0.8
B0B891	CTL0779	*rplX*	LSU ribosomal protein L24P	12608	10.4	67	6	90	4	25	83	4	19	0.9
B0B892	CTL0780	*rplN*	LSU ribosomal protein L14P	13443	9.8	108	11	97	4	10	85	4	56	0.9
B0B893	CTL0781	*rpsQ*	SSU ribosomal protein S17P	9645	10.5	12	4	9	1	ND	6	1	ND	0.7
B0B894	CTL0782	*rpmC*	LSU ribosomal protein L29P	8295	9.9	20	5	52	1	ND	16	3	5	0.3
B0B895	CTL0783	*rplP*	LSU ribosomal protein L16P	15775	11.3	62	6	152	4	27	111	4	14	0.7
B0B896	CTL0784	*rpsC*	SSU ribosomal protein S3P	24343	10.0	166	15	293	4	151	167	4	40	0.6
B0B897	CTL0785	*rplV*	LSU ribosomal protein L22P	12455	11.3	164	7	60	4	22	63	4	26	1.1
B0B898	CTL0786	*rpsS*	SSU ribosomal protein S19P	10233	10.8	40	5	41	1	ND	47	3	10	1.1
B0B899	CTL0787	*rplB*	LSU ribosomal protein L2P	31473	10.6	155	17	180	4	71	158	4	72	0.9
B0B8A0	CTL0788	*rplW*	LSU ribosomal protein L23P	12223	9.9	62	6	55	4	24	87	4	52	1.6
B0B8A1	CTL0789	*rplD*	LSU ribosomal protein L1E	24590	9.8	150	17	202	4	116	169	4	68	0.8
B0B8A2	CTL0790	*rplC*	LSU ribosomal protein L3P	23480	9.7	106	10	115	4	55	99	4	43	0.9

B0B8B6	CTL0804	*aspS*	aspartyl-tRNA synthetase	66272	5.3	326	34	117	4	40	93	4	20	0.8
B0B8B7	CTL0805	*hisS*	histidyl-tRNA synthetase	49093	6.7	215	25	93	4	82	43	3	11	0.5
B0B8E8	CTL0837	*pepP*	proline dipeptidase	39330	5.4	138	22	81	4	49	30	3	13	0.4
B0B8F5	CTL0844	*thrS*	threonyl-tRNA synthetase	72630	5.7	413	31	102	4	24	114	4	125	1.1
B0B8F9	CTL0848	*trpS*	tryptophanyl-tRNA synthetase	39380	6.5	181	24	197	4	59	247	4	76	1.3
B0B8G9	CTL0859	*dsdD*	thiol:disulfide interchange protein	76171	5.8	174	23	93	4	41	40	3	13	0.4
B0B8H8	CTL0867	*groEL2*	60 kDa chaperonin GroEL2	58869	5.1	50	4	59	1	ND	0	0	ND	RB
B0B8K1	CTL0890	*rpsD*	SSU ribosomal protein S4P	23660	10.0	203	15	102	4	54	97	4	32	0.9
B0B8J2	CTL0881	*rpsT*	SSU ribosomal protein S20P	10826	11.2	3	2	0	0	ND	ND	1	ND	EB
B0B7R1	CTL0596	*rpsU*	SSU ribosomal protein S21P	6666	10.9	5	2	0	0	0	16	1	ND	EB
Transport and binding Proteins														
B0B8R1	CTL0054		ABC transporter ATP-binding protein	28450	5.1	163	12	149	4	78	176	4	85	1.2
B0B8R2	CTL0055		conserved hypothetical protein	44583	6.0	151	18	45	4	28	22	2	0	0.5
B0B8R5	CTL0058	*dppF*	ABC transport protein_ ATPase component	31012	9.6	122	12	69	4	42	22	3	6	0.3
B0B8R6	CTL0059	*dppD*	ABC transport protein_ ATPase component	35998	6.4	62	6	29	2	13	0	0	ND	RB
B0B8V3	CTL0096		cation transporting ATPase	70483	6.8	229	20	269	4	162	98	4	4	0.4
B0B8W1	CTL0104	*dagA*	Na(+)-linked D-alanine glycine permease	48147	9.0	22	4	20	1	ND	0	0	ND	RB
B0B8W7	CTL0110		preprotein translocase	12814	9.5	56	5	74	4	9	39	3	18	0.5
B0B947	CTL0190		tyrosine-specific transport protein	43975	9.2	37	3	52	2	32	0	0	ND	RB
B0B968	CTL0211		putative membrane transport protein	40022	9.2	19	3	13	1	ND	0	0	ND	RB
B0B986	CTL0231		sulfate transporter	61653	8.5	45	5	47	2	34	0	0	ND	RB
B0B9E1	CTL0289		putative membrane transport/efflux protein	36145	7.0	22	3	15	3	13	0	0	ND	RB
B0B9H5	CTL0323		ABC transport protein	37063	5.0	304	19	950	4	209	379	4	102	0.4

B0B9H6	CTL0324		ABC transport protein_ ATP-binding component	28948	6.7	137	8	126	4	80	37	2	5	0.3
B0B9N7	CTL0384		putative ABC-membrane transport protein_ inner membrane component	23602	9.0	69	5	29	4	6	10	2	6	0.4
B0B9N8	CTL0385		ABC transporter_ ATP-binding component	25536	5.8	82	7	26	3	7	20	2	0	0.7
B0B9P7	CTL0394	*oppA*	oligopeptide transport system binding protein	48310	6.3	121	13	68	3	35	18	1	ND	0.3
B0B9R0	CTL0407		lipoprotein release ATP-binding component	25028	6.5	95	7	28	4	12	16	4	1	0.6
B0B9S5	CTL0427	*oppA2*	Oligopeptide transport system binding protein	60307	6.2	62	7	21	1	ND	0	0	ND	RB
B0B9U4	CTL0446	*mgtE*	magnesium transport protein	51411	4.9	142	12	69	4	54	15	2	0	0.2
B0B9Y1	CTL0482		putative sodium:dicarboxylate symport protein	45049	8.4	37	4	96	3	50	20	1	ND	0.2
B0B7R7	CTL0602		ABC transporter_ ATP-binding component	58876	5.3	239	24	89	4	46	46	4	3	0.5
B0B7R9	CTL0604		putative lipoprotein	63471	6.0	40	3	0	0	ND	33	1	ND	EB
B0B7S1	CTL0606	*secG*	preprotein translocase	11088	4.3	79	4	195	4	195	27	4	14	0.1
B0B7V0	CTL0636	*artJ*	arginine transport substrate-binding protein	28570	4.9	127	16	234	4	138	94	4	22	0.4
B0B7Y5	CTL0672		metal transporter_ metal-binding component	31462	6.9	49	4	15	2	15	0	0	ND	RB
Q6GX35	CTL0716	*tarp*	Translocated actin-recruiting phosphoprotein	103252	4.1	339	35	134	4	13	55	3	5	0.4
B0B852	CTL0740	*oppB2*	oligopeptide transport system membrane permease	53123	7.2	90	13	55	3	32	0	0	ND	RB
B0B853	CTL0741	*oppA4*	oligopeptide transport system_ binding protein	79929	5.0	270	37	97	4	59	22	1	ND	0.2
B0B859	CTL0747	*fliY*	glutamine-binding protein	28974	6.9	96	10	209	4	83	97	4	22	0.5

B0B8B8	CTL0806	*uhpC*	putative sugar phosphate permease	51668	8.6	107	16	143	4	98	29	1	ND	0.2
B0B8H3	CTL0862	*tolB*	outer membrane component of membrane transport system	47567	9.0	122	12	18	3	5	0	0	ND	RB

[a] = Molecular mass (kDa) were calculated using ProteinLynx Global Server Ver 2.3.
[b]= Isoelectric points were calculated using 'Compute p*I* ' (ExPASy bioinformatics resource portal)
[c]= Total number of peptides used to assign a protein.
[d]= Total number of unique peptides used to assign a protein.
[e]= Quantity of each assigned protein in EBs or RBs expressed as molecules per cell.
[f]= Calculated standard deviation of biological and technical replicates for each developmental form.