Expression and folding of recombinant chlamydial outer membrane proteins

Heather E Findlay

Thesis submitted for the degree of
Doctor of Philosophy
The University of Edinburgh

January 2004

Abstract

The *Chlamydiaceae* are obligate intracellular pathogens of eukaryotic cells. These few species are responsible for a wide variety of human and animal diseases, including trachoma, sexually transmitted genital infections, pneumonia and abortion. Female infertility secondary to chronic pelvic inflammation is a particularly important and relatively common complication of *C. trachomatis* infection. The cysteine-rich Major Outer Membrane Protein (MOMP) is the immunodominant chlamydial protein and a primary vaccine target. It is predicted to function as a general diffusion porin that may also, through intra- and/or intermolecular disulphide bond formation, play an important structural role in maintaining stability of the organism in the absence of a peptidoglycan layer. However, vaccine development and examining the specific biochemistry of MOMP *in situ* is problematic due to difficulties involved in culturing the organisms and the presence of other chlamydial membrane proteins.

A dual approach was taken to develop a recombinant system to express MOMP in *E. coli*. Full length MOMP expressed with the *E. coli ompT* signal sequence was shown to be successfully transported to the outer membrane of the cell. A proportion of the protein solubilised from the membrane eluted as trimers during size exclusion chromatography, and formed porin-like channels in planar lipid bilayers. Mature MOMP lacking a signal sequence accumulated in inclusion bodies when expressed in *E. coli*. These were denatured and refolded *in vitro* to produce higher order complexes of

MOMP composed of SDS resistant trimers. Cysteine residues were found to play a critical role in the stabilisation of both secondary and tertiary structures. The ability to express properly folded recombinant MOMPs, and other chlamydial outer membrane proteins, will allow more detailed analysis of the structural and functional roles of MOMPs, and contribute to the drive for vaccine development.

Declaration

I declare that this thesis has been completed by me and that, except where indicated to the contrary, the research documented is entirely my own.

Heather Findlay

Acknowledgements

I would like to thank all of my friends and family for their support over the last few years. Particular thanks are due to my supervisor Richard, Heather for all the help in the lab, and of course Andy for putting up with me.

Contents

List of Figures

List of Tables

Commonly used abbreviations

EB	elementary body
RB	reticulate body
COMC	chlamydial outer membrane complex
MOMP	major outer membrane protein
CRP	cysteine rich protein
PCR	polymerase chain reaction
SDS-PAGE	sodium dodecyl sulphate – polyacrylamide gel electrophoresis
CD	circular dichroism
DNA	deoxyribonucleic acid
EDTA	ethylenediamine tetraacetic acid
IPTG	isopropyl β-D-thiogalactopyranoside
PBS	phosphate buffered saline
GuHCl	guanidine hydrochloride
DTT	dithiothreitol
GSH	glutathionine – reduced
GSSG	glutathionine – oxidised
CMC	critical micelle concentration
DM	n-decyl β-D-maltopyranoside
DDM	n-dodecyl β-D-maltopyranoside
LDAO	lauryldimethylamine-N-oxide

Chapter 1: Introduction

1.1 Taxonomy and diseases of the Chlamydiales

The Chlamydiales are an order of bacteria, unique in that all known representatives are obligate intracellular pathogens of eukaryotic cells. Other well-known intracellular eubactria (eg. *Rickettsia*) are found in major bacterial groups where there are many extracellular members, whereas the Chlamydiales are phylogenetically isolated (figure 1.1A) (Preston et al.1998). For a long time, the order contained only one family and one genus (Chlamydiaceae and *Chlamydia*), but recently the availability of molecular techniques allowed a more thorough analysis that resulted in a reclassification of the members of the order (Everett et al.1999). Chlamydiales now contains four families; Chlamydiaceae, Parachlamydiaceae, Simkaniaceae and Waddliaceae (figure 1.1B) (Bush et al.2001). All four can be referred to as "chlamydiae." The family Chlamydiaceae is split into the genera *Chlamydia* and *Chlamydophila*, and the species within them are responsible for a wide variety of diseases in humans and animals (Everett.2002).

Chlamydia trachomatis is one of the most commonly known species of chlamydiae. The different serovars invade mainly cells of mucous membranes and lymphatic tissue. Trachoma, a disease that develops from infection of the eye, is a major cause of preventable blindness in developing countries with approximately 300 million cases

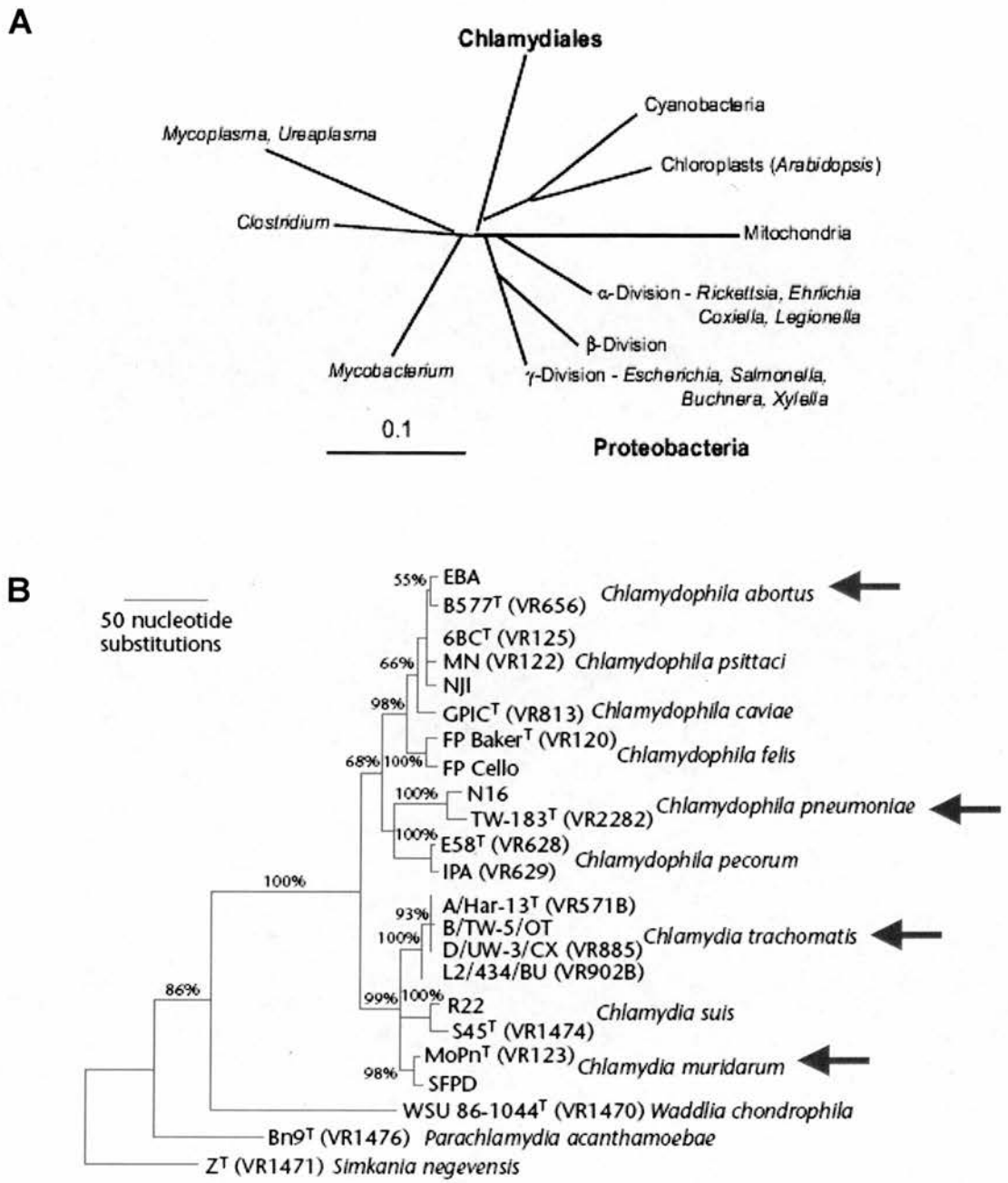


Figure 1.1: Phylogenetic relationships of the Chlamydiales. A. Consensus phylogenetic tree constructed from small ribosomal subunits. From Preston, Haubold and Rainey (1998). Bar indicates genetic distance, calculated by the Fitch-Margoliash algorithm. **B.** Evolutionary relationships among the Chlamydiales constructed from 16S-rRNA. From Bush and Everett (2001). Arrows indicate species of particular clinical or veterinary importance (see main text).

every year. *C. trachomatis* infection is the most common sexually transmitted disease (STD) worldwide, and can lead to serious complications such as ectopic pregnancy, pelvic inflammatory disease and female infertility. Infections are often asymptomatic, and so can go untreated for prolonged periods until the secondary disease is manifested. It can also cause neonatal inclusion conjunctivitis, pneumonia and some forms of arthritis. *Chlamydia muridarum* infection produces pneumonia in mice, and is important for its use as a model for the study of infection. An enteric isolate has been found in hamsters. *Chlamydia suis* infects swine, where it can cause conjunctivitis, enteritis and pneumonia.

Chlamydophila species also cause a wide range of disease. *Chlamydophila pneumoniae* has been shown to infect humans, koalas and horses. In humans, most infections produce acute or chronic bronchitis and pneumonia. It is widespread in the community with seroprevalence in >60% of adults. *Ch. pneumoniae* has also been associated with obstructive pulmonary disease, artherosclerosis (Ward.1995) and Alzheimers disease (Balin et al.1998). *Chlamydophila abortus* is endemic in ruminants, where the bacteria efficiently colonise the placenta leading to late-term abortion. This presents a major problem in the farming community. There are occasional incidences of zoonotic abortion in women working with infected animals. *Chlamydophila psittaci* infects birds where the disease is often systematic with most organs affected. The infection is passed in the eggs and is readily transmissible to humans. *Chlamydophila pecorum* has been isolated from ruminants, koalas and swine. It causes abortion, conjunctivitits, enteritis, pneumonia and polyarthritis. *Chlamydophila felis* is endemic in domesticated cats.

Infection causes conjunctivitis rhinitis and respiratory problems. *Chlamydophila caviae* causes conjunctivitis in guinea pigs.

Together the species of Chlamydiaceae are a major cause of disease worldwide. In humans, infections are often asymptomatic and persistent. Obtaining a better understanding of chlamydial biology is essential for the development of treatment and vaccines.

1.2 The developmental lifecycle

The Chlamydiales have a distinctive dimorphic lifecycle, typically of 2-3 days duration (figure 1.2) (Everett.2002). The infectious agent is the elementary body (EB). EBs are small, metabolically inactive, electron dense chlamydial forms, usually of 0.2-0.6μm diameter. The EBs attach to the host cell, then are endocytosed into vesicles known as "inclusions." There is no acidification or lysosomal fusion of these inclusions. Once internalised, the EBs differentiate into the metabolically active reticulate bodies (RBs), which can be up to 1.5μm in diameter. Nutrients are absorbed from the host cell, and the RBs undergo several rounds of replication. RBs then differentiate back into EBs, which are released from the cell either by host cell lysis, or the fusion of the inclusion and plasma membranes.

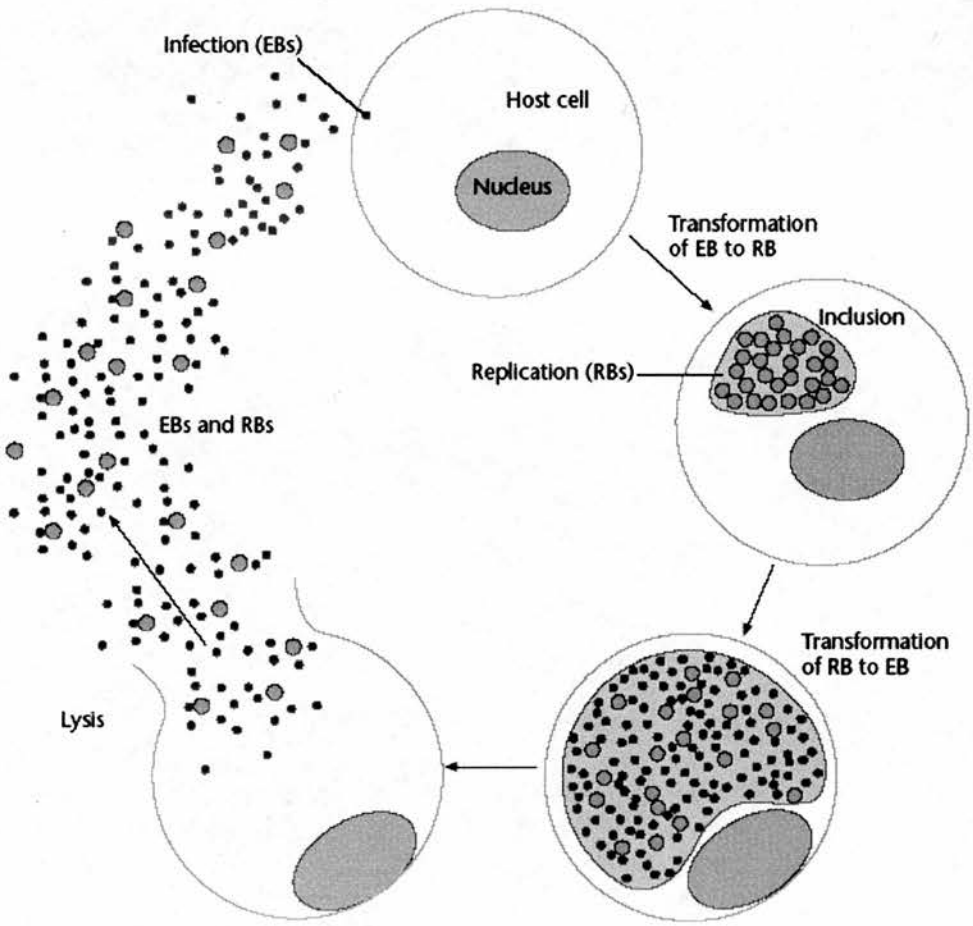


Figure 1.2: The Chlamydiales developmental cycle. Ususally takes 2-3 days. EBs are elementary bodies, RBs are reticulate bodies. From Everett (2002).

Chlamydiae do not require a secondary host as the EBs are highly stable in an extracellular environment, due to an unusual membrane structure. Chlamydiae are Gram-negative and have both inner and outer membranes however, unlike other Gram-negative bacteria, there is no detectable peptidoglycan layer. Instead there is a highly cross-linked network of disulphide bonds between envelope proteins containing numerous cysteine residues (Hatch.1996). Extracting EBs with the mild detergent Sarkosyl leaves proteinaceous chlamydial outer membrane complexes (COMCs) that resemble intact, but empty, EBs (Caldwell et al.1981). COMCs are predominantly composed of three proteins; the Major Outer Membrane Protein, OmcA and OmcB.

1.3 Chlamydial envelope proteins

1.3.1 The Major Outer Membrane Protein (MOMP)

MOMP was first reported by Caldwell *et al.* as a 40kDa, surface-exposed, chlamydial protein and the main protein in the *C. trachomatis* COMC, comprising over 60% of the total (Caldwell et al.1981). It was subsequently confirmed to be present in the other chlamydial species, and expressed throughout the developmental cycle (Salari et al.1981; Hatch et al.1981). Liposome-swelling assays and electrophysiological analysis suggested that MOMP functioned as a porin, similar to those in other bacterial membranes, eg. OmpF in *E. coli* (Bavoil et al.1984; Wyllie et al.1998). Circular

dichroism analysis of purified MOMP supported this hypothesis, with the 62% β-sheet measured comparative to known porins.

The gene encoding MOMP (*ompA*) was first cloned and sequenced by Stephens *et al.* (Stephens et al.1986). Subsequent analysis of the *ompA* genes and translated amino acid sequences from the different species and serovars of chlamydiae revealed that MOMP is composed of five highly conserved regions interspersed with four "variable sequence" domains (figure 1.3). Epitope-mapping has shown that these domains correspond to the primary immunogenic regions within MOMP and are responsible for the antigenic differences between species and serovars (Baehr et al.1988; Yuan et al.1989). This has led to the hypothesis that the VS domains are surfaced-exposed loops within the predicted porin structure.

MOMP is an unusual porin, in that it contains between 7 and 9 well-conserved cysteine residues that play an important functional role. In the extracellular EBs, MOMP is oxidised and highly cross-linked, and the outer membrane is largely impermeable and very osmotically stable. In contrast RBs are osmotically fragile and most of the MOMP is monomeric (Hatch et al.1984). The pore-forming ability of MOMP was increased almost tenfold when fully reduced (Bavoil et al.1984). Together these suggest that MOMP as a porin can be opened and closed by the breaking and reforming of disulphide bonds. This is linked to the developmental stage of the bacteria. EBs treated with 10mM dithiothreitol (DTT) took on some of the characteristics of RBs,

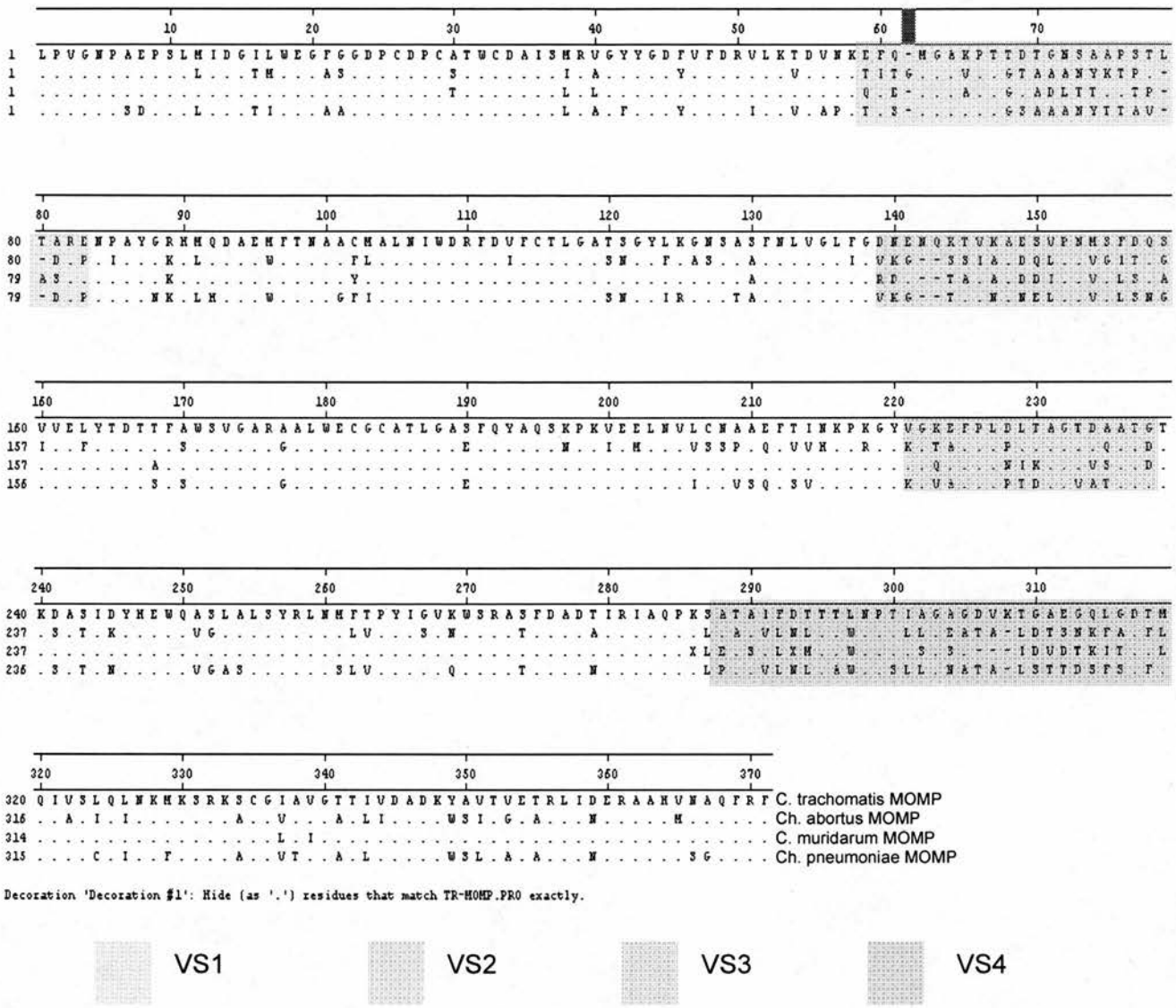


Figure 1.3: Alignment of MOMP from four chlamydial species. Only those amino acids that do not match the *C. trachomatis* sequence are displayed. The four variable sequence (VS) domains are highlighted. Alignment produced using Lasergene software.

namely increased metabolic activity, reduced infectivity, decreased osmotic stability and a change in staining properties (Hackstadt et al.1985)). The reducing agent was not sufficient in itself to initiate differentiation of the EBs into RBs, however the reduction of the disulphide network, perhaps as a result of reducing conditions within the inclusion vesicle, is likely to be an important early step of the process. The increased flexibility of the membrane would be essential for the RBs to grow and divide. MOMP solubilised from EBs and RBs in the presence of reducing agent ran as a monomer on SDS-PAGE. However MOMP solubilised in non-reducing conditions was detected as monomers, dimers, trimers, tetramers and as larger complexes that did not enter the polyacrylamide gel (Newhall et al.1983; Hackstadt et al.1985). Purified MOMP has been shown to run as a probable trimer on SDS-PAGE (McCafferty et al.1995; Wyllie et al.1998). The precise oligomeric state of MOMP seems to be complex, involving disulphide bonds, and differing between chlamydial species and developmental stages.

MOMP may have a further role in chlamydial biology as an adhesin during the initial attachment of the EB to the host cell. It seems likely that attachment is a complex process involving numerous chlamydial and host molecules. N-acetylneurminc acid, N-acetylglucosamine, heparan sulphate and a glucosamino glycan produced by the chlamydiae themselves have all been implicated (Zhang et al.1992; Taraktchoglou et al.2001; Wuppermann et al.2001). There is also evidence that MOMP plays a role in attachment. Trypsin digestion of *C. trachomatis* MOMP and antibody-blocking of VS2 and VS4 have been shown to inhibit the infectivity of EBs (Su et al.1988; Su et al.1990b). Additionally, MOMP is N-glycosylated with mannose oligosaccharide, a

sugar that can competitively inhibit the attachment of EBs to HeLa cells. Mannose-binding protein also has a protective effect (Swanson et al.1994; Swanson et al.1998). A comparison of infection of mannose-receptor positive and negative mouse macrophages showed different preferences for binding by several chlamydial species (Kuo et al.2002). It seems likely that the species of chlamydiae would display differences in attachment as they have evolved to predominantly infect different cell and tissue types. These findings indicate that MOMP may be involved in both specific and non-specific interactions during attachment to the host cell. Although it is likely that other molecules are also involved, the prevalence of MOMP on the surface of the bacteria would make its role significant.

Antigenic analysis has demonstrated that MOMP is the immunodominant chlamydial protein (Zhang et al.1987) and as such is considered to be a primary vaccine target. Early studies showed that antibodies raised against *C. trachomatis* MOMP had a neutralising effect on the infectivity of EB in cell culture (Caldwell et al.1982). Several approaches have been taken to creating a vaccine based on MOMP with varying degrees of success including intact EBs, COMCs, MOMP and DNA vaccination with *ompA* (Sandbulte et al.1996; Zhang et al.1997; Penttila et al.2000; Dong-Ji et al.2000; Pal et al.2001), and although no fully protective vaccine has yet been developed encouraging progress has been made (Christiansen et al.2002). It has been shown that T_{H1} immune responses and interferon-γ (IFN-γ) play a more critical role both in clearance of primary infection and protection against secondary infection than humoral immunity (Johansson

et al.1997a; Johansson et al.1997b). Several epitopes that stimulate T-cell proliferation have been identified in MOMP, again within the VS domains (Su et al.1990a; Allen et al.1991; Ishizaki et al.1992), but these have been focussed on obtaining high antibody titre, a T_{H2} response. However, a T-cell epitope has been identified on MOMP that elicits a proliferation of CD4+ and CD8+ T-cells and the production of IFN-γ (Stagg et al.1993; Knight et al.1995), that may be useful for further vaccine development.

1.3.2 The Cysteine Rich Proteins (CRPs)

The other main components of the COMC are the two cysteine rich proteins, OmcA and OmcB, in an ratio of approximately 1 OmcB:2 OmcA:5 MOMP (Everett et al.1991). Both genes are encoded by a bicistronic operon and are only expressed late in the developmental cycle as the RBs are reorganising into EBs (Newhall.1987; Lambden et al.1990). In *Ch. psittaci,* OmcA is a small protein of 12kDa, yet contains 14 cysteine residues and is predicted to be a lipoprotein (Everett et al.1994). OmcB is a 60kDa protein that is post-translationally modified to give a doublet on SDS-PAGE (Allen et al.1989), and contains 37 cysteine residues. The sequences of both proteins are well conserved (Watson et al.1989; de la Maza et al.1991). Although both proteins are in the Sarkosyl-insoluble COMCs, this is as a result of the disulphide cross-linking and not because they are integral membrane proteins. Under reducing conditions, OmcB partitioned into the aqueous rather than the detergent phase, and though OmcA remained associated with the membrane, it was soluble in Sarkosyl (Everett et al.1995). Neither

protein could be detected by immunogold labelling on the cell surface of intact EBs (Collett et al.1989; Watson et al.1994; Mygind et al.1998). It has therefore been proposed that in EBs OmcB is extensively cross-linked in the periplasm, and forms disulphide bonds to MOMP and to OmcA, which is associated with the inner leaflet of the bilayer (Everett et al.1995). In RBs the CRPs are not expressed and MOMP is reduced and functioning as a porin, allowing the influx of nutrients from the host cell and the expansion of the bacterial cell.

1.4 Recombinant MOMP

A major obstacle to the study of MOMP is the difficulty in obtaining useable quantities of purified protein. As chlamydiae must be grown intracellularly and at a high level of containment, culturing the organisms in large numbers is both technically difficult and expensive. Additionally, the interactions of MOMP with the other CRPs in the COMC make it difficult to ensure that experimental data are attributable to MOMP alone. For these reasons a recombinant system of expression would be preferable.

Several attempts have been made to express full length MOMP in *E. coli*. Most have taken the approach of cloning the full length *ompA* gene including the signal sequence, in order to target the translated protein to the *E. coli* outer membrane. Dascher *et al.* and Manning *et al.* demonstrated the expression, translocation across the inner membrane, and signal peptide cleavage of MOMP (Dascher et al.1993; Manning et al.1993). However, both groups found that the protein was not surface expressed and concluded it

was instead misfolded and aggregated in the periplasm. Similarly Kaul *et al.* produced translocated and processed MOMP that was associated with the outer membrane (Kaul et al.1990). However they provided no evidence of surface exposure and did not test for native conformation of the protein. Koehler *et al.* showed a portion of expressed MOMP was surface-exposed by immunodetection, however there was a dramatic loss of cell viability in this system during expression (Koehler et al.1992), so cell lysis or disrupted membranes may have affected the data. The results therefore should be interpreted with caution. Finally, Jones *et al* produced a synthetic MOMP gene expressed in *E. coli*. Again the protein was localised to the outer membrane, but neither surface-exposure nor native conformation was examined (Jones et al.2000). Wyllie *et al.* produced truncated versions of *Ch. abortus* and *Ch. pneumoniae* MOMP and successfully extracted small quantities from inclusion bodies in detergent (Wyllie et al.1999). However, this method does not provide enough protein for many downstream analysis techniques.

Expression systems other than *E. coli* have been tried. Recombinant MOMP has been expressed in COS cells (Vanrompay et al.1998) and in *Vibrio cholerae* (Eko et al.2003) for use as vaccine delivery vechicles. Both showed promising signs of surface exposure, but face the same difficulties in obtaining native conformation as MOMP expressed in *E. coli*.

A practical method of expression of recombinant MOMP has yet to be developed. The availability of such a system would greatly aid the development of vaccines and the study of MOMP and chlamydial biology.

1.5 PorB

PorB is another putative porin expressed in the chlamydial outer membrane. It is smaller than MOMP with a molecular weight of 37kDa. It is much less abundant in the membrane than MOMP but is also expressed throughout the developmental cycle (Kubo et al.2000). Immunodetection experiments on intact EBs confirmed PorB was surface-exposed and antibodies raised against the antigenic regions of the protein were able to neutralise *in vitro* infectivity (Kawa et al.2002). Antibodies against PorB have been identified in human sera (Sanchez-Campillo et al.1999). It is therefore considered to be a second possible vaccine target. Although there is much less PorB in the chlamydial outer membrane than MOMP, it has the advantage of better sequence conservation between serovars (Kubo et al.2000).

Due to its low abundance, it has been proposed that PorB is a substrate-specific rather than a general diffusion porin. Recombinantly expressed PorB was shown to be specific for 2-oxogluterate and other select dicarboxylates (Kubo et al.2001). However, in this system the PorB was expressed with a C-terminal His tag. As the N- and C-terminals of porins generally meet in a β-strand it is difficult to see how the His tag could fail to

affect folding and/or function of the porin. The precise specificity of PorB should therefore not be assumed until confirmed by alternate methods.

PorB is an attractive secondary vaccination target due to its immunogenicity and relative conservation. Additionally, its role as a substrate-specific porin may be important in chlamydial-host interactions.

1.6 Study aims

The central aim of this study was to establish a recombinant system for the expression of MOMP in its native fold. Protein obtained would then be used to investigate the structural and functional properties of MOMP. If possible, the system could be expanded to also express other chlamydial outer membrane proteins.

Two basic approaches were taken simultaneously to expressing MOMP in *E. coli*. The first approach was to express MOMP without a leader sequence as insoluble inclusion bodies in the bacterial cytoplasm. Protein would then be solubilised and denatured, and refolded *in vitro*. The second approach was to express MOMP directly to the *E. coli* outer membrane, targeting it using either MOMP's own leader sequence or a leader sequence from a known *E. coli* outer membrane protein. Developing two different methods was intended to increase the likelihood of success, as MOMP has proved difficult to produce recombinantly in the past.

This thesis outlines the methodology used and the level of success of these expression systems. The different properties of MOMP within these systems are also examined. The study furthers the understanding of the biochemistry of chlamydial MOMP and contributes to the drive for vaccine development.

Chapter 2: Materials and methods

2.1 Materials

Unless otherwise indicated, chemicals were supplied by either Sigma Aldrich or Fisher, detergents and lipids were supplied by Anatrace, and enzymes were supplied by New England Biolabs Inc.

2.2 Prediction algorithms

Three artificial neural network-based secondary structure prediction algorithms were used to analyse MOMP. Firstly, the Diederichs outer membrane protein topology prediction program, available at:

http://strucbio.biologie.uni-konstanz.de/~kay/om_topo_predict2.html

using the new set of weights (Diederichs et al.1998). Secondly, the TMBETA transmembrane β-strand prediction program (Gromiha et al.2004), available at:

http://psfs.cbrc.jp/tmbeta-net/

and finally the B2TMPRED transmembrane β-strand prediction program (Jacoboni et al.2001), available at:

http://gpcr.biocomp.unibo.it/cgi/predictors/outer/pred_outercgi.cgi

2.3 Agarose gel electrophoresis

DNA was separated by agarose gel electrophoresis. Gels were prepared by mixing 1% (w/v) agarose powder in TBE (45mM Tris-borate 1mM EDTA) and heating until the agarose dissolved. 0.5μg/ml ethidium bromide was added, and the molten agarose poured into moulds and allowed to set. DNA was mixed 1:1 with loading buffer (75mM Tris-HCl pH 7.6, 30% (v/v) glycerol, 1.25mg/ml bromophenol blue) and loaded into the wells. Electrophoresis was carried out at 150V using a Biorad PowerPac 3000, in a Biorad Sub Cell GT tank. Gels were viewed on a UVP GelDoc UV transilluminator.

2.4 Polymerase chain reaction (PCR)

2.4.1 Gene amplification

The *porB* gene, with and without its leader sequence, was amplified directly from *C. muridarum* genomic DNA (kindly provided by Dr Richard S Stephens and colleagues, UC Berkeley, CA). Primers were designed at the 5′ and 3′ ends of the gene. The sense primers contained a 5′ overhang of the restriction site *Nde*-I (using the start codon as its internal ATG) plus enough extra bases to allow subsequent restriction digest (New England Biolabs Inc.2002). The same anti-sense primer was used for both constructs and recognised downstream of the stop codon and included the restriction site *Nco*-I plus extra bases to provide an overhang for digest.

PCR was carried out using approximately 10ng of genomic DNA, 25pmol sense primer and 25pmol anti-sense primer in a 50µl reaction also containing 20mM Tris-HCl pH8.8, 10mM $(NH_4)_2SO_4$, 10mM KCl, 0.1% (v/v) Triton X-100, 0.1mg/ml BSA, 2mM $MgSO_4$, 0.2mM dNTPs and 3U *Pfu* polymerase. A 5min denaturing hold at 94°C was followed by 25 cycles of 30sec at 94°C, 30s at the annealing temperature (see appendix B) and 2min at 72°C, then a final extension of 7min at 72°C.

2.4.2 Site-directed mutagenesis (SDM)

2.4.2.1 SDM of plasmid DNA

The *ompA* gene from *C. trachomatis* had already been cloned into the pSTBlue-1 plasmid (R.H. Ashley, unpublished data). For cloning into the expression vector the restriction site at the 5′ end of the gene was changed to *Nde*-I and an internal *Nde*-I site was removed. These mutations were introduced using the QuikChange method (Stratagene). Overlapping primers were designed to contain the desired mutation with approximately 10-20 base pairs either side of the mutation site. ~10ng of plasmid, 12.5pmol sense primer and 12.5pmol anti-sense primer was added to a 50 µl reaction mix (as in section 2.4.1) and amplified around the entire plasmid with a standard protocol of 18 cycles consisting of 30sec at 94°C, 30sec at the annealing temperature and 10min at 72°C.

The product was digested with *Dpn*-I for 1hr at 37°C to degrade the template DNA then transformed into Novablue cells (Novagen).

The same method was used to create the cysteine-less mutant and the internal mutagenesis in *C. trachomatis ompA*. All nine cysteine residues in *C. trachomatis* MOMP were mutated to alanine residues in the pSTBlue-1 cloning vector, before subcloning into the pET22b(+) expression vector. Also in the cloning vector, unique restriction sites were introduced either end of each VS domain, allowing subsequent digestion and removal from the gene (see appendix B).

2.4.2.2 SDM from genomic DNA

The *ompA* gene of *C. muridarum* was amplified directly from genomic DNA, but required the removal of an internal *Nde*-I site during initial amplification to allow subsequent cloning into the pET22b(+) vector. Four primers were designed; two cloning primers and two overlapping mutation primers as described above, making a single base pair change to destroy the *Nde*-I site. The sense cloning primer contained an *Nde*-I restriction site and the anti-sense cloning primer contained a *Bam*H I restriction site. One reaction was set up using the sense cloning primer and the anti-sense mutation primer, and a second using the sense mutation primer and the anti-sense cloning primer. Both reactions were carried out in the conditions described in the previous section, over 18 cycles of 30sec at 94°C, 30sec at the annealing temperature and 2min at 72°C. The products were separated by agarose gel electrophoresis then purified by gel extraction.

1μl of each gel-purified product was added to a third PCR reaction without any primers but the same other conditions, and amplified for 8 cycles. The reaction mix was then spiked with 25pmol of each of the two cloning primers and continued for another 20 cycles.

2.4.3 Gene extension

The native chlamydial and *ompT E.coli* leader sequences were added to the 5′ end of genes by sequential PCR. A series of three overlapping sense primers were made, each adding more of the leader sequence on the 5′ end of the gene. The third primer contained the start codon and an *Nde*-I restriction site. One anti-sense primer was made containing an *Nco*-I site. 12.5pmol of sense primer 1 and 12.5pmol anti-sense primer were added to a 50μl reaction tube also containing ~10ng template DNA and the same buffer conditions as before. Amplification was by the standard 18 cycle protocol. The product was separated by agarose electrophoresis and purified by gel extraction. 1μl of this was used as the template for the second round of PCR using sense primer 2. This was repeated with sense primer 3 to give the complete product.

2.5 Gel purification

Bands were excised from the agarose gel and purified using a DNA-binding Perfectprep™ Gel Cleanup Kit (Eppendorf) according to the manufacturer's instructions. DNA was routinely eluted in the smaller recommended volume of 30µl.

2.6 Restriction digestion

2.6.1 Sequential digest

The pET22b(+) vector was prepared for cloning by long sequential digest, to maximise the yield of double digested DNA, when the restriction sites are close to each other. Similarly, the sticky ends of genes for cloning directly into pET22b(+) were prepared by this method because of the reduced efficiency of digestion near the end of DNA strands. A 20µl reaction was set up containing 3µg vector, or 10µl gel-purified insert, 50mM K-acetate, 20mM Tris-acetate, 10mM Mg-acetate, 1mM DTT, pH 7.9 and 10U *Nco*-I and incubated at 37°C overnight. With the *C. muridarum* clones, a similar reaction was set up with vector/insert, 150mM NaCl, 10mM Tris-HCl, 10mM $MgCl_2$, 1mM DTT, 0.1mg/ml BSA, pH 7.9 and 10U *Bam*H-I and also incubated overnight at 37°C. In each case, 10U of *Nde*-I were then added to the reaction before incubating for a further 8 hours at 37°C.

2.6.2 Double digestion

For subcloning, the DNA was cut out by double digest. 3μg of plasmid was added to a 20μl reaction buffered as for sequential digestion, but with10U *Nco*-I and 10U *Nde*-I added, then incubated for 3hr at 37°C. The products were separated by agarose gel electrophoresis and the insert gel purified.

2.6.3 Domain digest

The VS domains were digested out of the *C. trachomatis ompA* gene by adding 3μg of DNA to a 20μl reaction containing 2μl 10x reaction buffer and 20U of the relevant restriction enzyme and incubated at 37°C or 50°C (*Bcl*-I) for 3hr.

2.7 Alkaline phosphatase

Phosphate groups were removed from the cut ends of vectors to prevent re-ligation giving rise to false positives. 1U of shrimp alkaline phosphatase (Helena Biosciences) was added per pmol end of vector in 50mM Tris-HCl pH 8.5, 5mM $MgCl_2$ according to the equation:

$$pmol\ ends\ (linear\ double\ stranded\ DNA) = \frac{\mu gDNA}{kbDNA} \times 3.04$$

and incubated for 10min at 37°C. The alkaline phosphatase was removed using a Perfectprep™ Gel Cleanup column according to the manufacturer's instructions, eluting in 30µl.

2.8 Ligations

2.8.1 Cloning/subcloning

Alkaline phosphatased pET22b(+) and the gel purified insert genes were ligated in a 10µl reaction containing 1µl vector, 50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 10mM DTT, 1mM ATP, 25µg/ml BSA, 3µl insert and 3U of T4 DNA Ligase (New England Biolabs), and incubated overnight at 16°C. The ligase was inactivated by heating for 10min at 70°C, then 3U *Msc*-I was added and the reaction incubated at 37°C for 3hr. In the pET22b(+) multiple cloning site, the *Msc*-I site is between the *Nde*-I and *Nco*-I restriction sites, and therefore the enzyme will only digest empty vectors that would give rise to negative colonies.

2.8.2 Domain deletion

Digested *C. trachomatis ompA* in vector pSTBlue-1 was re-circularised without the deleted domain, in a 10µl ligation reaction containing 1µl DNA, in the same reaction

conditions as above, and incubated at 16°C overnight. The mutagenised *ompA* gene was then subcloned into pET22b(+).

2.9 Preparation of media and agar

2.9.1 Luria-Bertani medium and agar

Luria Bertani (LB) medium was prepared by mixing 0.5% (w/v) yeast extract, 1% (w/v) bactotryptone and 0.5% (w/v) NaCl in dH_2O, and buffering to pH 7.4 using NaOH. LB agar was made by adding 15% (w/v) agar powder to LB media. The media and agar were aliquoted into flasks and autoclaved. To make agar plates, ampicillin was added to the concentration of 100μg/ml to the liquid agar then poured out approximately 20ml per plate into petri dishes, the surface flame sterilised, and allowed to set.

2.9.2 SOC medium

SOC medium was made by mixing 0.5% (w/v) yeast extract, 2% (w/v) bactotryptone and 0.05% (w/v) NaCl in dH_2O and buffering to pH 7.0 using NaOH. The media was aliquoted into flasks and autoclaved. Directly before use 5ml sterile 2M $MgCl_2$ and 20ml sterile 1M glucose were added.

2.10 Cell stocks

2.10.1 Glycerol stocks

Bacteria were stored as glycerol stocks at -70°C. To make these, a single colony of transformed bacteria was picked with a toothpick or a pipette tip and grown overnight in 5ml LB media, containing antibiotic where appropriate. This culture was re-seeded 1:100 into fresh media and grown until mid-exponential phase (OD_{600} = 0.6). Sterile glycerol was added to 10% (v/v) and the cells were aliquoted into sterile 2ml tubes and frozen rapidly in a mix of dry ice and isopropanol.

2.10.2 Competent cells

Cells were scraped from a glycerol stock and added to 5ml of LB media and grown at 37°C overnight. For Novablue (Novagen) and Δomp8 cells, 10μg/ml tetracycline was added to the media. 2.5ml of this culture was added to 250ml fresh LB media, again containing tetracycline where appropriate, and grown with shaking at 37°C until it reached OD_{600} ~0.4-0.6. The cells were harvested by centrifuging for 5min at 4500 x g at 4°C in a sterile container. The resulting pellet was re-suspended in 100ml TFB1 buffer (30mM potassium acetate, 10mM $CaCl_2$, 50mM $MnCl_2$, 100mM $RbCl_2$, 15% (v/v) glycerol, pH 5.8) and incubated on ice for 5min. The cells were pelleted as before then re-suspended in 10ml TFB2 buffer (10mM MOPS, 75mM $CaCl_2$, 10mM $RbCl_2$,

15% (vv) glycerol, pH 6.5) and incubated on ice for 15-60min. 200µl aliquots were transferred into sterile 1.5ml tubes and rapidly frozen in dry ice/isopropanol then stored at -70°C.

2.11 Transformations

Aliquots of competent cells were thawed on ice, then 1-5µl DNA was added and the cells incubated on ice for 30min. The bacteria were heat shocked for 45sec at 42°C then incubated on ice for a further 2mins. 200µl of fresh medium was added, followed by incubation at 37°C for 15-30min with shaking. 100µl of transformed cells were plated out onto agar plates containing 100µg/ml ampicillin, and incubated overnight at 37°C.

2.12 DNA preparation

Mini, midi and maxi preparations of DNA (with maximum yields of 20, 200 and 500µg respectively) were made using the QIAprep® Miniprep Kit, HiSpeed™ Plasmid Midi Kit and QIAfilter™ Plasmid Maxi Kit (Qiagen), all according to the manufacturers instructions.

2.13 Screening

2.13.1 Colony PCR

Transformed colonies were screened for the presence of the inserted gene by PCR amplification using the relevant cloning primers. Each reaction totalled 50µl and contained 10mM Tris-HCl pH 9.0, 50mM KCl, 1% (v/v) Triton-X-100, 1.5mM $MgCl_2$, 25pmol sense primer, 25pmol anti-sense primer, 1µl dNTPs and 2.5U *Taq* DNA polymerase (Promega). A colony was picked using a toothpick/pipette tip and spotted onto an agar plate, which was incubated overnight at 37°C. Remaining cells were then shaken off into the PCR reaction mix. Amplification consisted of a 5min hold at 94°C, 25 cycles of 30sec at 94°C, 30sec at the annealing temperature and 30sec at 72°C, and finally a 7min hold at 72°C. Positive colonies were determined by agarose electrophoresis of the PCR products, then cultured from the agar plate.

2.13.2 Hemi-nested screen

The successful deletion of a domain was tested by a hemi-nested PCR. Reaction mixes were set up containing 10mM Tris-HCl pH 9.0, 50mM KCl, 1% (v/v) Triton-X-100, 1.5mM $MgCl_2$, 1µl dNTPs and 0.5µl *Taq* polymerase. 25pmol 5′ sense mutagenesis primer, 25 pmol 3′ anti-sense cloning primer and ~10ng DNA was added. Clones where the section between the two mutagenesis sites has been digested away were unable to

amplify compared to the positive control. Amplification conditions were as in the previous section.

2.14 Protein expression

2.14.1 Mature protein (lacking a signal sequence)

Constructs without signal sequences were transformed into BL21(DE3) cells (Novagen) and Δomp8 cells, and stored as glycerol stocks.

A scraping of glycerol stock was added to 5ml of LB media, containing 100μg/ml ampicillin. Δomp8 cells also had 10μg/ml tetracycline added. This was grown overnight at 37°C with shaking, then reseeded 1:100 into 250ml fresh LB media containing ampicillin. This was incubated at 37°C until the OD_{600} = 0.6. Cells were harvested by centrifugation at 6000 x g for 5mins in a Beckman Avanti™ J-25 centrifuge, and re-suspended into fresh LB media containing ampicillin. The culture was replaced in the 37°C incubator and allowed to equilibrate to temperature for 10min. Isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1mM to induce the plasmid, and the culture incubated for a further 3hr at 37°C.

2.14.2 Full length protein (with a signal sequence)

Full length constructs were also transformed into BL21s and Δomp8 cells and glycerol stocks prepared. An overnight culture in LB was grown to saturation overnight as for mature protein. This was re-seeded 1:100 into 1 litre of SOC medium containing 100μg/ml ampicillin and incubated at 37°C until the OD_{600} = 0.6. A media change was carried out as described in the previous section. The cells were resuspended into 1 litre of fresh SOC with antibiotic and incubated at 16°C for 10min. IPTG was added to a final concentration of 0.1mM and the culture incubated at 16°C overnight.

2.15 Inclusion body preparation

Cells were harvested from induced cultures by centrifugation at 6000 x g for 5min and washed in 50ml phosphate buffered saline (PBS). The pellet was resuspended in 5ml of TEN buffer (50mM Tris, 10mM EDTA, 100mM NaCl, pH 8.0) and 0.2mg/ml lysozyme added. After a 30min incubation at room temperature, cells were lysed by sonicating with 6 x 15sec blasts at 6μm using a Sanyo Soniprep 150 sonicator.

Membrane-associated proteins in the cell lysate were solubilised by incubation for 30min at 37°C with 0.8mg/ml sodium deoxycholate. Nucleic acids were degraded by incubation for 15min at room temperature with 20U/ml Benzonase (Novagen). The inclusion bodies were isolated by centrifugation at 12,000 x g for 10min. The resulting

whitish pellet was washed repeatedly with 10ml IB wash buffer (TEN buffer, with 2M urea, 1% (v/v) Triton-X-100) until the supernatant was colourless. A final wash with 10ml TEN buffer removed traces of urea and detergent. The IB pellet was dissolved in 2ml of 6M GuHCl with 10mM DTT. This solution was ultracentrifuged for 20min at 100,000 rpm (~300,000 x g) in a Beckman TL100 Ultracentrifuge to pellet any large aggregates and remaining cellular debris.

2.16 Outer membrane protein isolation

Cells were harvested and lysed by sonication as in the inclusion body preparation. The cell lysate was incubated with with 20U/ml Benzonase for 15min at room temperature. The outer membrane fragments were pelleted by centrifugation at 15,000 x g for 10min and the supernatant discarded. The pellet was washed twice in 20ml TEN buffer. Membrane proteins were extracted by resuspending the pellet in 6ml solubilisation buffer (50mM Tris pH 8, 1mM EDTA, 50mM NaCl, 10mM DTT, 1% (v/v) detergent) and incubating at 37°C for 1hr. The solution was clarified by ultracentrifuging for 20min at 100,000 rpm.

2.17 Protein concentration assay

The concentration of protein was determined using the Marwell-Lowry assay (Markwell et al.1978) to avoid interference from detergent and lipid. Three samples of a 1:10 dilution of denatured protein, of 30µl, 15µl and 7.5µl were aliquoted into 1.5ml tubes.

In addition, a set of standards were set up containing 0, 5, 10, 15, 20, 25μg bovine serum albumin. To each tube was added 50μl of 10mg/ml Na-deoxycholate, before vortexing. 1ml of 10% (w/v) trichloroacetic acid was added to each tube, which were vortexed again then centrifuged for 2min at 13,000 rpm in a Biofuge pico (Heraeus Instruments) benchtop centrifuge. The supernatant was discarded. Alkaline copper reagent was freshly prepared from a 100:1 mixture of solution A (200mM Na_2CO_3, 100mM NaOH, 7mM K Na Tartrate, 1% (w/v) SDS) with solution B (4% (w/v) $CuSO_4.5H_2O$). The pellets were resuspended in 1ml alkaline copper reagent. 100μl of a 1:1 dilution of Folin Reagent in dH_2O was added with each tube being vortexed immediately. After 1hr incubation at room temperature the A_{750} was measured. The standards absorbances were plotted and the protein concentration calculated.

2.18 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated by polyacrylamide gel electrophoresis using the Biorad Protean 3 Mini Gel System. 10% or 12% separating gels were prepared from 40% (w/v) acrylamide/bis-acrylamide (37.5:1) solution in 375mM Tris pH 8.8, 0.1% (w/v) sodium dodecyl sulphate (SDS), 0.1% (w/v) ammonium persulphate (APS) and 0.1% (w/v) N,N,N′,N′-tetramethylethylenediamine (TEMED). Stacking gels were composed of 4% (w/v) acrylamide/bisacrylamide, 125mM Tris pH 6.5, 0.1% SDS (w/v), 0.1% (w/v) APS and 0.1% (w/v) TEMED. Gels were run at 150V on a Biorad PowerPac 300 for 60-

120min. Precision Plus Unstained Protein Markers and Kaleidoscope Pre-stained Protein Markers (BioRad) were used as molecular weight standards.

2.19 Staining

2.19.1 Coomassie stain

Polyacrylamide gels were stained for 30min in 40% (v/v) methanol, 10% (v/v) acetic acid, 0.25% (w/v) Brilliant Blue R-250, then destained by several changes of 40% (v/v) methanol, 10% (v/v) acetic acid.

2.19.2 Colloidal stain

Polyacrylamide gels were stained with GelCode (Pierce) for 1hr, then destained by several changes of dH_2O.

2.20 Western blot

2.20.1 Wet transfer

Gels were transferred on to Hybond-P polyvinylidene difluoride (PVDF) membranes (Amersham Biosciences) pre-activated by washing for 10sec in methanol. Transfers

were carried out on ice in a Mini Trans-Blot Cell tank transfer system (Biorad) for 60min at 100V on a PowerPac 300 power supply.

2.20.2 Blotting

Membranes were blocked for 30min at room temperature or overnight at 4°C in 5% (w/v) Marvel in PBS-T (0.005% (v/v) Tween 20 in PBS) then incubated in a 1:5000 dilution of goat anti-trachomatis MOMP antibody (Fitzgerald International, Inc) for 1hr at room temperature. The primary antibody was removed with 2 x 30sec washes and 3 x 5min washes of PBS-T. Membranes were incubated in a 1:10,000 dilution of HRP-conjugated anti-goat/sheep antibody (Sigma) for 1hr at room temperature. The secondary antibody was washed off as before.

2.20.3 Detection

Western blots were visualised by electrochemiluminescence (ECL). Equal volumes of ECL solution 1 (100mM Tris-HCl pH 8.5, 2.5mM luminol, 0.4mM coumaric acid) and solution 2 (100mM Tris-HCl pH 8.5, 0.019% (v/v) H_2O_2) were combined and the membrane incubated in the mixture for 1min. In a dark room the membrane was exposed to medical film (Konica) for 15sec to 30min. The film was developed in a Konica SRX-101A X-ray developer.

2.21 Dot Blots

2.21.1 Whole cell immunoblots

A fresh glycerol scraping was added to 5ml of LB media containing 100μg/ml ampicillin and incubated at 37°C overnight. This was re-seeded 1:100 into 10ml fresh media and antibiotic and grown at 37°C until OD = 0.6. The media was changed by centrifuging the culture for 10min at 6000 x g, discarding the supernatant and resuspending the cells in fresh media and antibiotic. The culture was incubated at 25°C for 10min, then 1mM IPTG was added and it was incubated for a further 2hr.

Intact cells were harvested by centrifuging for 5min at 4500 x g and washed in 1ml PBS. The pellets were resuspended in 200μl PBS. 10μl was spotted on to nitrocellulose membrane (Biorad) and allowed to dry. The membrane was then blocked, blotted and exposed as for Western blotting.

2.21.2 Dot blots

Preactivated PVDF membrane was allowed to air dry. 10μl protein samples were dotted onto the membrane. After drying, the protein was fixed by a 5min incubation in methanol and the membrane rinsed in dH_2O. The membrane was then blocked, blotted and exposed as for Western blotting.

2.22 Chromatography

A Pharmacia FPLC™ system was used for Fast Performance Liquid Chromatography. Gel filtration chromatography was carried out on a HR 26/60 HiLoad Superdex 200 prep grade column (Amersham Biosciences). The column was pre-equilibrated with 2 column volumes of equilibration buffer (50mM Tris pH 8.0, 1mM EDTA, 50mM NaCl, 5mM DTT, 0.05% (w/v) LDAO or Zwittergent 3,14). 2ml of solubilised outer membrane protein was applied and the column run with equilibration buffer for 800min at a flow rate of 0.5ml/min. 5ml fractions were collected on a FRAC-200 fraction collector.

2.23 Planar lipid bilayer

A suspension of diphytanoyl phosphatidylcholine in decane was used to form bilayers across a 0.3mm hole in a polystyrene partition separating two solution filled chambers. Bilayers used had a capacitance of 200-300pF and a conductance of 5-10pS. Both chambers were filled with 50mM KCl, 10mM Tris, pH 7.4 and 1mM DTT and the "*cis*" chamber was voltage-clamped at 0mV using an Axopatch 200B amplifier. The "*trans*" chamber was grounded. Channel incorporation was induced by adding aliquots of 3M KCl to the *cis* chamber to give a concentration gradient of 500:50 mM KCl *cis:trans*, before adding ~1ng/ml MOMP to the *cis* chamber, then repeatedly switching the membrane potential between +60 and –60mV. Following the appearance of unit transmembrane currents, implying channel insertion into the membrane, the *cis* chamber

was perfused with a minimum of 10 volumes of 50mM KCl, 10mM Tris,pH 7.4, to minimise the incorporation of further channels. Subsequent changes in salt concentration were effected by the addition of 3M KCl aliquots or perfusion, as required.

Traces were low-pass filtered 50Hz-1kHz (8 pole Bessel filter), and digitally recorded and analysed using pClamp 8 software.

2.24 Circular dichroism

Samples were dialysed into 50mM sodium phosphate pH 8.0, containing 2mM DTT and detergent at a concentration slightly above its critical micelle concentration (CMC). Circular dichroism spectra were obtained in a Jasco J-810 spectropolarimeter using 0.05 cm cells, scanning between 260nm and 190nm at 10nm/min and a data pitch of 0.2nm. For each sample, 5 spectra were accumulated, the buffer-only baseline spectrum was subtracted and the data were analysed using CDSSTR (Compton et al.1986; Manavalan et al.1987), CONTINLL (Provencher et al.1981) and SELCON3 (Sreerama et al.1993) and reference set 7 (Sreerama et al.2000a; Sreerama et al.2000b) from the Dichroweb website (Lobley et al.2002).

Chapter 3: Secondary structure and topology predictions of *C. trachomatis* MOMP

3.1 Neural Network Topology Prediction

The sequence of *C. trachomatis* MOMP was analysed using an artificial neural network that had been trained with a set of outer membrane porins of known, high-resolution structure (Diederichs et al.1998). The neural network predicts the z-coordinate of porin amino acids in a frame where the xy plane is the outer membrane and the transmembrane pore is aligned along the z-axis. Low z-values, <0.4, are predicted to correspond to residues in periplasmic turns and high z-values, >0.6, to residues in extracellular loops. Residues predicted to be in the transmembrane β-strands have values between these extremes.

A graph of these calculated z-coordinates against the sequence residue numbers was plotted (figure 3.1A, see appendix A.1 for original data). The resulting curve showed several clear maxima and minima, suggesting the position of the turns and loops. Of particular interest was the association of the VS domains with extracellular regions of the protein, as has been frequently inferred from other experimental data. However, there were regions of the plot where the positioning of loops and strands was less clear, so further prediction methods were applied to the sequence.

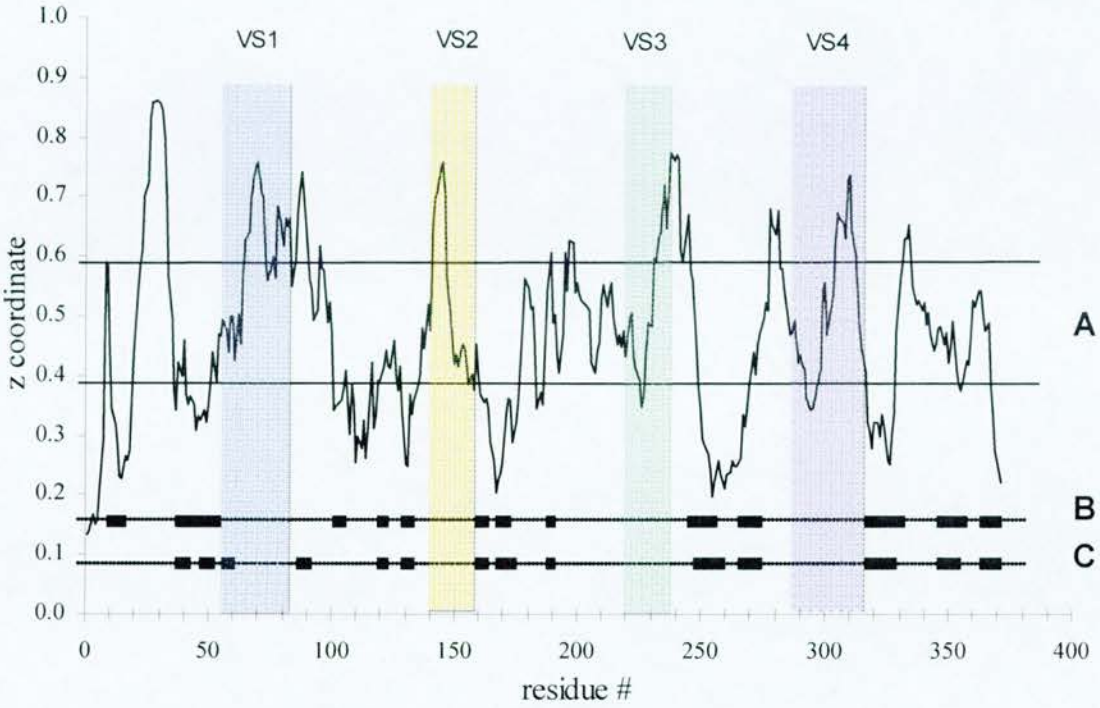


Figure 3.1: Secondary structure predictions of *C. trachomatis* MOMP. A. Diederichs topology prediction. The lines at 0.4 and 0.6 indicate the borders between predicted periplasmic and transmembrane residues, and extracellular and transmembrane residues respectively. **B.** TMBETA transmembrane β-strand prediction. Blocks indicated predicted strands. **C.** B2TMPRED transmembrane β-strand prediction. The four VS domains are high-lighted.

3.2 Transmembrane β-strand Prediction

Two different prediction algorithms were used to identify potential transmembrane β-strands in *C. trachomatis* MOMP; TMBETA (Gromiha et al.2004) and B2TMPRED (Jacoboni et al.2001). Both these programs were again artificial neural network based. The results were mapped out against the residue number allowing comparisons to be drawn (figure 3.1B, 3.1C, appendices A.2 and A.3 for full data). TMBETA predicted 13 transmembrane strands, and B2TMPRED 14 strands. Many of these were identified by both algorithms, however there were some differences between the two outcomes, for example a strand was predicted between residues 11-20 in TMBETA that was not identified by B2TMPRED. Again it was noticeable that the VS domains were almost entirely free from β-strands in these predictions.

3.3 Topology Prediction of *C. trachomatis* MOMP

The results of the above prediction algorithms were combined and adapted to give a model of the topology of *C. trachomatis* MOMP (figure 3.2). The two strand predicton programs between them identified 16 potential β-strands, however the strand inside the VS1 domain identified only in B2TMPRED was discarded as a possibility, as that region is considered highly likely to be extracellular. An extra strand was inserted between L203 and F212 to bring the chain back across the membrane so that all four VS domains

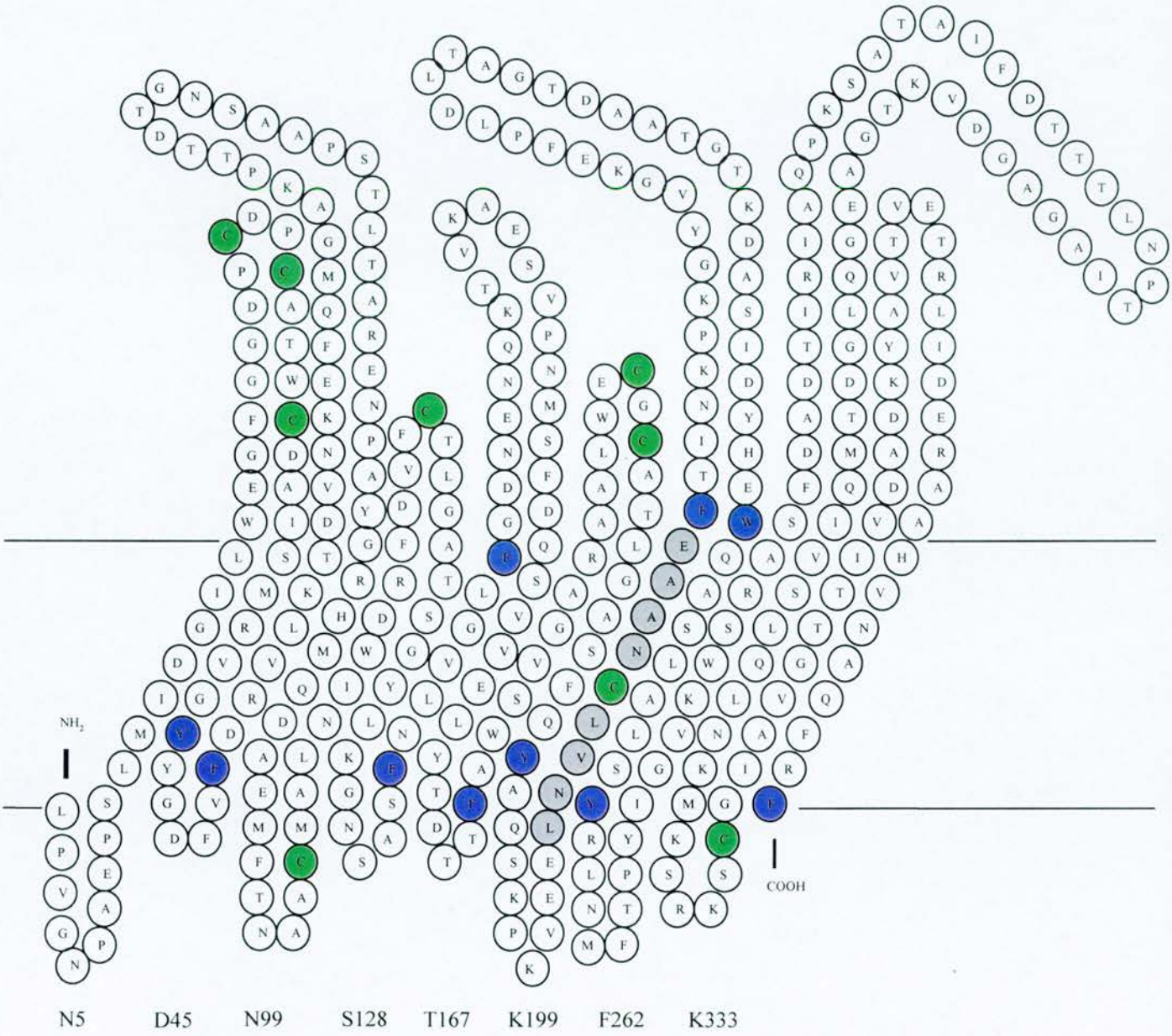


Figure 3.2: The predicted topology for *C. trachomatis* MOMP. Cysteine residues are highlighted in green. Outward facing "aromatic belt" residues are highlighted in blue. The manually determined transmembrane strand is shaded, and a selection of periplasmic residues are labelled.

were orientated to the extracellular side of the membrane. The strand was positioned to conform to standard observations of transmembrane β-strands, and so that the periplasic loop was shorter than the extracellular loop.

The final model corresponded well with general construction principles for β-barrels (Schulz.2000). MOMP is predicted to be a 16-stranded β-barrel, as has generally been suggested before. The N and C termini join in a strand at the periplasmic barrel end. The periplasmic turns are short and the extracellular loops long with high sequence variability. The outer barrel surface largely consists of aliphatic side chains, but towards the end of the strands there are frequently aromatic residues that form two "aromatic belts" that contact the nonpolar-polar interface of the bilayer. Twelve charged residues and numerous polar side chains project into the pore to form the hydrophilic passage through the bilayer. The external loops contain a high number of charged residues. Of the nine cysteine residues, six are on extracellular loops, two are on periplasmic turns and only one is found in a transmembrane strand. The periplasmic cysteines are on almost opposite sides of the barrel, and would therefore be unable to form an intra-molecular disulphide bond, though they could potentially form inter-molecular bonds. The transmembrane cysteine projects into the barrel pore, and could only be involved in bonding if loops were to fold into the barrel. Most cysteine residues are on extracellular loops, where they may interact with each other intra- and inter-molecularly and with other cysteine-rich surface exposed proteins in the chlamydial outer membrane.

3.4 Discussion

The combined topology prediction differed in several significant aspects from a topology prediction that had been made previously for *C. muridarum* MOMP (Rodriguez-Maranon et al.2002) based on Diederichs and hyrophobicity plots. Both concluded that MOMP was a 16 stranded β-barrel, however the prediction presented here assigned overall slightly longer strands, with an average transmembrane length of nine amino acids compared to eight. Consequently, the extracellular loops are not so long. In this study the VS domains have been assigned to loops 2, 4, 6 and 7, compared to loops 2, 3, 5 and 6 in *C. muridarum* MOMP. Critically, there were no periplasmic cysteine residues in the *C. muridarum* MOMP prediction. In the current model of chlamydial outer membrane structure, this would prevent important interactions with OmcB and OmcA, and could also impact subunit oligomerisation. Finally, there are no proline residues in strands in this *C. trachomatis* MOMP topology, which have been shown to be hardly tolerated in β-barrel transmembrane strands (Koebnik.1999). There are two strands containing proline in the *C. muridarum* MOMP prediction.

A hydrophobicity plot was not used as part of this topology predicton as the standard programs can have problems identifying transmembrane β-strands. While they can sometimes identify the classic amphipathic strands in porins with alternating hydrophilic and hydrophobic residues, other strands have different features. The constraint on the membrane-facing residues on the β-sheet to be hydrophobic is significantly stronger

than that of the pore-facing residues to be hydrophilic and strands with fewer polar residues can be misinterpreted. Strand 15 is such an example, with inward-facing residues of I, T, V and I. The artificial neural network algorithms, however, are trained on all known porin β-strands and both TMBETA and B2TMPRED strongly identified strand 15.

Secondary structure and topology models provide further evidence that MOMP is a 16-stranded β-barrel porin. The availability of such a model additionally allows more rational design of experiments to examine the structure and function of MOMP. Due to the high sequence similarity in other chlamydial MOMPs, especially outside the VS domains, it is expected that this prediction can also serve as a model for the other species.

Chapter 4: Preparation of a series of chlamydial porin constructs

4.1 Mature protein constructs

The first approach to the expression of recombinant MOMP began with expressing the protein as leaderless constructs that would be directed to inclusion bodies. Several *ompA* constructs from different chlamydiae species were made. In each case the gene was cloned without its leader sequence, to give mature MOMP as the translated product. A methionine codon was introduced at the 5′ end of the sequence to provide a new transcription initiation site. Genes were cloned into the multiple cloning site of the T7 expression vector pET22b(+) (Novagen), using the *Nde* I restriction site directly after the ribosome binding site (figure 4.1). The ATG start codon contained within the *Nde* I site is positioned at an optimal distance from the ribosome binding site for transcription initiation.

A mature *Ch. abortus ompA* construct in the pET22b(+) vector had been made previously (Hughes et al.2001) as had mature *Ch. pneumoniae ompA*. A mature *C. trachomatis ompA* (serovar D) construct was available in the pSTBlue-1 cloning vector (Novagen), but needed modification for use in the expression vector. To achieve this, an *Nde*-I restriction site was introduced at the 5′ end of the gene by site-directed mutagenesis (figure 4.2). The same method was used to remove an internal *Nde*-I site

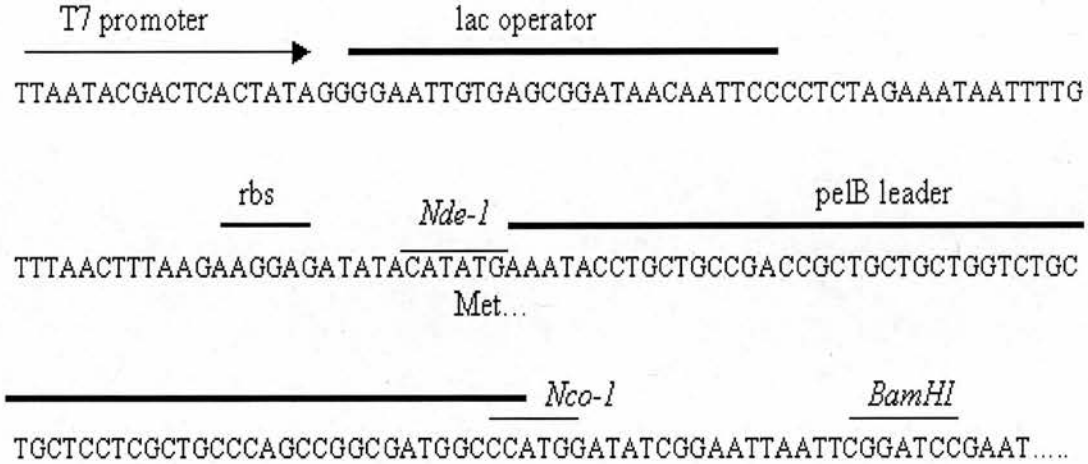


Figure 4.1: The multiple cloning site of expression vector pET22b(+). All constructs were cloned between either *Nde*-I and *Nco*-I, or *Nde*-I and *BamH* I. The ribosome binding site is also shown (rbs).

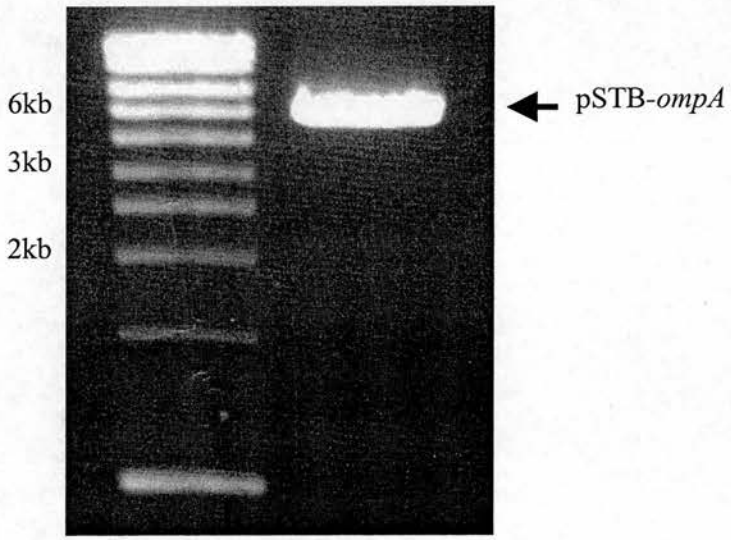


Figure 4.2: Site-directed mutagenesis of *C. trachomatis ompA*. Overlapping primers were used to introduce an *Nde-I* restriction site at the 5' end of the *ompA* gene in a pSTBlue-1 vector.

from the gene. A single thymine to cytosine change (nucleotide 270) resulted in a null mutation that destroyed the recognition sequence but ensured no change was made to the primary amino acid sequence. The mutagenised insert was then restricted with *Nde*-I and *Nco*-I to provide a 1.2kbp insert, that was subsequently ligated into the pET22b(+) vector.

A mature *C. muridarum ompA* clone was created directly from genomic DNA. Similarly to the *C. trachomatis* construct, *C. muridarum ompA* contained an internal *Nde*-I site that required removal before cloning. An equivalent null mutation was introduced during amplification from the genome using a two step PCR method (figure 4.3). In the first step, *ompA* was amplified in two parts, 5′ and 3′ of the mutagenesis site, each side using a cloning primer and a mutagenesis primer that contained the required nucleotide substitution. The products of these reactions were used in the second PCR. The two parts were self-primed for 8 cycles, resulting in a relatively small number of copies of the entire gene containing the mutation, before being spiked with the two cloning primers. The major product of this PCR was the full length *ompA* containing the mutation. The cloning primers were designed to include *Nde*-I and *Bam*H-I restriction sites, plus enough extra bases to allow subsequent digestion to create sticky ends (New England Biolabs Inc.2002). The complete *C. muridarum ompA* gene was then cloned directly into the pET22b(+) vector.

In addition to the *ompA* constructs, *porB* was also cloned. The gene, minus its leader sequence, was amplified from *C. muridarum* genomic DNA, giving a 1.1kbp product

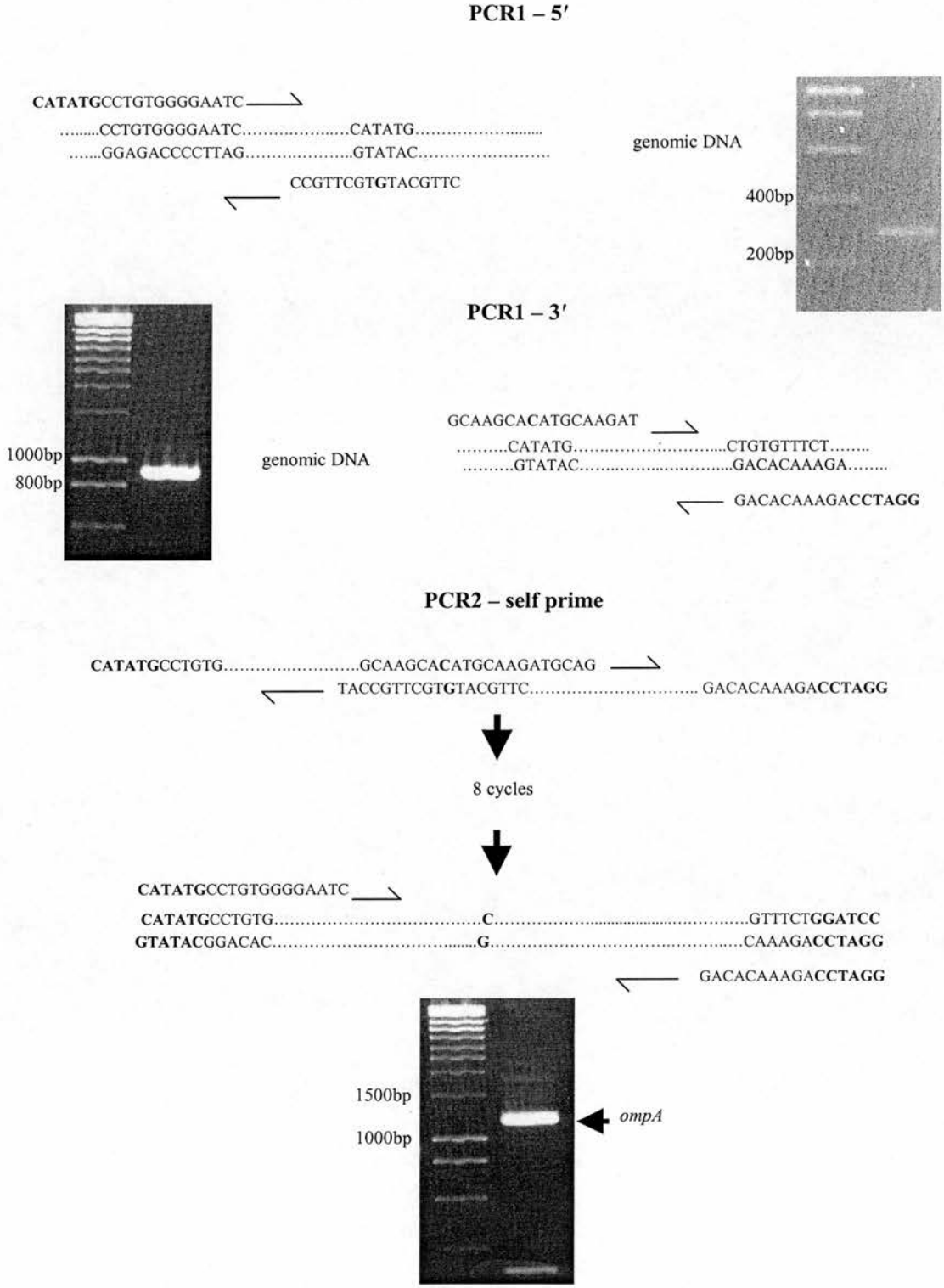


Figure 4.3: Schematic of *C. muridarum ompA* amplification. The gene was amplified by PCR in two sections, either side of an internal *Nde-I* restriction site (CATATG), while introducing a null mutation (PCR1). The products were used to amplify the entire *ompA* gene by self-priming then spiking the reaction with the cloning primers (PCR2).

that was digested to create specific ends and cloned directly into the *Nde*-I and *Nco*-I sites in the pET22b(+) vector (figure 4.4). All constructs were confirmed as correct by sequencing (MWG Biotech).

4.2 Cysteine mutagenesis

A cysteine-free mutant of *C. trachomatis ompA* (*M9ompA*) was made by site-directed mutagenesis. *C. trachomatis ompA* contains nine cysteine residues; C26, C29, C33, C102, C115, C182, C184, C207 and C335. Sequential rounds of mutagenesis were used to convert all nine to alanine residues in the pSTB-*ompA* construct before subcloning into the expression vector as before. Again the correct construct was confirmed by sequencing.

4.3 Full length constructs

The second approach used to express recombinant MOMP was designed to direct the protein to the outer membrane of *E. coli.* This required the inclusion of a signal sequence, so that the protein would be transported across the inner membrane of the cell. Signal sequences are cleaved by signal peptidase in the periplasm (Paetzel et al.2000), following which the protein is folded and inserted into the outer membrane. Full length *C. trachomatis ompA* constructs, including either the native chlamydial leader or an *E. coli* leader (from the *ompT* gene), were made by gene extension PCR using the mature *ompA* construct as a template (fig 4.5). Each primer in turn was used to amplify the

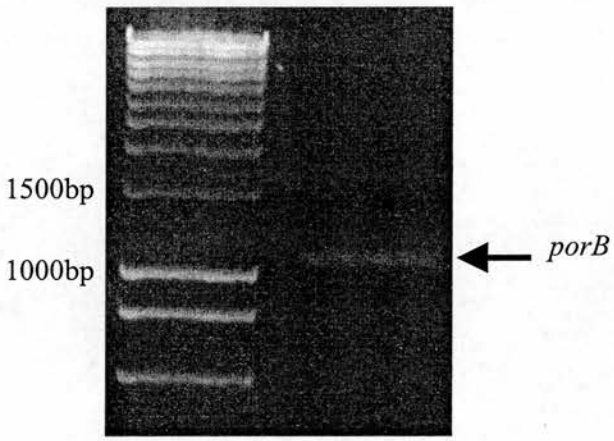


Figure 4.4: Amplification of *porB*. The *porB* gene was amplified directly from *C. muridarum* genomic DNA and cloned into pET22b(+).

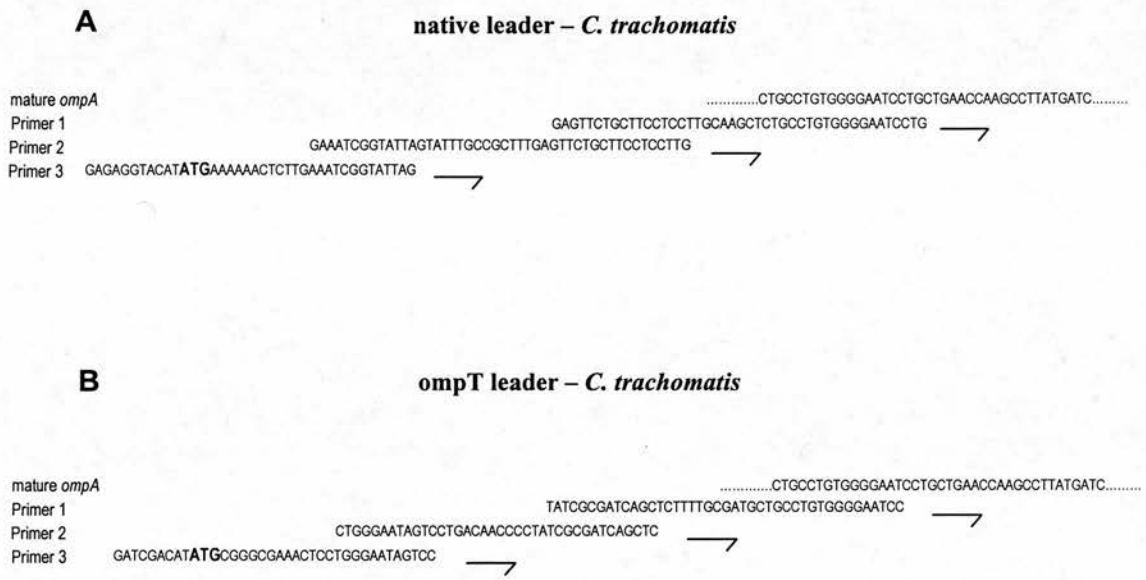


Figure 4.5: Schematic of full length *ompA* produced by gene extension. A. The native chlamydial leader sequence was added to mature *trachomatis ompA* by three sequential rounds of PCR, each extending the DNA further upstream. **B.** An *E. coli* leader sequence, from the *ompT* gene, was added to the mature *ompA* gene by the same method.

ompA gene with a common anti-sense primer containing an *Nco*-I restriction site. Three successive rounds of PCR resulted in the full length gene, including an *Nde*-I restriction site at the 5′ end. The final product was restricted and cloned into pET22b(+) as described previously.

The same method was used to create native leader and *ompT* leader clones of *Ch. abortus ompA* and the cysteine-free mutant *C. trachomatis M9ompA*, and an *ompT* version of full length *C. muridarum porB*, again using the mature constructs as templates. Native leader clones of *C. muridarum ompA* and *porB* were made by direct amplification from genomic DNA. This was carried out in exactly the same way as for the mature constructs but using a sense primer positioned to include the signal sequence in the amplified product.

4.4 Internal mutagenesis of *C. trachomatis* MOMP

A series of deletion mutant clones were made to investigate the influence of the VS domains on the structure and function of MOMP. Each of the four VS domains was removed from the full length *C. trachomatis ompA* gene using the native chlamydial leader. Site-directed mutagenesis was used to introduce an identical restriction site, in frame, at either end of a VS domain, allowing subsequent digestion of the domain from the gene, before re-ligating the ends back together. The regions of protein removed by this were: VS1, G63 to Y87; VS2, E141 to F156; VS3, Y220 to G238; VS4, D278 to T318 (see appendix C). The SDM employed to do this resulted in some slight

mutations. These were fairly conservative (M62T for VS1, T239V for VS3 and A277V for VS4) except for the mutation G219D in VS3. However, the topology prediction places this residue in an external loop, where the extra charge in unlikely to cause problems. Successful removal of each VS domain was confirmed by hemi-nested PCR (figure 4.6). The full length gene was amplified with both the cloning primers, to demonstrate successful subcloning, and with the 5′ sense mutagenesis primer and anti-sense cloning primer. As the 3′ half of the mutagenesis primer extends into the VS domain, successful amplification is only achieved if the domain is still present and digestion has failed, as in lane 1. The clones in lanes 2-5 are positive for the desired construct.

4.5 Discussion

A dual approach was taken to express recombinant MOMP for structural and functional studies. Several different constructs were made of the *ompA* gene and a second probable porin, *porB,* including some with internal mutagenesis. A variety of different molecular biological methods were used to produce these constructs. The DNA polymerase *Pfu* was used throughout for two reasons. Firstly, as a proofreading thermophilic polymerase *Pfu* has a low error rate (1.3×10^{-6} errors/nucleotide/cycle) compared to standard *Taq* DNA polymerase (8.0×10^{-6} errors/nucleotide/cycle) (Cline et al.1996). This reduced the likelihood of unwanted mutations being introduced into the DNA during amplification. All genes were sequenced and confirmed as correct, and it was

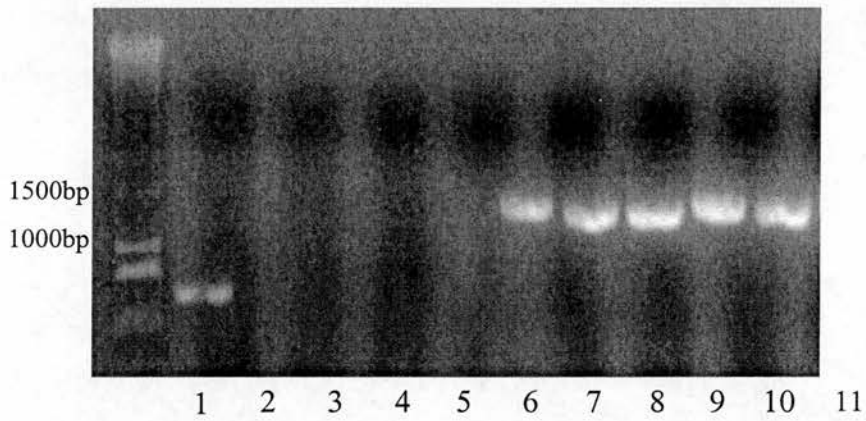


Figure 4.6: Hemi-nested PCR of *C. trachomatis ompA* VS2 deletion construct. 1-5: ΔVS2 clones amplified with VS2 sense and cloning anti-sense primers. 6-10: ΔVS2 clones amplified with cloning-sense and cloning-anti-sense primers.

rare for such a secondary mutation have occurred. As a further precaution, all mutagenised inserts were subcloned into fresh pET22b(+) vector to ensure against mutations that may have been introduced into other critical parts of the plasmid. Secondly, *Pfu* produces blunt ended DNA. This is necessary for the QuikChange site-directed mutagenesis protocol, so that extra nucleotides are not introduced in the middle of the gene at the end of each cycle of PCR. Primer design was also important, in particular for site-directed mutagenesis and gene extension. Mutagenesis primers were designed to change a maximum of three nucleotides in one round of PCR. Where more changes were required, for example to create a restriction site, a second pair of mutagenesis primers was designed to introduce these in a second round of PCR. The primers included 10-20 nucleotides either side of the mutation to ensure they annealed to the template despite the central mismatch. With the gene extension primers a balance had to be achieved between minimising the number of rounds of PCR required to add the leader sequence, to avoid introducing mutations, yet also avoiding primers of too great a length. Long primers are more likely to have internal secondary structure that can make annealing and therefore amplification difficult. Primers approximately 40 nucleotides in length were used, with around 20 nucleotides overlapping with the template and a further 20 nucleotides being added per round of PCR.

Access to a large range of *ompA* clones (table 4.1) has several advantages. The two basic approaches used to express the proteins, combined with the variety of species the gene was cloned from, serve to increase the chance of producing correctly structured protein. Additionally, where comparisons are possible between these proteins,

Species	Mature clones	Leadered clones	Mutagenesis
C. trachomatis	*ompA*	*oT-ompA, nL-ompA*	*M9ompA, oT-M9ompA, nL-M9ompA, nL-ompA-ΔVS1, nL-ompA-ΔVS2, nL-ompA-ΔVS3, nL-ompA-ΔVS4*
Ch. abortus	*ompA*	*oT-ompA, nL-ompA*	-
C. muridarum	*ompA, porB*	*nL-ompA, oT-porB, nL-porB*	-
Ch. pneumoniae	*ompA*	-	-

Table 4.1: Complete series of chlamydial protein constructs: *ompA* is the gene that encodes MOMP, *porB* encodes PorB, *oT* is the *E. coli* signal sequence from the *ompT* gene, *nL* is the native chlamydial leader, *M9ompA* is the cysteine-less mutant of *C. trachomatis* and *ΔVS1-4* are the domain deletion constructs.

information can be obtained relative to structure, function, folding and assembly of MOMP. The cysteine-free mutant and the internal mutagenesis relating to the VS domains of *C. trachomatis* MOMP allow investigation into the roles of these potentially crucial aspects of the protein. Recombinant *C. muridarum* MOMP would be useful for its role in the animal model for *C. trachomatis*. Finally, the *porB* construct provides an alternative target for study where the smaller predicted size, reduced variability of sequence and potential selectivity are of interest (Kubo et al.2000; Kubo et al.2001).

Chapter 5: Expression of recombinant chlamydial porins in *E. coli*

5.1 Mature protein constructs

5.1.1 Porin expression in BL21 cells

The pET22b(+) vector constructs were transformed into BL21(DE3) cells, a standard expression strain of *E. coli* with the genotype *E. coli* B F^- *dcm ompT hsdS*($r_B^- m_B^-$) *gal* λ(DE3). The DE3 lysogen consists of the T7 RNA polymerase gene under the control of a *lac* promoter. On the addition of IPTG the polymerase is expressed, which in turn transcribes the porin gene. The pET22b(+) vector also contains a copy of the *lac* repressor, which inhibits basal protein expression in the absence of IPTG.

In initial expression tests of *Ch. abortus* MOMP little protein was detected. With ampicillin resistant vectors loss of plasmid can reduce the efficiency of expression. This is possible because the amp^R product, β-lactamase, is secreted into the culture media where it can break down all the ampicillin without requiring the antibiotic to enter the cell. As the β-lactamase builds up in the media, bacteria are able to eject the plasmid without resulting cell death. The plasmid-free bacteria are not producing the over-expressed protein and so have a growth advantage over the remaining expressing bacteria and overtake the culture. Therefore a plasmid stability test was conducted. The impact on the expression of MOMP was determined by sampling the inducing bacterial

	percentage amp^R colonies
2 hours after induction	76 ± 8
3 hours after induction	17 ± 11
3 hours after induction with media change	92 ± 6

Table 5.1: Plasmid stability test of *Ch. abortus* mature MOMP construct. Percentage of colonies on an ampicillin plate compared to a non-ampicillin plate (±SD, n=3), relative to an initial amp^R of 100%.

culture at different time points and testing for presence of plasmid (table 5.1). An equal volume of each sample was spread on agar plates with and without antibiotic, and grown at 37°C overnight. A comparison of the number of colonies on the different plates shows the proportion of cells in the original sample that still retained the plasmid. Plasmid instability was found to be a major determinant of MOMP expression, as only 17% of the bacteria retained ampicillin resistance after three hours of protein expression. To overcome this, a complete media change was carried out between the bacterial growth and protein induction. This resulted in an increase of plasmid retention to 92%. The media change was included in the protocol of all future expressions.

C. trachomatis MOMP began to express within the first thirty minutes of the addition of IPTG and continued to accumulate throughout the three hour induction, giving a broad band at 40kDa (figure 5.1). The band was confirmed as MOMP by western blotting.

5.1.2 Inclusion body (IB) preparation and expression in Δomp8 cells

The recombinant protein comprised a large proportion of the total protein in the cell lysate. The IBs were thoroughly washed with buffer containing both 1% (v/v) Triton X-100 and 2M urea (figure 5.2A). This reduced the amount of contaminating proteins but some remained, most notably a tentative OmpF band. Endogenous *E. coli* porins in the sample are difficult to remove completely, and could interfere significantly with later experiments. To avoid this problem, pET-*ompA* was transformed into Δomp8 cells, a

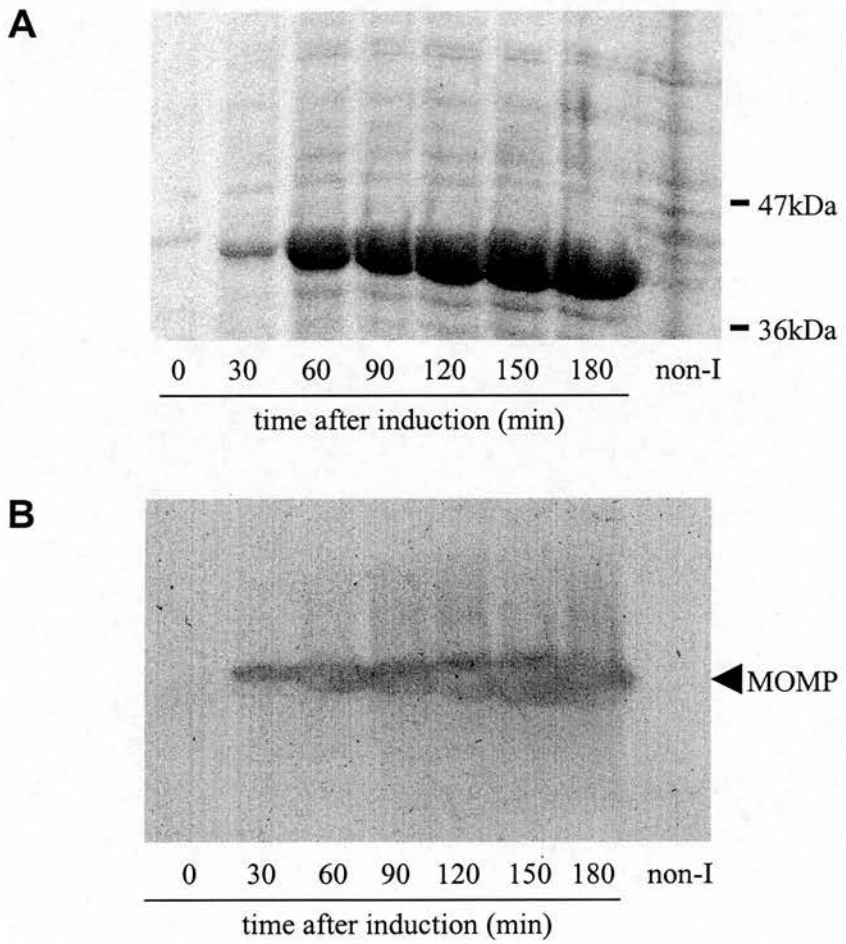


Figure 5.1: Expression time course of *C. trachomatis* MOMP. A. 12% (w/v) SDS-PAGE stained with coomassie blue. **B.** Western blot. All samples are protein from 100µl cells, lysed and boiled for 5min.

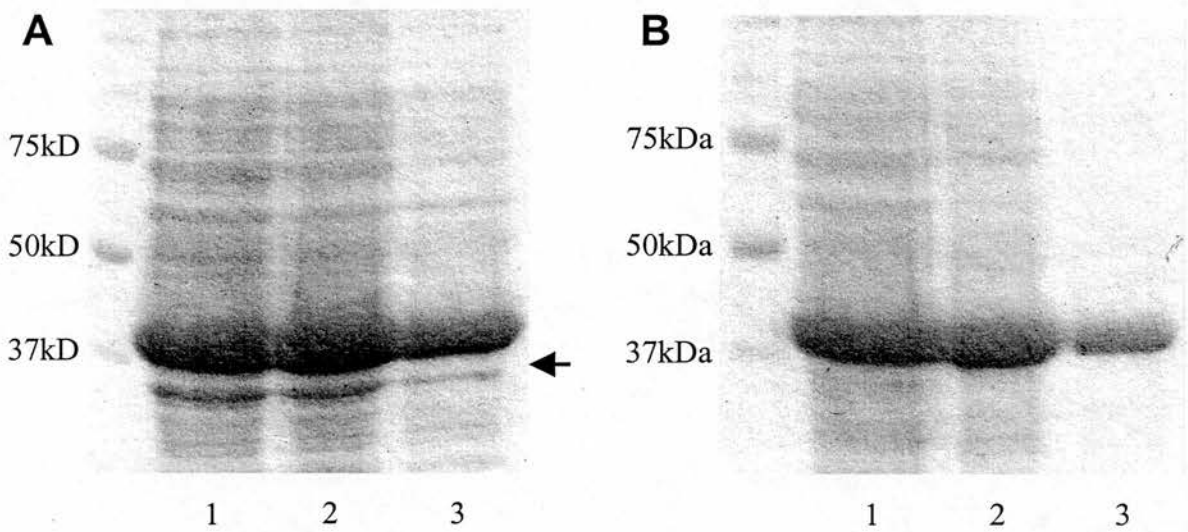


Figure 5.2: Inclusion body preparation of *C. muridarum* MOMP. A. MOMP expressed in standard BL21(DE3) bacteria. **B.** MOMP expressed in Δomp8 cells. Lane 1 – cell lysate, lane 2 – crude inclusion bodies, lane 3 – washed inclusion bodies. 12% (w/v) SDS-PAGE, stained with coomassie blue. All samples were boiled for 5mins. Arrow indicates tentative OmpF band.

porin-deficient strain of BL21s (Prilipov et al.1998). Δomp8 cells are derived from standard BL21(DE3) *E. coli* that have been mutated to give a genotype of BL21(DE3) *ΔlamB ompF::*Tn*5 ΔompC ΔompA*, with the major porin genes individually knocked out. MOMP was expressed successfully in these cells, though with a slightly reduced yield. However the washed inclusion body sample was much purer and free from contaminant native porin (figure 5.2B).

Protein was recovered from the inclusion bodies by solubilisation in a chaotropic agent with an excess of DTT (figure 5.3). Urea at concentrations up to 8M was relatively inefficient, and preferentially solubilised remaining contaminants. In contrast 6M guanidine hydrochloride was more effective and fully solubilised the recombinant protein.

5.2 Full length constructs (with leader sequences)

5.2.1 Growth of *E. coli* expressing leadered MOMP

Full length pET-*ompA* constructs were also transformed into BL21 cells and expressed at 37°C. With this expression method, not only must the cells produce the MOMP but it must also be transported across the inner membrane, the signal sequence digested and then the resulting mature protein inserted into the outer membrane. This is likely to put

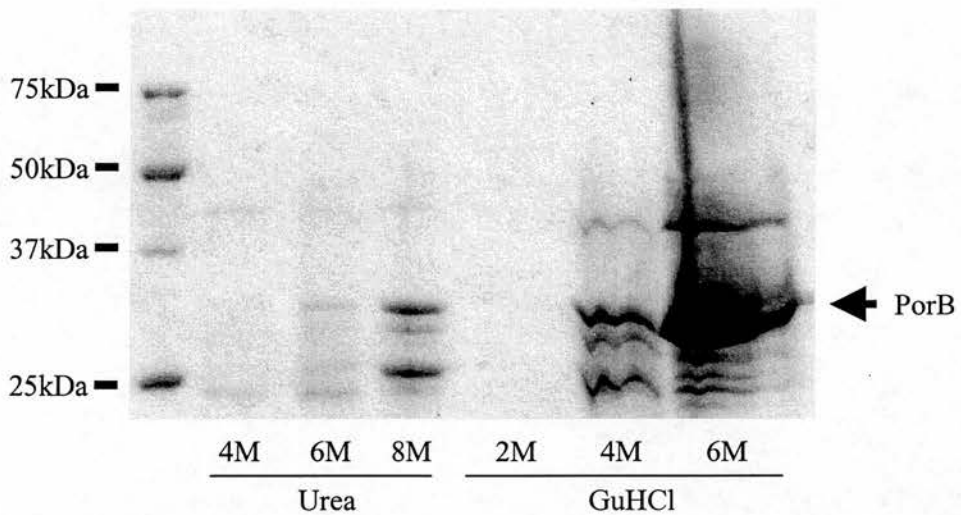


Figure 5.3: Solubilisation of PorB. Inclusion bodies of PorB, expressed in Δomp8s, solubilised in increasing concentrations of two chaotropic agents. 10% (w/v) SDS-PAGE, stained with Coomassie blue.

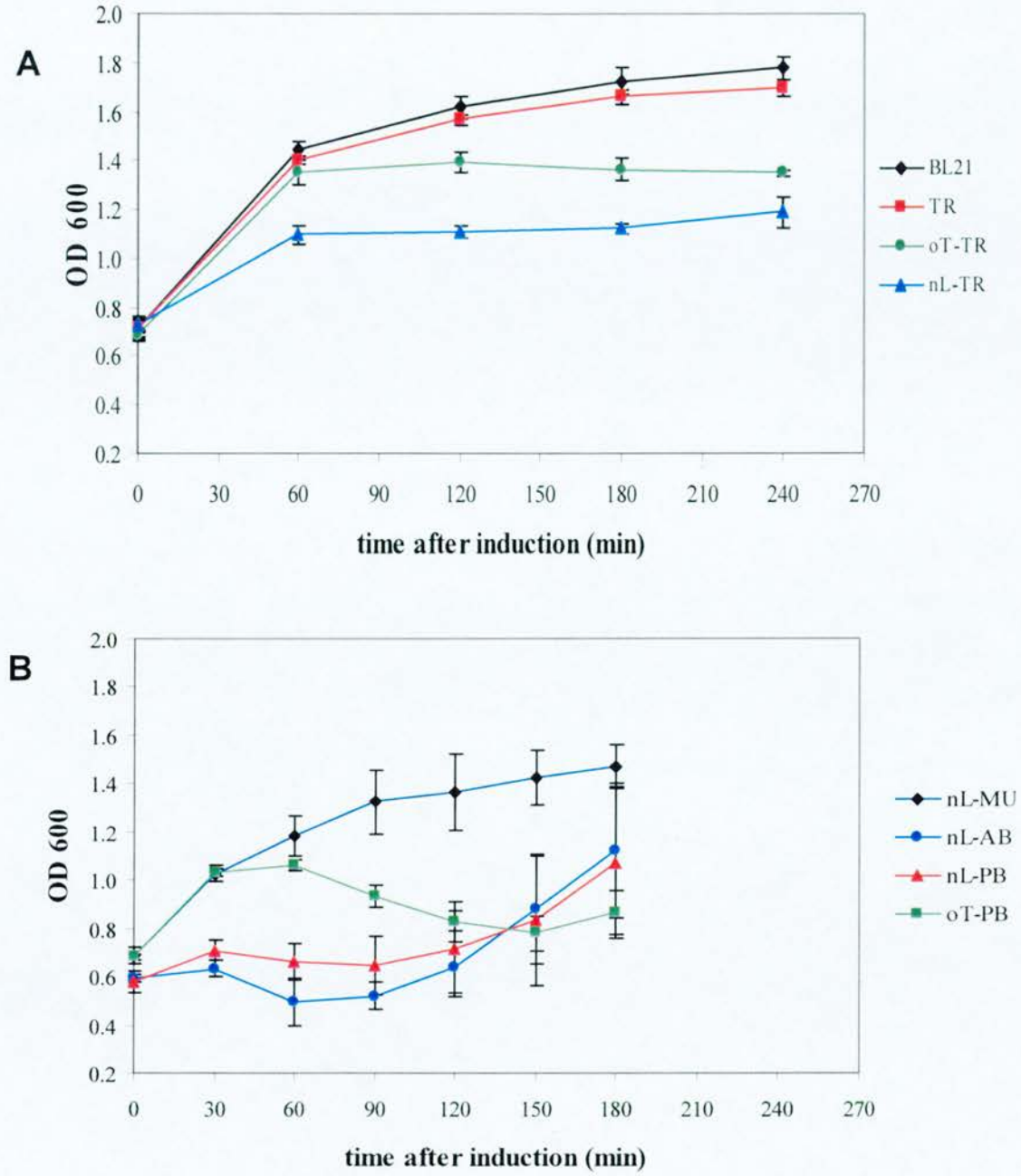


Figure 5.4: Growth curves of BL21s expressing full length porins at 37°C. A. Comparison of *C. trachomatis* MOMP expressed with native (nL-TR) and *E. coli* (oT-TR) leaders and with mature, leaderless MOMP (TR). **B.** Growth curve of *E. coli* during the expression of *C. muridarum* and *Ch. abortus* MOMP and PorB with a native leader (nL-MU, nL-AB and nL-PB respectively) and PorB with an *E. coli* leader (oT-PB). (all values are shown as ±SD, n=4).

a greater strain on the cell than that of mature MOMP expression and so compromise viability or growth. Therefore a time course of optical densities throughout induction were measured, allowing the growth curves of the cultures to be plotted (figure 5.4). Expression of mature, leaderless *C. trachomatis* MOMP did not measurably inhibit the growth of the bacteria compared with the cells alone. There is a much greater effect with expression of full length MOMP. In ompT-MOMP expression the initial growth rate was comparable to that of the cell only and mature MOMP expression, however the culture saturated at a lower density. This suggests that whilst expressing the protein presented no difficulty to the bacteria, the build up of MOMP within the cell increased the likelihood of cell lysis. During native leadered-MOMP expression the initial growth rate was inhibited and the saturation density was even lower. Full length constructs of the other chlamydial genes were variable in their viability (figure 5.4B). Bacteria expressing *C. muridarum* MOMP had a slower initial growth rate than those expressing *C. trachomatis* MOMP, however the early, low density saturation was not seen and the culture continued to slowly grow throughout the induction. In contrast, bacteria had such difficulties with the *Ch. abortus* MOMP and *C. muridarum* PorB that there was barely any initial growth, and the culture density soon began to drop as the cells lysed. At the end of the induction, no recombinant protein was detected. Selection pressure was so strong that after only 60-90mins plasmid-free cells began rapidly to take over the culture, seen in the rapid increase in density in the latter half of the induction. PorB expression was improved by replacing the native chlamydial leader with the *ompT* leader. Initial growth rate was recovered and the decrease in optical density occurred later in the induction and continued more slowly and for longer before the antibiotic

selectivity was overcome. No significant difference was seen with *Ch. abortus* MOMP when expressed with the *ompT* leader (not shown). Together, these results indicated that leadered MOMP expression can slow the rate of cell division, particularly in those constructs with the native leader sequence. For *C. trachomatis* MOMP and PorB this was improved by substituting the native leader with the ompT leader. For most constructs a second problem, of cell death as the recombinant protein increased in the cell, resulted in a suppressed density of culture saturation.

5.2.2 Optimisation of expression in Δomp8 cells

It was desirable to express the full length constructs in Δomp8 cells, as contaminant *E. coli* porins would be particularly likely to interfere with subsequent work in this expression system. However, problems with cell death were even more pronounced in this cell strain than in standard BL21 cells. The expression conditions were therefore optimised to give the best possible yield of protein. Inductions were carried out on both native and ompT leadered *C. trachomatis* MOMP, rapidly at 37°C with 1mM IPTG and slowly at 16°C with 0.1mM IPTG (figure 5.5A). Both versions of MOMP were expressed strongly at 37°C with about half of the protein being processed, as evidenced by the doublet band of MOMP with and without the signal sequence (~2kDa difference). When native-MOMP was expressed at 16°C there was a slight decrease in total protein but the proportion of protein that was processed was unchanged. There was a larger decrease in protein production with slow induction of ompT-MOMP, but in this case all

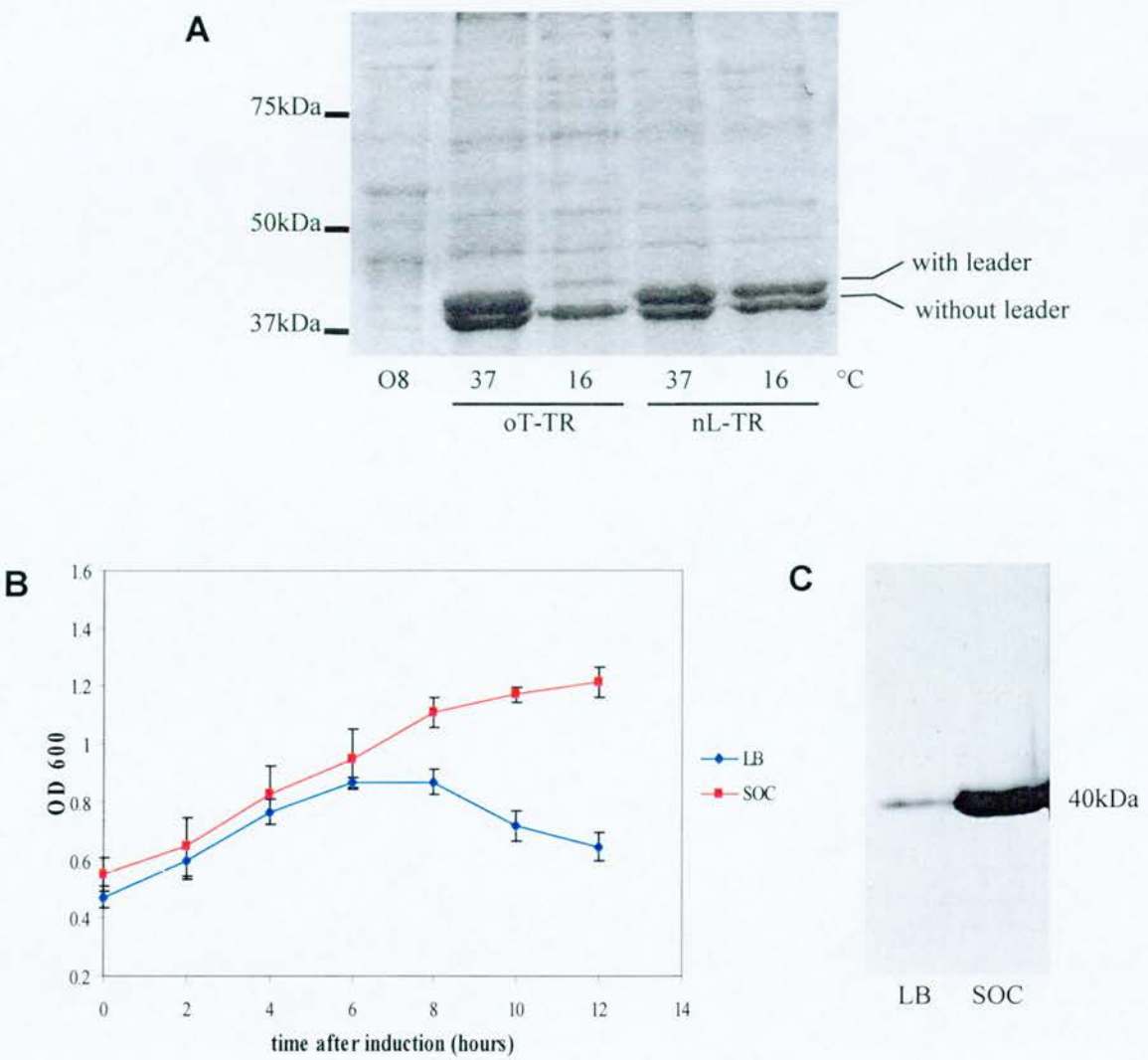


Figure 5.5: Optimisation of expression of full length *C. trachomatis* MOMP. A. Outer membranes from 0.5ml of Δomp8 cells (O8) with native and *ompT* MOMP expressed at 37°C and 16°C in LB medium. Upper arrow indicates full length MOMP, lower arrow indicates processed, mature MOMP. 12% (w/v) SDS-PAGE, stained with Coomassie blue. **B.** Growth curve of oT-TR expressed in LB and SOC media at 16°C in Δomp8 cells. **C.** oT-TR MOMP expressed in Δomp8 cells grown in LB and SOC, and extracted from the outer membrane in 1% (w/v) OG. Protein from 0.75ml cells. Western blot.

of it was processed to the mature form. Induction was also carried out in different growth media to try to further improve protein yield. "Slow" expression of ompT-MOMP was carried out in both LB and SOC media (figure 5.5B). The cultures grew at a similar rate for the first 6 hours of induction, at which point the LB culture levelled off at OD_{600} ~0.85, after which the cells began to lyse and the density to fall. In contrast, the SOC culture continued to grow and remained unsaturated twelve hours after induction. As expected, the yield of MOMP was increased in the SOC culture (figure 5.5C).

5.3 Discussion

An important consideration when expressing chlamydial porin constructs in *E. coli* has been to overcome problems associated with the toxicity of the recombinant protein to the bacteria. The pET vectors are T7 *lac* based and pET22b(+) has in addition an extra layer of restriction on basal (non-induced) protein production by having the *lac* repressor encoded on the plasmid. This is a very tightly controlled system, yet plasmid instability was still a problem. However, it was found that the incorporation of a complete media change into the protocol was largely able to prevent this. Using a kanamycin resistant instead of an ampicillin resistant plasmid at the outset would have provided a surer way of preventing plasmid loss, however the Δomp8 cells have chromosomal kanamycin resistance, so unfortunately that was not possible.

The expression of full length protein was particularly problematic, as the same toxicity and plasmid instability issues were compounded by the problem of cell lysis as the amount of processed protein increased. The level of lysis varied greatly between constructs. Expression of *Ch. abortus* MOMP and *C. muridarum* PorB were rapidly fatal to the bacteria, as indicated by the finding that antibiotic selectivity was overcome after only ~1 hour of induction. The effect was such that no recombinant protein was detectable at the end of the 3 hour expression (not shown). *C. trachomatis* MOMP was relatively non-toxic to the cells, and in bacteria expressing *C. muridarum* MOMP cell lysis did not overtake growth at any point in the induction. In some cases substituting the chlamydial leader sequence for a leader from an *E. coli* outer membrane protein - OmpT - reduced cell lysis, and improved cell survival is the likely cause of the larger yield of recombinant protein seen in the oT-TR versus nL-TR outer membrane preparations (figure 5.5A).

It is probable that as a result of all the absent porins Δomp8 cells have a much less stable outer membrane than unmodified BL21 cells. This is reflected in a generally higher tendency towards lysis. It is perhaps unsurprising that the overproduction of a protein that can effectively create holes in the outer membrane can be fatal to the cell. For this reason SOC medium was explored as an alternative growth medium. SOC medium is typically used in the transformation of bacteria with DNA. The exact mechanism of transformation is poorly understood, but it does involve permeablisation of the membrane to allow entry of the DNA. The conditions in SOC medium are such that this

is less likely to result in cell death. It was considered that the expression of recombinant porin in Δomp8 cells might be analogous to the situation involved in transformation. This seems to have been borne out in the dramatic improvement in cell survival when grown in SOC compared to LB, without any difference in the initial growth rate between the two cultures.

Recombinant MOMP and PorB have been successfully expressed by both approaches. All the constructs expressed well in the mature form to inclusion bodies, with yields of 80-150mg protein per litre culture. Protein in the cleaned and solubilised inclusion bodies had a high level of purity and was suitable for future *in vitro* refolding experiments. There was, however, more variability with the full length MOMP approach to expression, and it was not possible to express *Ch. abortus* in a form that could be processed to the outer membrane at all. Of the proteins tested, *C. trachomatis* and *C. muridarum* MOMP showed best expression, using ompT and native leaders respectively. *C. trachomatis* ompT-MOMP was expressed optimally in Δomp8 cells in SOC medium containing 0.1mM IPTG at 16°C. As problems of the translocation and processing of large quantities of recombinant protein, and of Δomp8 cell lysis apply to all constructs, this optimisation is likely to generally improve expression. At this stage the fully processed proteins can be extracted from the outer membrane using detergents, for further analysis and purification.

Chapter 6: Characterisation of outer membrane expressed MOMP

6.1 Surface expression of MOMP

MOMP was successfully expressed in BL21 and Δomp8 cells with leader sequences and translocated to the periplasm. The signal sequence was digested to give the mature form, which was isolated as part of the membrane fraction of the cell. However, MOMP could be merely associated with the membrane peripherally, rather than fully inserted into the outer membrane as an integral membrane protein. To examine whether the protein had truly inserted into the outer membrane, whole cell immunoblots were used to test for the presence of MOMP on the surface of expressing cells.

BL21 cells and BL21s transformed with mature, leaderless *C. trachomatis* MOMP (TR), ompT-MOMP (oT-TR) and native leadered-MOMP (nL-TR) were induced for 2 hours at 25°C. Intact cells were harvested by a short and slow centrifugation, and resuspended in PBS pH 7.4. A sample of each was spotted and dried onto nitrocellulose membrane. Nitrocellulose was used as no methanol was needed at any stage. Methanol could have permeablised the outer membrane and distorted the results. MOMP was detected using the anti-*C. trachomatis* MOMP polyclonal antibody. No non-specific binding was picked up on the BL21 cell only control (figure 6.1). Additionally there was no binding to MOMP expressed in the mature form, indicating the drying and blotting protocol did not cause cell lysis. Both the ompT and native leadered MOMP were detected, ompT-

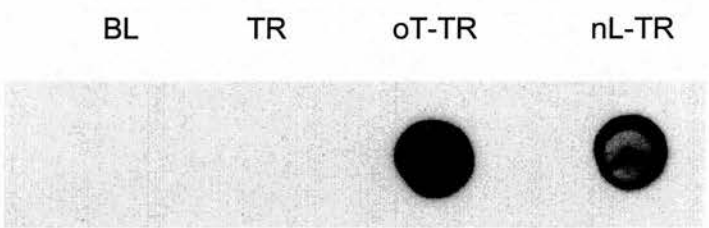


Figure 6.1: Whole cell immunoblot of leadered *C. trachomatis* MOMP. BL21 cells, and BL21s expressing mature *C. trachomatis* MOMP and ompT and native leadered MOMP dried on to nitrocellulose membrane. Surface-exposed MOMP detected by blotting with anti-*C. trachomatis* MOMP polyclonal antibody.

MOMP in greater amounts. This demonstrated that in both cases the protein was successfully inserted into the outer membrane. It is unlikely that this could have happened without some folding of the protein. Interestingly, with equivalent immunoblots for MOMP expressed at 37°C, no protein was detected on the cell surface, although translocated and processed protein was found in the membrane fraction. Unfortunately, it proved impossible to repeat these experiments in Δomp8 cells, as these more fragile bacteria lysed during the protocol so that results could not be conclusive.

The effect of internal mutagenesis of *C. trachomatis* MOMP on membrane insertion was examined (figure 6.2). Cysteine-less MOMP, M9-MOMP, was unaffected by the mutagenesis and was detected on the cell surface of bacteria expressing both the ompT and native leadered versions. This suggests that disulphide bonds are not essential for folding and membrane insertion. The importance of the VS domains was also examined. Four native leadered *C. trachomatis* MOMPs, each lacking one of the VS domains, were expressed and the cells blotted. In each case MOMP was detected on the cell surface, indicating the VS domains are extracellular loops as predicted, rather than transmembrane β-strands and are also, at least individually, not essential for folding and insertion into the outer membrane. The variable intensity of signal from the mutant constructs could be as a result of differences in the rate of membrane insertion, differences between populations of the polyclonal antibody that react to the missing regions, or some combination of both.

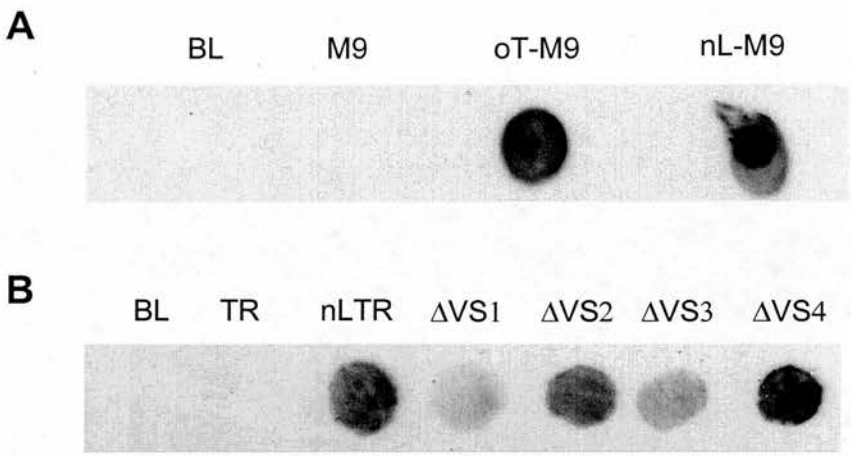


Figure 6.2: Internal mutagenesis of *C. trachomatis* MOMP. A. Whole cell immunoblots of the cysteine-less (M9-MOMP) constructs; cell only, mature M9-MOMP, ompT leadered M9-MOMP and native leadered M9-MOMP. Surface-exposed MOMP detected by blotting with anti-*C. trachomatis* MOMP polyclonal antibody. **B.** Whole cell immunoblots of VS domain deletion constructs, all expressed with the native leader sequence. Surface-exposed MOMP detected by blotting with anti-*C. trachomatis* MOMP polyclonal antibody.

6.2 Oligomerisation in the outer membrane

Chlamydial MOMPs are predicted to be trimeric porins, as MOMP purified from the chlamydial membrane ran as SDS-resistant putative trimers on electrophoresis (Wyllie et al.1998). However, in the chlamydial membrane MOMP is part of an extensive network of cross-linked proteins, which combined with the difficulties in purifying large quantities of MOMP has made it difficult to examine its intra- and inter-subunit interactions in isolation from other cysteine-rich chlamydial outer membrane proteins. The oligomeric state of MOMP expressed in *E. coli* membranes was therefore investigated.

Outer membranes were prepared from Δomp8 cells expressing *C. trachomatis* MOMP, grown in SOC medium and induced at 16°C. Samples were incubated in SDS loading buffer with or without reducing agent at room temperature for 10mins. Proteins were separated by SDS-PAGE and detected by immunoblotting (figure 6.3). Fully reduced MOMP ran as a broad single band at 40kDa but oxidised MOMP, incubated without reducing agent, ran as a smear. Several stronger, distinct bands were, however, detected within the smear. Denatured, monomeric MOMP runs as a band at 40kDa (DM) as in the reduced sample. Another strong band at ~35kDa is also likely to be monomeric MOMP, but folded and running ahead of its molecular weight as has been seen with other porins (Jansen et al.2000; Conlan et al.2003). Additional bands are seen of

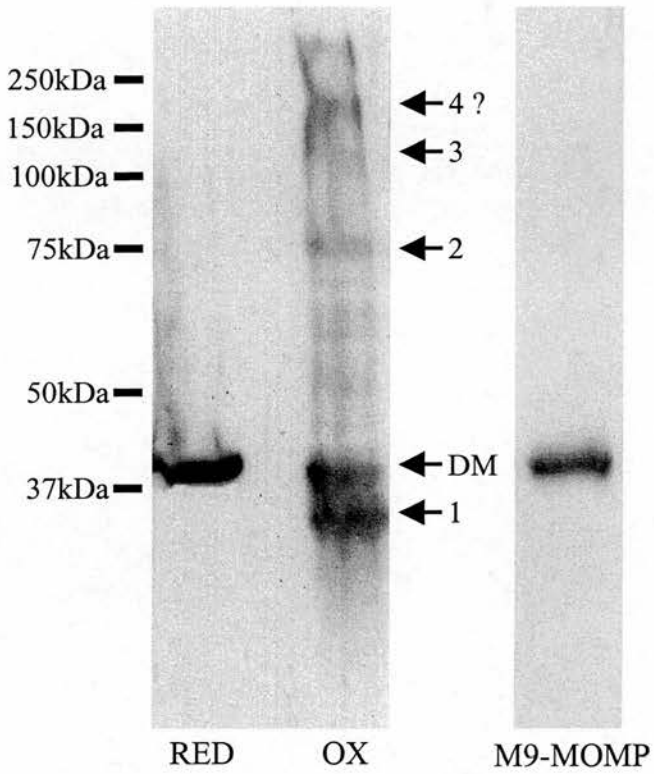


Figure 6.3: The role of disulphide bonds in MOMP oligomerisation. *C. trachomatis* MOMP outer membrane protein separated by SDS-PAGE with (RED) and without (OX) reducing agent in the sample, and cysteine-free M9-MOMP without reducing agent. Western blot from the gel. Predicted denatured monomer (DM), monomer (1), dimer (2) and trimer (3) bands, and putative tetramer band.

higher molecular weight that would correspond to dimers and trimers (~80kDa, 120kDa). Above this size the molecular weights of bands are difficult to determine accurately, however, the band below the 250kDa marker could plausibly be a tetramer with other higher order complexes seen at the top of the gel.

The presence of a folded monomer band in the oxidised but not the reduced sample indicated that at least one intra-molecular disulphide bond was involved in the structure of MOMP. The ability of the cysteine-less mutant to be expressed and inserted into the outer membrane suggests that this disulphide bond is likely to be involved in the stabilisation of the tertiary structure, so that the protein is SDS-resistant, rather than essential for the folding itself. MOMP was also present in the outer membrane in a variety of oligomeric forms, with two, three or more monomers associated together. It is unclear whether the subunits are connected by disulphide bonds, as the destabilisation of the monomer subunits by reduction could result in subunit dissociation even if cysteine residues were not involved.

6.3 Size exclusion chromatography

Protein from the outer membrane was solubilised in 1% (w/v) Zwittergent 3,14 and 10mM DTT, then separated by size exclusion chromatography in 0.05% (w/v) Zwittergent 3,14 and 5mM DTT (figure 6.4A). 10µl of each 5ml fraction was dotted on

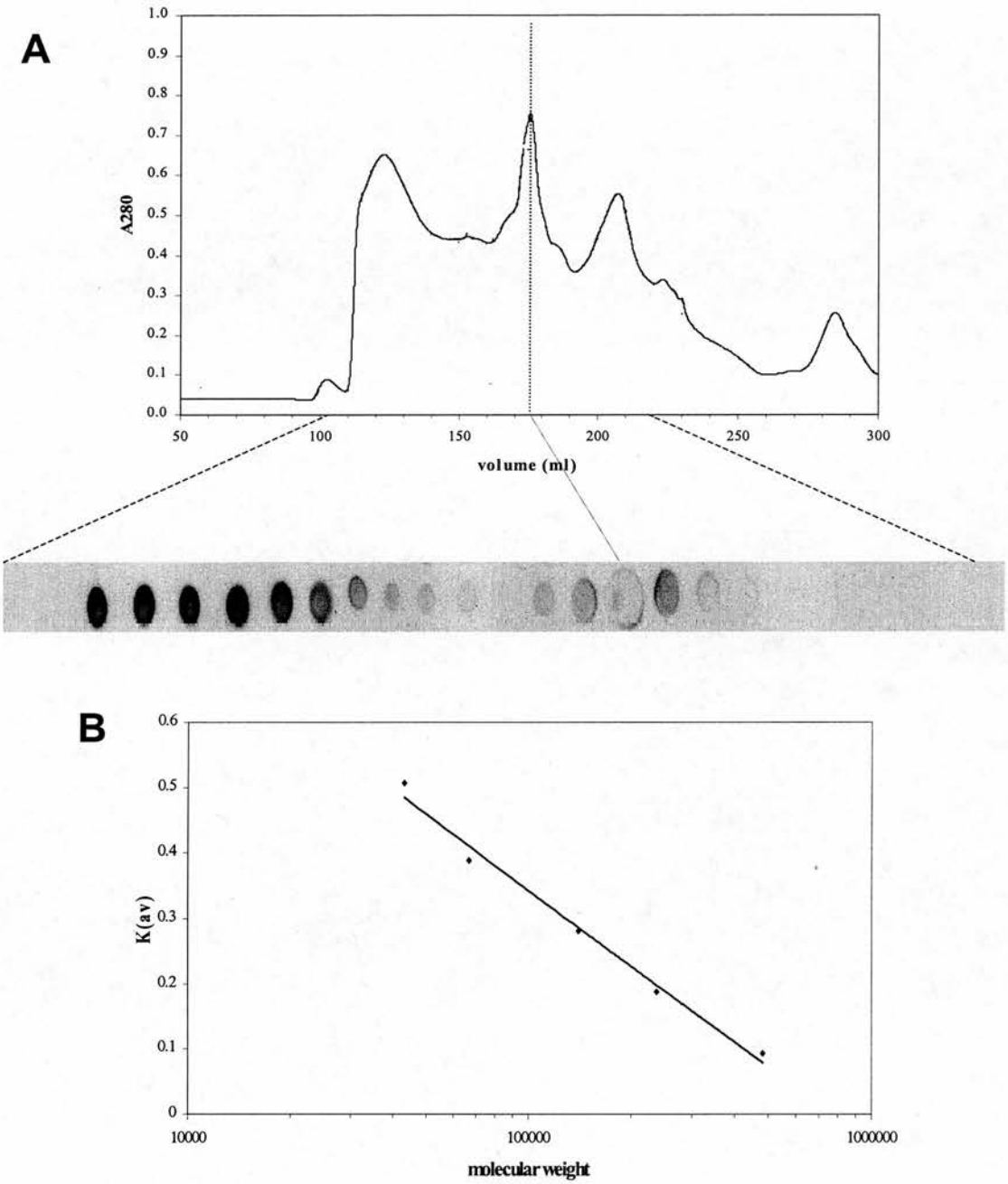


Figure 6.4: Size exclusion chromatography of *C. trachomatis* MOMP. A. Separation of zwittergent 3,14-solubilised MOMP, with dot blots of collected 5ml fractions. **B.** Calibration of column in the presence of zwittergent 3,14. Standards were apoferritin, pyruvate kinase, lactate dehydrogenase, bovine serum albumin, ovalbumin.

PVDF membrane and fractions containing MOMP were detected by immunoblotting. MOMP eluted in two peaks. A large proportion of the protein eluted in a broad peak with the void volume, with a slow tail off over the next 50ml. A significant fraction of the protein eluted in a second peak, with its maximum at a volume of 177ml. The column had been calibrated in the same buffering conditions (figure 6.4B), including detergent, and the peak was calculated to correspond to 125kDa. This indicated a portion of the MOMP in the sample was in trimeric form.

The experiment was repeated using protein solubilised in 1% (w/v) LDAO and DTT, then separated by chromatography as before but with 0.05% (w/v) LDAO as the detergent in the buffer (figure 6.5A). MOMP again eluted in two sections. A much larger proportion of MOMP eluted with the void volume than with Zwittergent 3,14-solubilised MOMP. The correspondingly small amount of MOMP that eluted later was not enough to create a visible peak on the trace against background protein, however the dot blots showed a peak with a maximum at 193ml or 78kDa. Although most of the MOMP in this LDAO sample existed as very large complexes or aggregates, some was eluting as a dimer. *C. trachomatis* M9-MOMP was also expressed and extracted in Zwittergent 3,14. On separation by size exclusion chromatography, M9-MOMP eluted more evenly than wild type MOMP (figure 6.5B). The peak at the void volume was lower and broader, and M9-MOMP continued to be detected in each fraction over the next 80ml. However, a second maximum could be seen both in the A_{280} trace and in the dot blots at 168ml or 163kDa, corresponding to a tetramer of M9-MOMP. It seems clear

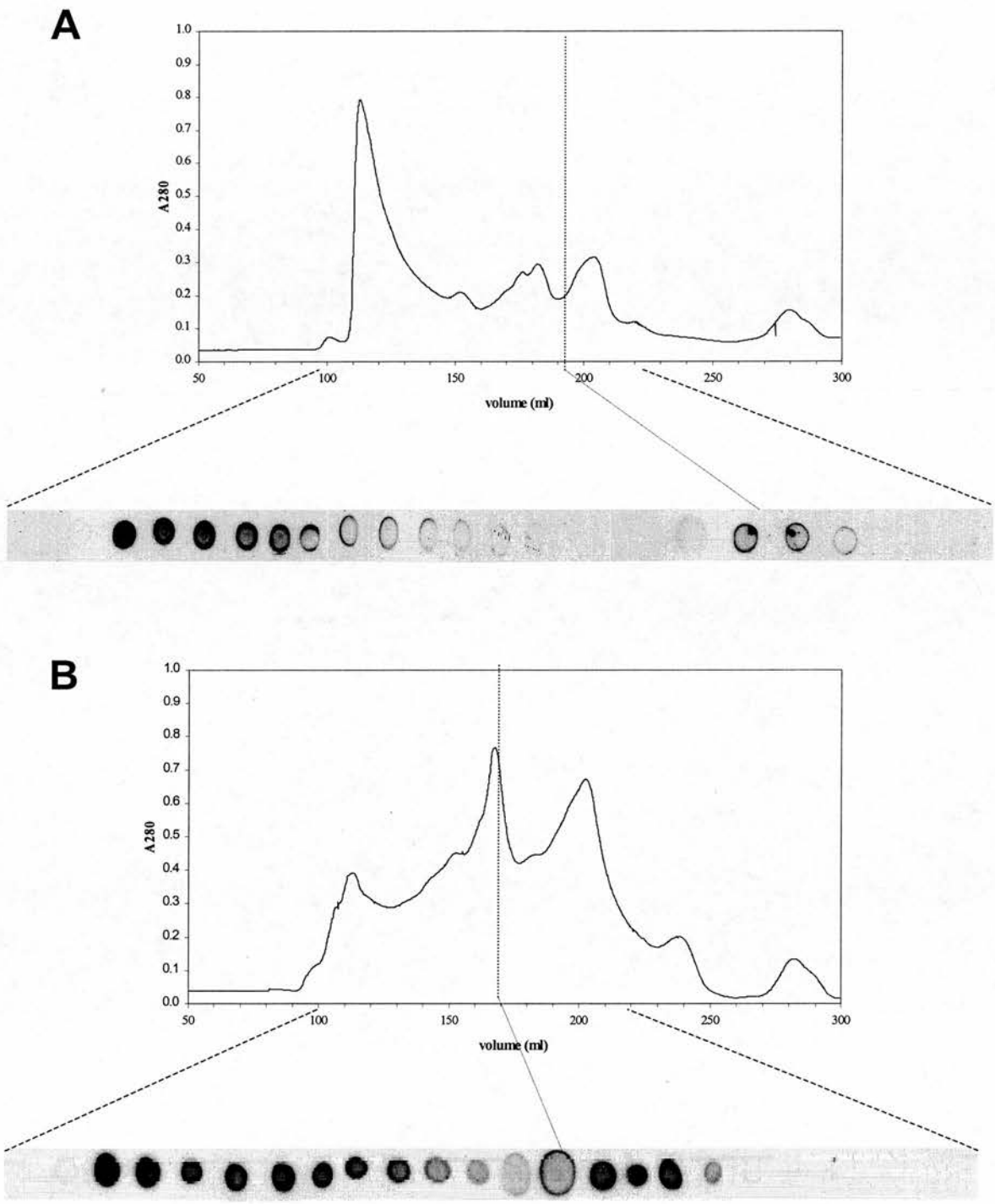


Figure 6.5: Size exclusion chromatography of *C. trachomatis* MOMP and M9-MOMP. **A.** Separation of LDAO-solubilised MOMP with dot blots of collected 5ml fractions. **B.** Separation of Zwittergent 3,14-solubilised M9-MOMP with dot blots of collected 5ml fractions.

that the oligomerisation of MOMP is highly condition dependent. The difference between MOMP and M9-MOMP extracted in Zwittergent 3,14 indicates that disulphide bonds have some role in this. Although the MOMP sample was in sufficient DTT concentrations to fully reduce all the cysteine residues, not all were necessarily solvent accessible. This could account for the larger proportion of very high molecular weight protein eluting with the void volume compared to cysteine-less MOMP. However, disulphide bonds are not the only factor in oligomerisation, as seen with the distinct difference between MOMP solubilised in two alternative detergents, and that M9-MOMP did not elute as monomers. In these circumstances it is difficult to define the true multimeric state of MOMP.

6.4 Planar lipid bilayer analysis of *C. trachomatis* MOMP

In addition to examining the structure of recombinant MOMP it was important to consider function. To this end, chromatography purified MOMP was reconstituted into planar lipid bilayers. Δomp8 cells containing empty vector were grown and expressed, and the outer membranes extracted and purified by chromatography. Fractions matching the volumes where MOMP had eluted were retained. These were added to the *cis* chamber of the bilayer setup. In six independent attempts, no channels incorporated into the bilayer from these controls. Additionally, no channels were obtained when MOMP samples from the first peak at the void volume were added. This was perhaps unsurprising, as the protein there was likely to be misfolded aggregate or in complexes

simply too large to enter the bilayer. However, when MOMP from the second peak was used, channels were incorporated. On the addition of MOMP in Zwittergent 3,14, porin-like unit conductances were observed in the bilayer (figure 6.6). With the *cis* chamber voltage clamped at +40mV, the gating was frequent, however as the voltage was increased the channels spent more time closed. The bilayer also became less stable, though the channel openings were still clearly distinguishable. MOMP in LDAO gave very similar channels. Overall, the conductance of the channels was 0.46 ± 0.07nS in 500:500mM KCl (± SD, n=7). This value would be appropriate for a porin, and is consistent with data from *Ch. abortus* MOMP (Wyllie et al.1998).

6.5 Discussion

Previous attempts to express MOMP recombinantly have shown that association with the outer membrane is not sufficient to demonstrate the full insertion of porin into the membrane. While the surface exposure of MOMP does not prove it to have a fully stable, native structure, it does indicate that folding is well advanced. Investigations into the *in vivo* folding pathways of other porins suggest that full insertion into the membrane is a relatively late step in the process (Jansen et al.2000). When performing the whole cell immunoblots to test for surface exposure, keeping the bacterial membrane intact during the experiment was of primary importance. Cells were taken from the inducing culture before saturation so that most cells would be intact. The bacteria were then washed gently in PBS to remove any lysed cells or MOMP that had been released into the medium. As PVDF membranes must be pre-activated in methanol, nitrocellulose

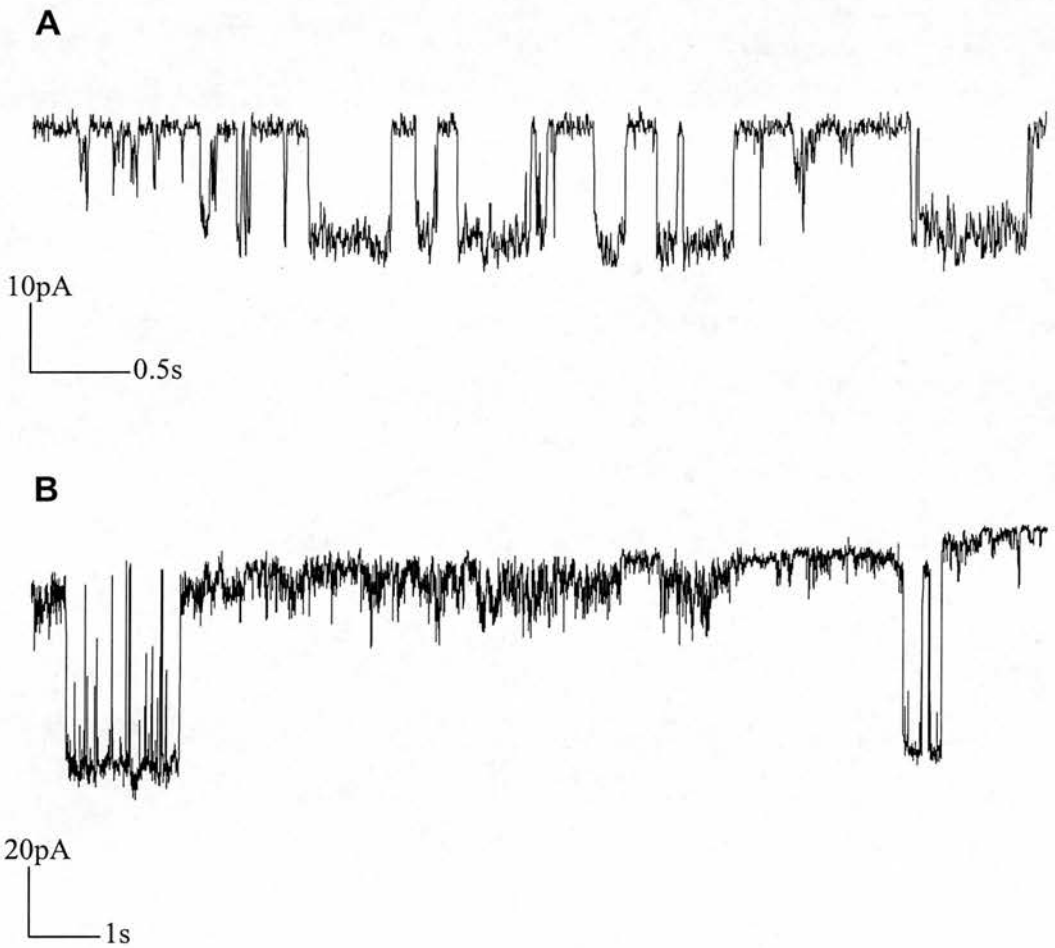


Figure 6.6: ***C. trachomatis*** **MOMP reconstituted into planar lipid bilayers.** Recordings made in symmetrical 500mM KCl, 2mM DTT. Downward currents represent channel openings. **A.** Voltage-clamped at +40mV. **B.** Voltage-clamped at +100mV.

membranes were used in these experiments, to ensure no lingering methanol could permeablise the outer membrane to antibody. Finally, a control for cell lysis during the protocol was provided, in the form of BL21 cells expressing the mature versions of the constructs. Deliberate lysis of these control cells resulted in a positive signal on the immunoblot (not shown). Unfortunately the Δomp8 cells could not be used in these experiments as the more fragile membrane meant that they did not survive the protocol without rupturing. However, as MOMP produced in Δomp8 cells has in all other aspects displayed the same characteristics as that produced in standard BL21s, it is likely to also be surface exposed. The signal from cells expressing oT-TR was stronger than the nL-TR in both wild type and M9-MOMP. This is most likely due to the previously demonstrated better processing of protein with the *E. coli* leader compared to the chlamydial leader, resulting in a larger quantity of mature MOMP available for folding and membrane insertion. Of particular interest is that immunoblots consistently failed when the bacteria were grown at 37°C, although processed, mature MOMP is present at this temperature. This suggests that, although the cell can translocate and digest MOMP efficiently at the higher temperature, at some point in the folding/insertion pathway the process breaks down and the protein accumulates in the periplasm. This could be where previous attempts to express MOMP in this way have failed. Membrane-bound intermediates in the *E. coli* porin OmpA folding pathway have also been shown to be stabilised at lower temperatures (Kleinschmidt et al.1996). The lower temperature of expression in the method presented here could slow the periplasmic aggregation of

MOMP, and help stabilise folding intermediates, sufficiently to allow a portion of the protein to complete the folding pathway and insert into the membrane.

The oligomerisation of MOMP is complex. When outer membranes were solubilised directly into SDS-loading buffer without prior reduction, several bands were seen, corresponding to multiples of one, two, three and more subunits of MOMP. When reduced, a strong, denatured monomeric band at 40kDa was observed. The same result is shown when *C. trachomatis* chlamydial EBs are solubilised directly into SDS-loading buffer (Newhall et al.1983), indicating MOMP in the *E. coli* outer membrane is in a similar environment as in the extracellular form of the native bacteria. MOMP solubilised from *Ch. abortus* in the detergent octyl-β-D-glucopyranoside or non-denaturing concentrations of SDS runs on SDS-PAGE as an SDS-resistant 100kDa oligomer, in the presence of DTT (Wyllie et al.1998). Here, a portion of Zwittergent 3,14-solubilised MOMP has been shown to elute from size exclusion chromatography in DTT at the approximate molecular weight of a trimer. This oligomeric protein was not SDS-resistant. Changing the detergent to LDAO resulted in a change of elution pattern, where the non-complexed MOMP eluted with the molecular weight of a dimer.

Despite the presence of sufficient DTT in these experiments to fully reduce all disulphide bonds, the cysteine residues are still involved in the structure, as seen by the differences between elution patterns of wild type and M9-MOMP. There is less protein in the early void volume peak with M9-MOMP. This very high molecular weight peak

is probably composed of two distinct populations of MOMP. Some may be misfolded protein that did not complete the folding and membrane insertion pathway and instead aggregated in the periplasm, as has been seen with other attempts at producing recombinant MOMP in this way. This would be the case for M9-MOMP as well as for the wild type protein. The second population may be of highly disulphide bonded oligomers of MOMP in the outer membrane. There would be none of this group in the M9-MOMP sample, explaining the decreased void volume peak. Although the non-complexed M9-MOMP peaks to a maximum at the molecular weight of a tetramer, the pattern is less well defined than for wild type MOMP with protein detectable across a wide range.

It seems clear that the quaternary structure of MOMP is highly condition dependent, both in terms of lipid and/or detergent environment and in the redox state of the cysteine residues. The latter is related to the former by solvent accessibility for reducing agents. The dimeric and trimeric MOMP obtained here have not been SDS-resistant when reduced. This could be attributable to differences between the *Ch. abortus* MOMP that previously demonstrated possible trimers, and the *C. trachomatis* MOMP used here. The oligomers of *C. trachomatis* MOMP observed in EBs were not SDS-resistant when reduced (Hackstadt et al.1985). Repeat experiments need to be extended across a range of species so that data from different MOMPs can be combined and compared effectively. Alternatively, MOMP as a defined trimer may be purely artifactual. It has been shown here that simply changing the solubilising detergent can affect the number

of subunits associating. This raises concerns that detergent solubilised MOMP may not be a true reflection of its situation *in vivo*.

Importantly, recombinant MOMP produced here has been shown to form porin-like channels in a planar lipid bilayer, comparable in size to native MOMP. More analysis, in particular on the effect of redox potential, is needed to fully characterise the channels and interpret their other properties.

Chapter 7: *In vitro* refolding of chlamydial outer membrane proteins

7.1 Investigation of refolding conditions

Initial attempts at refolding MOMP from guanidine solubilised inclusion bodies were made according to protocols developed for other porins (Minetti et al.1997; Jansen et al.2000), without success (not shown). A screening process was therefore employed to establish folding conditions empirically, using the FoldIt Protein Folding Kit (Hampton Research). As MOMP has no enzymatic activity some other form of assay was required to measure success. Using a lipid bilayer or vesicle based assay to test for channel function was inappropriate, as the lipidic environment of the protein in these systems could have either an enhancing or deleterious effect on the folding of the protein that may be unrelated to the initial folding conditions. However, misfolded proteins tend to aggregate and precipitate from solution, and this can be easily measured with a visible spectrometer. Initial screens were carried out with 0.1% (w/v) LDAO and 2mM DTT added to all FoldIt reagent combinations and diluting denatured MOMP into the solutions. These screens were assessed by measuring the absorbance of the solutions at 400nm, with high absorption indicating marked precipitation and low absorption showing that the protein remained in solution. A comparison of the absorbance readings with different reagents provides a numerical analysis of the tendency of each reagent to promote or inhibit misfolding and aggregation of the protein.

Numerous conditions were investigated using this screen; including pH, protein concentration, salt concentration, the presence of an osmolyte (PEG 3350), a chaotrope (GuHCl), the addition of cations (Ca^{2+} and Mg^{2+}) or chelator (EDTA) and the presence of nonpolar (sucrose) and polar (arginine) additives (see appendix D.1 for combinations). The absorptions of all the samples containing each individual reagent (eg. GuHCl) were added together, then presented as a proportion of the total absorption of all the samples (table 7.1, see appendix D.2 for original readings). As each reagent was present in half of the screening combinations, a reagent that had no effect on folding should have a relative total absorbance of 0.5. Values >0.5 show a tendency to increase misfolding and aggregation, while values <0.5 show a tendency to inhibit them. The further from 0.5 the value is, the larger the impact on folding. Some conditions stood out in particular. *C. muridarum* MOMP showed a strong preference for pH 8.2, low protein concentration and the presence of a chaotropic agent in its refolding buffer. Milder preferences were seen for low salt concentration, and the presence of a chelating agent and a polar additive. Similar results were seen when the cysteine-less *C. trachomatis* M9-MOMP was screened, though with generally milder preferences: perhaps a reflection of a decreased tendency towards aggregation without any possibility of disulphide bond formation. The screen was repeated by dialysis instead of dilution, but the protein always showed higher absorption at 400nm than that observed in the equivalent dilution, so that method was not pursued. These basic conditions were used to further explore the refolding conditions of MOMP in a second round of screening.

	Relative Total Absorbance	
Condition	**MP**	**M9**
pH 8.2	0.06	0.07
pH 6.5	0.94	0.93
low [protein]	*0.12*	*0.34*
high [protein]	*0.88*	*0.66*
high NaCl	0.59	0.57
low NaCl	0.41	0.43
+ osmolyte	*0.54*	*0.58*
- osmolyte	*0.46*	*0.42*
+ chaotrope	0.13	0.34
- chaotrope	0.87	0.66
cation	*0.57*	*0.55*
chelator	*0.43*	*0.45*
+ nonpolar	0.51	0.48
- nonpolar	0.49	0.52
+ polar additive	*0.42*	*0.45*
- polar additive	*0.58*	*0.55*

Table 7.1: FoldIt screen of refolding from denatured *C. muridarum* MOMP (MP) and *C. trachomatis* M9-MOMP. MOMP diluted into FoldIt screening reagents were mixed overnight at 4°C. The absorbances at 400nm were measured and added for each condition. Data are presented relative to the total absorbance of all samples.

7.2 Circular dichroism (CD) analysis of refolding conditions

While lack of aggregation was a useful assay for the rapid elimination of incorrectly folded samples in the initial screens, a more positive assay was required for further analysis of refolding. Circular dichroism is a spectroscopic technique that measures the difference in absorption of left and right circularly polarised radiation by chiral molecules. Proteins are inherently so due to the chirality of all the composite amino acids, except glycine. The far UV region corresponds to absorption by peptide bonds and analysis of the CD spectrum of proteins can determine the content of regular secondary structural features (Kelly et al.2000). For example, α-helical structure displays characteristic double minima at 208 and 222nm, whereas β-sheet structure shows a single minimum at 217nm and has a mean residue ellipticity of zero at a longer wavelength than α-helix (figure 7.1). Native MOMP was shown to have similar secondary structure to other, known porins in CD analysis (Wyllie et al.1998) therefore the technique was adopted as an assay to examine the refolding of MOMP.

C. trachomatis M9-MOMP was diluted to approximately 0.5mg/ml into refolding buffers containing the reagents suggested by the above screening (50mM Tris-HCl pH 8.2, 1mM EDTA, 10mM NaCl, 500mM GuHCl, 500mM L-Arginine) plus various detergents at concentrations of 1, 10 and 50 times the critical micelle concentration (CMC). The refolding mixtures were incubated overnight at 4°C with stirring. All samples containing the detergent zwittergent 3,14 and n-octyl-β-D-glucopyranoside

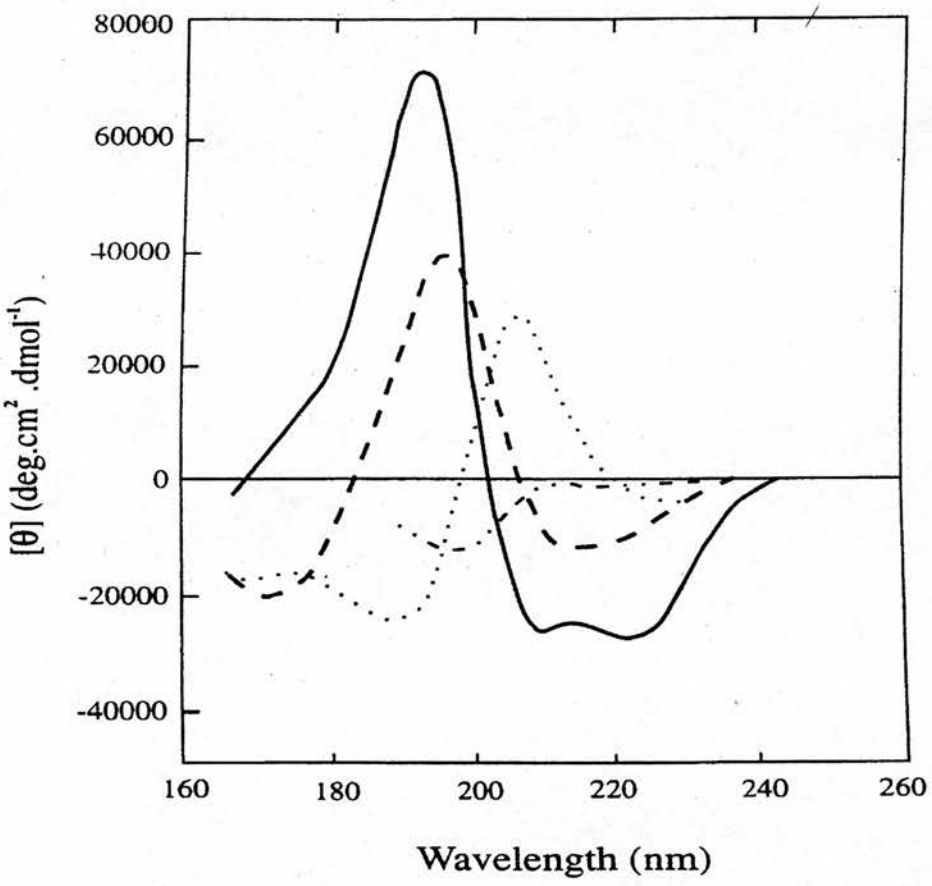
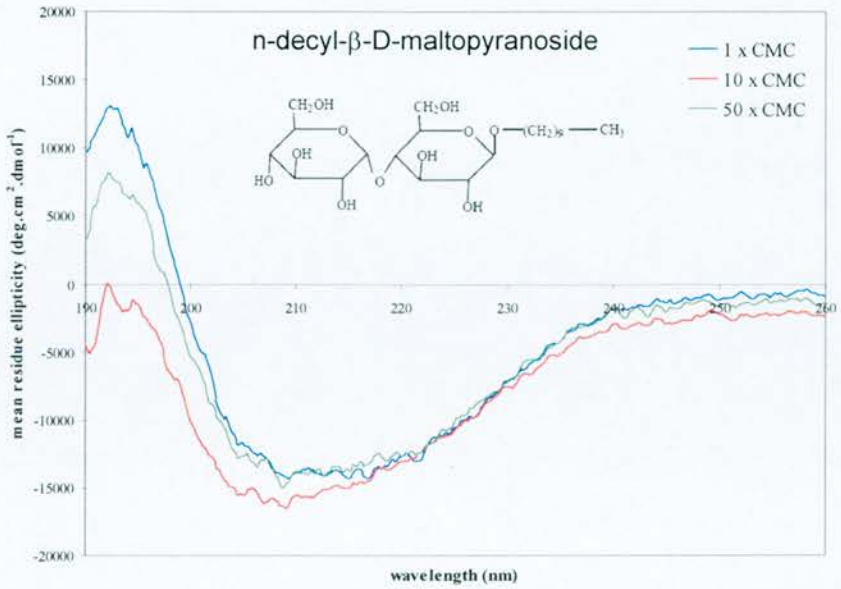


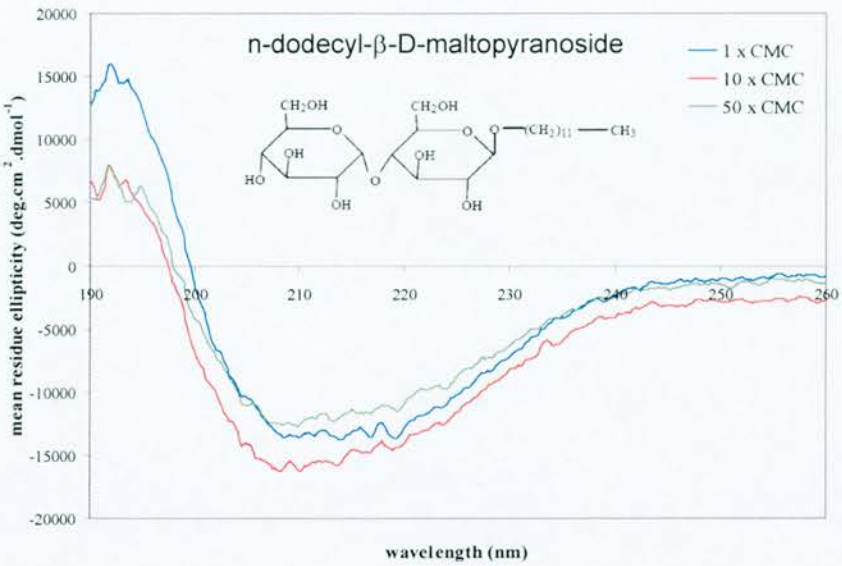
Figure 7.1: Far UV CD spectra of various secondary structures. Solid line shows α-helix, long dashes anti-parallel β-sheet, dots type I β-turn and dots and dashes irregular structure. From Kelly and Price (2000).

completely precipitated at this stage. Where the protein was still in solution and the absorbance at 400nm measured <0.05, the sample was dialysed with two changes to remove the guanidine and arginine, and to reduce the detergent concentration to just above the CMC so the background detergent absorption was minimal for the CD analysis. The sample was then dialysed with two changes into CD buffer, containing 50mM sodium phosphate pH 8.0 and just above one times the CMC of the appropriate detergent. CD spectra were measured (figure 7.2A,B,D and E) and the results analysed using the CONTINLL, SELCON3 and CDSSTR programs. The results of all three programs were similar, therefore only those from CDSSTR are shown (figure 7.2C and F). MOMP diluted into the non-ionic detergents n-decyl-β-D-maltopyranoside (DM) and n-dodecyl-β-D-maltopyranoside (DDM) showed spectra indicative of a large portion of α-helical structure. This could be seen in the distinctive minima at 208nm and in the trace crossing the axis below 200nm wavelength. This was borne out in the computer analysis, with more than a third of the overall structure assigned to α-helix, generally more than was assigned to β-structure. MOMP diluted into LDAO also showed a spectrum characteristic of significant helical structure. On analysis the proportion of structure assigned to α-helix was less than with DM and DDM, however this was balanced by an increase in irregular structure. Higher concentrations of LDAO increased the proportion of irregular structure in the protein further. Finally, MOMP was diluted into refolding buffer containing Cymal-6. The spectra showed less pronounced helical features, with higher concentrations of detergent showing strong β-structure content. The ellipticity at 208nm was increased with increasing detergent, and

A



B



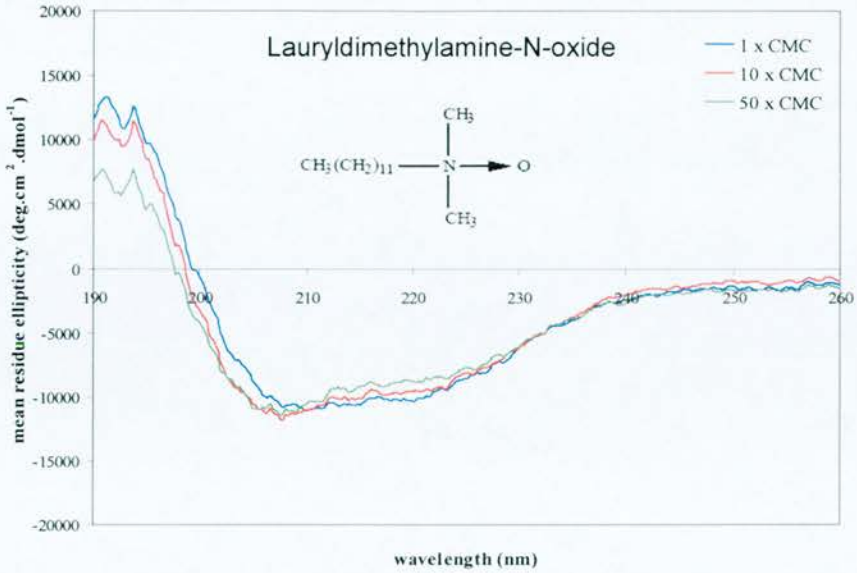
C

DM	α	β	irregular
1 x CMC	38%	35%	27%
10 x CMC	38%	31%	31%
50 x CMC	40%	37%	25%

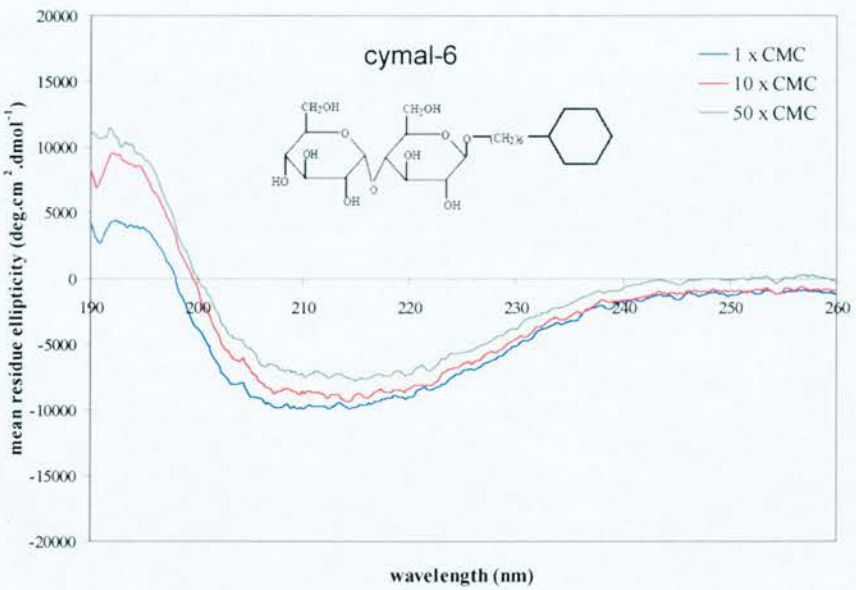
DDM	α	β	irregular
1 x CMC	36%	31%	33%
10 x CMC	45%	36%	18%
50 x CMC	34%	36%	30%

Figure 7.2: Circular dichroism analysis of *C. trachomatis* M9-MOMP refolding in various detergent environments. **A.** CD spectra of MOMP refolded in different concentrations of DM. **B.** In DDM. **C.** CDSSTR analysis of secondary structure contents from A and B. **(D, E and F overleaf)**

D



E



F

LDAO	α	β	irregular
1 x CMC	30%	35%	36%
10 x CMC	31%	31%	39%
50 x CMC	24%	29%	47%

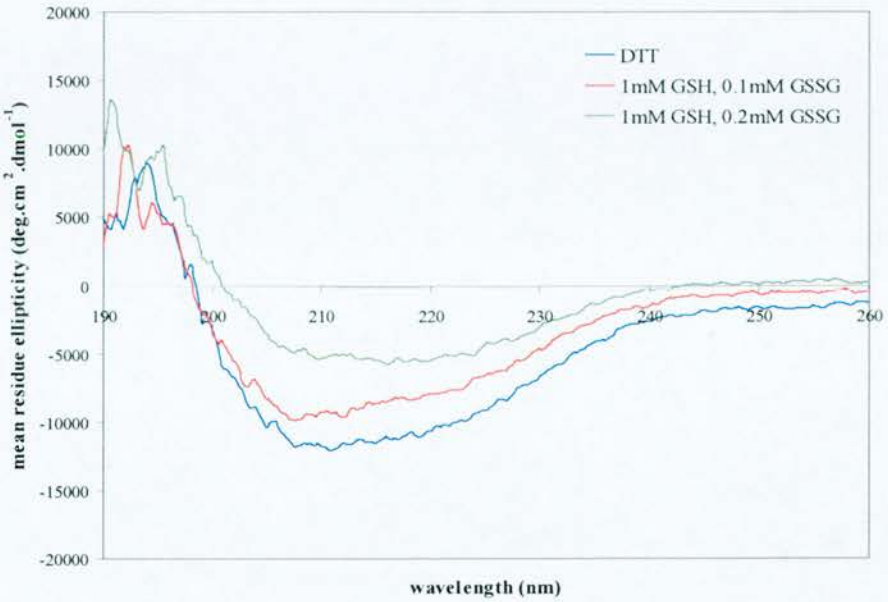
CYMAL-6	α	β	irregular
1 x CMC	23%	38%	40%
10 x CMC	22%	42%	37%
50 x CMC	17%	47%	34%

Figure 7.2: Circular dichroism analysis of *C. trachomatis* M9-MOMP refolding in various detergent environment (cont). **D.** CD spectra of MOMP refolding in different concentrations of LDAO. **E.** In Cymal-6. **F.** CDSSTR analysis of secondary structure contents from D and E.

in the 50 times CMC sample the minimum was at the 217nm associated with β-sheet. The computer analysis agreed, with less α- and more β-structural content as the Cymal-6 was increased. The effect of redox on secondary structure and folding was examined using *C. muridarum* MOMP, which contains 8 cysteine residues. MOMP was diluted into refolding buffers as before containing 1 x CMC LDAO and different redox conditions. Similarly to the M9-MOMP, *C. muridarum* MOMP with 2mM DTT showed a spectrum with the 208nm minimum, and analysis assigned the largest portion of the protein structure as irregular (figure 7.3). Refolding buffer containing 1mM:0.1mM reduced (GSH) to oxidised (GSSG) glutathionine showed only slightly improved structural assignments. However when the ratio was decreased to 1mM GSH:0.2mM GSSG there was a large change in both the CD spectrum and the computer analysis. The ellipticity at 208nm was greatly reduced, with the minimum now at the β-sheet indicative 217nm. This was accompanied by a shift to the right where the spectrum crosses the x-axis, to greater than 200nm, again a sign of increasing β-content. The analysis assigned only 7% of the structure to α-helical and 59% to β-structural content.

7.3 Refolded *C. muridarum* MOMP

The various conditions identified as inhibiting aggregation and promoting β-structure formation were combined and used to refold *C. muridarum* MOMP. MOMP was diluted dropwise 1:50 into refolding buffer containing Tris-HCl pH 8, EDTA, NaCl, GuHCl and arginine as detailed above, plus 1.5% (w/v) cymal-6, 1mM GSH and 0.2mM GSSG.



	α	β	irregular
DTT	24%	34%	43%
1mM GSH, 0.1mM GSSG	20%	40%	40%
1mM GSH, 0.2mM GSSG	7%	59%	33%

Figure 7.3. The effect of redox conditions on secondary structure. *C. muridarum* MOMP diluted into 1 x CMC LDAO refolding buffer containing either 2mM DTT or 1mM reduced glutathionine with 0.1 or 0.2mM oxidised glutathionine.

This refolding mix was incubated at 4°C overnight with gentle stirring. The protein was then dialysed against refolding buffer without GuHCl or arginine and with only 0.03% (w/v) Cymal-6 (just over the CMC). A portion of the sample was further dialysed into an appropriate CD buffer and the absorbance measured. The spectrum was strongly characteristic of anti-parallel β-sheet (figure 7.4). The computer analysis assigned 65% of the secondary structure as β-structure, with only 6% α-helical. The remaining 29% was irregular structure. These values are similar to that obtained from native *Ch. abortus* MOMP (Wyllie et al.1998).

While the correct formation of secondary structure is critical to folding, it does not necessarily prove the final tertiary fold has been made. Therefore the refolded MOMP was further characterised. When separated by non-reducing SDS-PAGE, MOMP ran as a very high molecular weight band that barely entered the separating gel, but fell apart into denatured monomers upon incubaton with reducing agent (not shown). Protease resistance is another indicator of porin β-barrel formation. MOMP briefly digested with proteinase K then separated by non-reducing SDS-PAGE ran as a heat modifiable, SDS-resistant trimer (figure 7.5). The trimer denatured between 65 and 70°C, a temperature similar to that seen with other trimeric porins. The proteinase K-treated MOMP also dissociated into denatured monomers in the presence of reducing agent. Taken together, these results indicate that MOMP has refolded into a high order complex of MOMP subunits composed of stable trimers. The stability of both the trimer and the larger complex is dependent on disulphide bond formation.

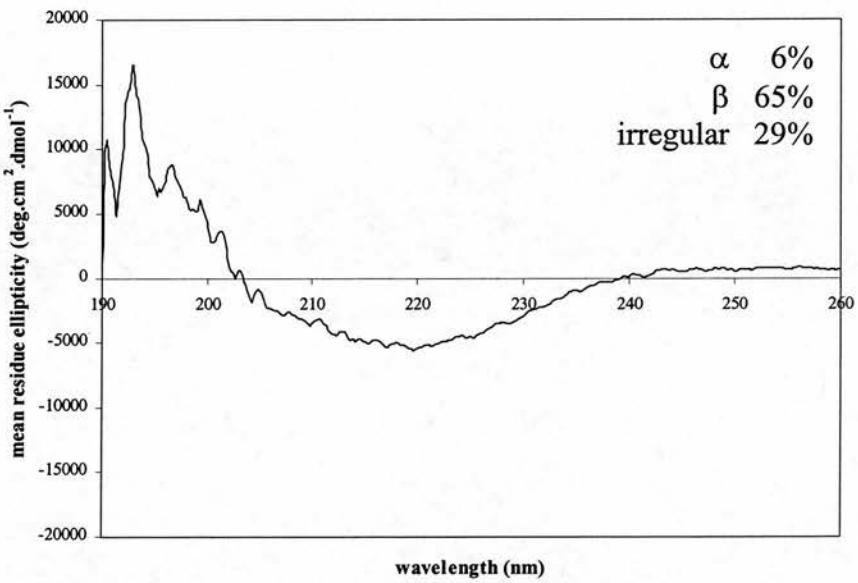


Figure 7.4: CD spectrum of refolded *C. muridarum* MOMP. MOMP in 50mM sodium phosphate, pH 8, 0.03% cymal-6, 1mM GSH and 0.2mM GSSG. Secondary structure assignments from CDSSTR analysis of the spectrum.

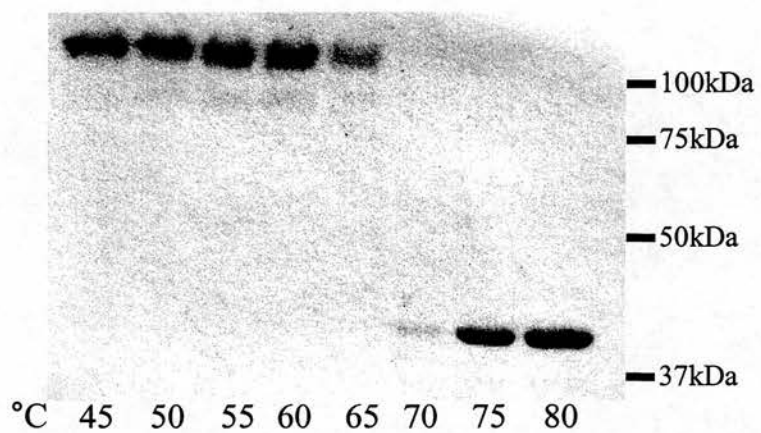


Figure 7.5: Thermostability of refolded *C. muridarum* MOMP. Proteinase K-treated MOMP samples were incubated at each temperature for 10min.10% (w/v) polyacrylamide, non-reducing SDS-PAGE. Stained with GelCode colloidal blue stain.

7.4 Discussion

The process of *in vitro* refolding of proteins from inclusion bodies is poorly understood and can be difficult to achieve, particularly with membrane proteins. As with crystallisation, even slight changes in sequence can affect the conditions required for success. Amongst porins, despite the strong structural similarity, there is very little sequence homology and it was perhaps unsurprising that protocols used for other porins did not work with MOMP. Establishing refolding conditions empirically can be very time-consuming due the huge range of combinations possible. Using the FoldIt screen together with spectroscopic measurement allowed the rapid optimisation of many of the basic conditions without wasting time or resources. The smaller range of combinations could then be analysed further by methods that give more information about the structural state of the protein. CD analysis was a useful tool at this stage. It is still relatively quick and easy to do yet can give useful information on the progress of protein folding. Unlike the standard spectroscopic measurement, CD can distinguish intermediate levels of folding at secondary structural level, allowing the rational optimisation of conditions. These two levels of screening suggested a set of conditions most favourable to the correct folding of MOMP.

The refolded MOMP showed characteristics that suggested the folding was correct. The secondary structure was comparable to that of native MOMP from *Ch. abortus*. Although no CD analysis of native *C. muridarum* was available, it is likely to be very

close to that of other species. *C. muridarum* and *Ch. abortus* share 62% sequence identity and much of the differences are concentrated in the VS domains that, as extracellular loops, probably are within the irregular portion of the secondary structure. In any case, both the native and recombinant MOMPs exhibit secondary structure characteristic of other known porins. Although MOMP ran as a very large molecular weight band on non-reducing SDS-PAGE there is evidence to suggest this was not just a misfolded aggregate. During initial screening MOMP readily precipitated when misfolded, and when these samples were separated by SDS-PAGE, the aggregate did not enter the gel and was unaffected by reducing agent (not shown). In contrast the refolded MOMP always remained soluble and disassociated into denatured monomer in the presence of reducing agent. On digestion with proteinase K, the large complex runs as a porin-like trimer, that is highly stable, SDS and heat resistant when oxidised. Although more analysis is required, these results suggest that MOMP has refolded correctly into β-barrels that form disulphide-mediated stable trimers. Further disulphide bond formation results in the oligomerisation of the trimers into higher order complexes of MOMP that may in fact be more truly representative of the *in vivo* state of MOMP than the smaller trimers shown.

Unfortunately it has not been possible to obtain planar lipid bilayer recordings of this *in vitro* refolded MOMP. The higher order complexes are simply too large to insert into the bilayer and trimers produced by proteinase K treatment only resulted in unstable, flickering openings without well-defined unit openings. It is possible that the protease

treatment caused some nicks to be made in the loops of the protein that allowed the breaking apart of the large complexes, but also adversely affected protein function.

Establishing an effective *in vitro* refolding method for recombinant MOMP would make a valuable contribution to chlamydial research. The main advantage of MOMP produced this way is the very large yields of protein that can readily be obtained. The protein produced in inclusion bodies is also easily purified and so it would make an excellent source of MOMP. This would make it possible to pursue crystallisation studies and more detailed analyses of MOMPs function as a porin and possible adhesin and its development as an immunological target.

Chapter 8: Conclusions

8.1 Summary

Chlamydial infections represent a major global health problem, with the different species responsible for a wide range of human and animal diseases. MOMP is the immunodominant chlamydial protein, with potential structural and functional roles as a porin, adhesin and, as part of the COMC, providing strength and osmotic stability to the cell. MOMP is therefore considered to be a primary vaccine target. Constraints in the culturing of the intracellular bacteria and difficulties in purifying MOMP and separating it from other outer membrane proteins have made it difficult to effectively study MOMP biochemistry and develop its vaccine potential. The establishment of a reliable recombinant expression system would greatly facilitate the furtherance of these aims.

A topology prediction for *C. trachomatis* MOMP has been presented here, based on analysis of the primary amino acid sequence by three artificial neural network programs. The resulting model complied well with the general rules for β-barrel membrane proteins, and with other experimental data concerning MOMP. MOMP was predicted to be a 16-stranded β-barrel with an average strand length of 9 amino acids, none of which contained proline residues. The "aromatic belts" of outward facing aromatic residues seen in other porins were also observed here. Crucially, the highly immunogenic VS domains were on the long extracellular loops and two of the cysteine residues were on

periplasmic turns, allowing important interactions with the periplasmic OmcB and OmcA on the inner leaflet of the outer membrane. Most other cysteine residues were on extracellular loops, and so could be available for intersubunit cross-linking.

A large range of chlamydial porin constructs have been made. These include *ompA* from a further two species of chlamydiae (*C. trachomatis* and *C. muridarum*) and *porB,* all without any signal sequence, to add to the two mature *ompA* constructs available in the lab (*Ch. abortus* and *Ch. pneumoniae*). Constructs containing *ompA* and *porB* genes including their native chlamydial signal sequence or an *E. coli* signal sequence have also been made (*ompA* from *C. trachomatis*, *C. muridarum* and *Ch. abortus*). A mutant *C. trachomatis ompA,* with all nine cysteine residues converted to alanine residues, has been produced in both the mature form of the gene without a signal sequence, and with both the native and *E. coli* signal sequences. Finally, internal mutagenesis on the *C. trachomatis ompA* with the chlamydial leader has been used to create four constructs, each lacking one of the MOMP VS domains.

Two different approaches to the production of recombinant MOMP have been developed using these constructs. The expression of mature MOMP and PorB as inclusion bodies was optimised to give a large yield of recombinant protein dissolved in chaotrope. The protein was expressed in porin-deficient Δomp8 BL21 cells and processed through a rigorous clean up protocol to achieve a high level of purity. High levels of recombinant protein could be produced from all mature constructs by this method. Recombinant

porin expressed with a signal sequence was targeted to the membrane fraction of the cell. This method of expression inhibited both the growth rate of the bacteria and cell survival. The use of the *E. coli* ompT signal sequence rather than the chlamydial leader helped to prevent this. There was considerable variability between constructs. *C. trachomatis* and *C. muridarum* MOMP expressed well by this method, PorB with the ompT leader expressed less well, and *Ch. abortus* MOMP could not be expressed at all. The production of the ompT leadered *C. trachomatis* MOMP was further optimised to increase the yield of protein fully processed to its mature form. This included reducing the temperature of induction to 16°C, and culturing the bacteria in an alternative growth medium.

Outer membrane-expressed *C. trachomatis* MOMP was demonstrated to be surface exposed in intact BL21 cells. Membrane insertion was temperature dependent. The oligomerisation of MOMP examined directly from the membrane was similar to what previously observed with MOMP in *C. trachomatis* EBs, and protein solubilised in detergent from the *E. coli* outer membrane formed porin-like channels in a planar lipid bilayer. Together these results strongly suggest that recombinant MOMP produced by this method is folding and inserting into the outer membrane in a native-like manner. However, the oligomerisation of detergent-solubilised MOMP is highly condition dependent, and the choice of detergent and redox potential influence subunit association. The cysteine-less MOMP was able to insert into the membrane, indicating disulphide bonds are not essential for at least the early part of the folding pathway. Similarly the

VS domain deletion mutants were showed to be surface exposed when expressed in BL21s, demonstrating that they are not, at least individually, essential for folding.

MOMP solubilised in GuHCl was screened to optimise conditions for refolding into its native tertiary structure. Numerous conditions were examined, including pH, salt concentration, addition of chaotrope, detergent choice and concentration and redox state. CD analysis of the final folded *C. muridarum* MOMP assigned β-structure of 65%, a value consistent with that obtained from native MOMP and other porins. This refolded MOMP was composed of SDS-resistant, heat-modifiable trimers that were further cross-linked into a large oligomers. Both trimer and oligomer dissociated into monomers in the presence of reducing agent, suggesting MOMP is forming disulphide bond mediated complexes, as in the native environment. Large quantities of refolded MOMP could be obtained by this method.

8.2 Discussion

Recombinant MOMP has been successfully produced by two different expression methods, with *in vivo* and *in vitro* (re)folding. There are advantages to each method, and the protein produced can be put to different uses. MOMP expressed via inclusion bodies has much higher yield and purity than outer membrane expressed MOMP. This makes it useful for bulk protein production and structural studies, which require larger quantities of protein and more homogenicity than that observed in the *in vivo* folded MOMP. It is

also the only route possible for *Ch. abortus* MOMP, which failed to express at all in the membrane. However, the large oligomers formed make it difficult to manipulate and difficult to study the biochemistry of MOMP. The conditions for *in vitro* refolding will not necessarily apply to MOMP from all the species expressed, and in particular are likely to be different for PorB. The screening and optimisation of folding would therefore need to be repeated, possibly for each protein. This would be highly time consuming, and yet still may not ultimately be successful for all constructs.

Membrane expressed protein may be more appropriate for examining the biology of MOMP. The yield of protein is not as high as for IB produced MOMP, but it is sufficient for biochemical applications. It has already been demonstrated that the system can tolerate substantial internal mutagenesis, and still process protein and insert it into the outer membrane. Protein appropriate for electrophysiological analysis can be produced by this method, allowing detailed analysis of the function of mutants, which cannot be obtained as effectively any other way. MOMP seems to behave in the membrane in a similar way to that in chlamydial EBs. As the EBs represent the infectious stage of the chlamydial lifecycle, this system could be useful for the development of vaccines against the organism, acting as cellular "mimics" for the chlamydiae themselves, but much easier to work with. This would also prove useful for studying the antigenic and adhesin properties of MOMP without interference from other chlamydial surface molecules.

Combining data from both systems gives some insights into the biochemistry of MOMP, in particular concerning the role of the conserved cysteine residues. CD analysis of refolding conditions demonstrated that some disulphide bonding stabilises the secondary structure of MOMP. Cysteine-less *C. trachomatis* MOMP (M9-MOMP), and *C. muridarum* MOMP in DTT, showed similar β-structural content (35% and 34% in 1 x CMC LDAO, respectively). Altering the redox potential surrounding *C. muridarum* MOMP resulted in a large increase in β-structure (59% in 1 x CMC LDAO). Disulphide bonds also stabilise the tertiary structure of MOMP. Reducing *C. trachomatis* MOMP from outer membranes resulted in the collapse of the SDS-resistant folded monomer to denatured monomer. The oligomers of various subunit multiples were also denatured by reducing agent, but not by boiling. Likewise, the large oligomers formed by MOMP when it was refolded from IBs dissociated into monomers when disulphide bonds were reduced. Together these data suggest that disulphide bonds are important for forming inter-subunit connections. Overall, cysteine residues are clearly important at several levels for the structure and/or stability of MOMP.

In detergent solubilised MOMP, it seems likely that a trimer is the preferable structure, as previous work, Zwittergent 3,14-solubilised MOMP and refolded MOMP have all displayed trimeric properties. In CD analysis LDAO was more destabilising to the protein than other detergents, which may explain the difference in subunit associations seen in LDAO-solubilised MOMP. However, this is not proven. MOMP examined directly from chlamydial and *E. coli* membranes showed no preference for trimers, and

results presented here indicated that detergent solubilised MOMP may not be fully accessible to DTT in the solution. Which disulphide bonds have remained intact in these experiments is critical to establishing the accuracy of the structural prediction. The notion of a trimeric MOMP should therefore be considered with caution until these difficulties have been resolved.

8.3 Future work

Refolded IB protein is already being subjected to further structural studies. Early crystallisation trials have been started, using both standard oligomeric, refolded MOMP and MOMP that has been reduced after refolding (to separate the oligomers, but hopefully without completely denaturing the protein). Obtaining a crystal structure for MOMP would be invaluable to the study of the protein and of chlamydial membrane structure. Refolding screening will also be started for other species of MOMP. The availability of several versions will increase the chances of successful crystallisation, and there may also be interesting differences between the species.

The outer membrane expression system will also be extended to include the different species of MOMP, PorB and mutants. Further electrophysiological analysis of *C. trachomatis* MOMP will be used to examine channel properties such as ion selectivity, subunit number and cooperation, and gating in the presence and absence of DTT. Some of these properties have been measured from the native porin, and so will further confirm correct folding and function. Others can be used to examine differences

between MOMP and other porins. *C. muridarum* MOMP and PorB can already be produced by this expression system, and will be analysed in a similar manner to *C. trachomatis* MOMP. The predicted selectivity of PorB can also be tested by these methods.

In order to further examine the role of cysteine residues and disulphide bonds in MOMP structure and function, a better method for controlling the reduction/oxidation level is required. Concerns have been expressed here about restricted access for reducing agent in the detergent solubilised experiments. Additionally, protein was oxidised simply by exposure to air in the absence of reducing agent. Alternative mechanisms for oxidising the protein in a more controlled manner should be explored. Alkylating agents could be used to block thiol groups without concerns about the deterioration of DTT over time in solution. The level of cysteine bond formation could be tested by several means, including mass spectroscopy. However, near UV CD analysis may be a more appropriate method, as it is faster and cheaper, and the sample can be recovered afterwards.

A major difficulty in working with and analysing MOMP in these systems is trying to resolve the influences of the predicted intra- and inter-molecular disulphide bonding. A useful development would be to create a mutant MOMP that contains only the cysteine residues for an intra-molecular bond, giving a stable monomer without the complications of extensive subunit cross-linking. This could be achieved by "adding back" pairs of cysteine residues to the M9-MOMP mutant. The topology prediction shows there are

relatively few candidates for such a bond. For example, the two periplasmic cysteine residues are on opposite sides of the β-barrel and could not form an intra-molecular bond. Such a model would be much more tractable than "wild type" protein, and could be used for several purposes. These could include determining the preferred oligomerisation of MOMP outside the highly cross-linked network of the chlamydial envelope, and providing a form of the protein that may be easier to use for crystallisation. Circular dichroism could again be used to monitor the formation of any disulphide bonds and stopped-flow CD analysis could provide interesting insights into the folding process.

8.4 Conclusion

This thesis has described the development of two different systems for the production of recombinant chlamydial MOMP. Both methods have generated folded MOMP with characteristics comparable to the native protein. The work here provides a strong base for further analysis of the properties of MOMPs, some of which are unique amongst porins, and a useful source of recombinant MOMPs for future vaccine development.

References

Allen JE, Locksley RM, and Stephens RS (1991) "A single peptide from the major outer membrane protein of Chlamydia trachomatis elicits T cell help for the production of antibodies to protective determinants" *J. Immunol.* **147** 674-679

Allen JE and Stephens RS (1989) "Identification by sequence analysis of two-site posttranslational processing of the cysteine-rich outer membrane protein 2 of Chlamydia trachomatis serovar L2" *J. Bacteriol.* **171** 285-291

Baehr W, Zhang YX, Joseph T, Su H, Nano FE, Everett KD, and Caldwell HD (1988) "Mapping antigenic domains expressed by Chlamydia trachomatis major outer membrane protein genes" *Proc. Natl. Acad. Sci. U. S. A* **85** 4000-4004

Balin BJ, Gerard HC, Arking EJ, Appelt DM, Branigan PJ, Abrams JT, Whittum-Hudson JA, and Hudson AP (1998) "Identification and localization of Chlamydia pneumoniae in the Alzheimer's brain" *Med. Microbiol. Immunol. (Berl)* **187** 23-42

Bavoil P, Ohlin A, and Schachter J (1984) "Role of disulfide bonding in outer membrane structure and permeability in Chlamydia trachomatis" *Infect. Immun.* **44** 479-485

Bush RM and Everett KD (2001) "Molecular evolution of the Chlamydiaceae" *Int. J. Syst. Evol. Microbiol.* **51** 203-220

Caldwell HD, Kromhout J, and Schachter J (1981) "Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis" *Infect. Immun.* **31** 1161-1176

Caldwell HD and Perry LJ (1982) "Neutralization of Chlamydia trachomatis infectivity with antibodies to the major outer membrane protein" *Infect. Immun.* **38** 745-754

Christiansen G and Birkelund S (2002) "Is a Chlamydia vaccine a reality?" *Best. Pract. Res. Clin. Obstet. Gynaecol.* **16** 889-900

Cline J, Braman JC, and Hogrefe HH (1996) "PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases" *Nucleic Acids Res.* **24** 3546-3551

Collett BA, Newhall WJ, Jersild RA, Jr., and Jones RB (1989) "Detection of surface-exposed epitopes on Chlamydia trachomatis by immune electron microscopy" *J. Gen. Microbiol.* **135 (Pt 1)** 85-94

Compton LA and Johnson WC, Jr. (1986) "Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication" *Anal. Biochem.* **155** 155-167

Conlan S and Bayley H (2003) "Folding of a monomeric porin, OmpG, in detergent solution" *Biochemistry* **42** 9453-9465

Dascher C, Roll D, and Bavoil PM (1993) "Expression and translocation of the chlamydial major outer membrane protein in Escherichia coli" *Microb. Pathog.* **15** 455-467

de la Maza LM, Fielder TJ, Carlson EJ, Markoff BA, and Peterson EM (1991) "Sequence diversity of the 60-kilodalton protein and of a putative 15-kilodalton protein between the trachoma and lymphogranuloma venereum biovars of Chlamydia trachomatis" *Infect. Immun.* **59** 1196-1201

Diederichs K, Freigang J, Umhau S, Zeth K, and Breed J (1998) "Prediction by a neural network of outer membrane beta-strand protein topology" *Protein Sci.* **7** 2413-2420

Dong-Ji Z, Yang X, Shen C, Lu H, Murdin A, and Brunham RC (2000) "Priming with Chlamydia trachomatis major outer membrane protein (MOMP) DNA followed by MOMP ISCOM boosting enhances protection and is associated with increased immunoglobulin A and Th1 cellular immune responses" *Infect. Immun.* **68** 3074-3078

Eko FO, Lubitz W, McMillan L, Ramey K, Moore TT, Ananaba GA, Lyn D, Black CM, and Igietseme JU (2003) "Recombinant Vibrio cholerae ghosts as a delivery vehicle for vaccinating against Chlamydia trachomatis" *Vaccine* **21** 1694-1703

Everett KD (2002) "Chlamydiae" *Encyclopedia of Life Sciences*

Everett KD, Bush RM, and Andersen AA (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" *Int. J. Syst. Bacteriol.* **49 Pt 2** 415-440

Everett KD, Desiderio DM, and Hatch TP (1994) "Characterization of lipoprotein EnvA in Chlamydia psittaci 6BC" *J. Bacteriol.* **176** 6082-6087

Everett KD and Hatch TP (1991) "Sequence analysis and lipid modification of the cysteine-rich envelope proteins of Chlamydia psittaci 6BC" *J. Bacteriol.* **173** 3821-3830

Everett KD and Hatch TP (1995) "Architecture of the cell envelope of Chlamydia psittaci 6BC" *J. Bacteriol.* **177** 877-882

Gromiha MM, Ahmad S, and Suwa M (2004) "Neural Network Based Prediction of Transmembrane Beta Strand Segments in Outer Membrane Proteins" *J. Comp. Chem* **(in press)**

Hackstadt T, Todd WJ, and Caldwell HD (1985) "Disulfide-mediated interactions of the chlamydial major outer membrane protein: role in the differentiation of chlamydiae?" *J. Bacteriol.* **161** 25-31

Hatch TP (1996) "Disulfide cross-linked envelope proteins: the functional equivalent of peptidoglycan in chlamydiae?" *J. Bacteriol.* **178** 1-5

Hatch TP, Allan I, and Pearce JH (1984) "Structural and polypeptide differences between envelopes of infective and reproductive life cycle forms of Chlamydia spp" *J. Bacteriol.* **157** 13-20

Hatch TP, Vance DW, Jr., and Al Hossainy E (1981) "Identification of a major envelope protein in Chlamydia spp" *J. Bacteriol.* **146** 426-429

Hughes ES, Shaw KM, and Ashley RH (2001) "Mutagenesis and functional reconstitution of chlamydial major outer membrane proteins: VS4 domains are not required for pore formation but modify channel function" *Infect. Immun.* **69** 1671-1678

Ishizaki M, Allen JE, Beatty PR, and Stephens RS (1992) "Immune specificity of murine T-cell lines to the major outer membrane protein of Chlamydia trachomatis" *Infect. Immun.* **60** 3714-3718

Jacoboni I, Martelli PL, Fariselli P, De P, V, and Casadio R (2001) "Prediction of the transmembrane regions of beta-barrel membrane proteins with a neural network-based predictor" *Protein Sci.* **10** 779-787

Jansen C, Heutink M, Tommassen J, and De Cock H (2000) "The assembly pathway of outer membrane protein PhoE of Escherichia coli" *Eur. J. Biochem.* **267** 3792-3800

Johansson M, Schon K, Ward M, and Lycke N (1997a) "Genital tract infection with Chlamydia trachomatis fails to induce protective immunity in gamma interferon receptor-deficient mice despite a strong local immunoglobulin A response" *Infect. Immun.* **65** 1032-1044

Johansson M, Ward M, and Lycke N (1997b) "B-cell-deficient mice develop complete immune protection against genital tract infection with Chlamydia trachomatis" *Immunology* **92** 422-428

Jones HM, Kubo A, and Stephens RS (2000) "Design, expression and functional characterization of a synthetic gene encoding the Chlamydia trachomatis major outer membrane protein" *Gene* **258** 173-181

Kaul R, Duncan MJ, Guest J, and Wenman WM (1990) "Expression of the Chlamydia trachomatis major outer membrane protein-encoding gene in Escherichia coli: role of the 3' end in mRNA stability" *Gene* **87** 97-103

Kawa DE and Stephens RS (2002) "Antigenic topology of chlamydial PorB protein and identification of targets for immune neutralization of infectivity" *J. Immunol.* **168** 5184-5191

Kelly SM and Price NC (2000) "The use of circular dichroism in the investigation of protein structure and function" *Curr. Protein Pept. Sci.* **1** 349-384

Kleinschmidt JH and Tamm LK (1996) "Folding intermediates of a beta-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism" *Biochemistry* **35** 12993-13000

Knight SC, Iqball S, Woods C, Stagg A, Ward ME, and Tuffrey M (1995) "A peptide of Chlamydia trachomatis shown to be a primary T-cell epitope in vitro induces cell-mediated immunity in vivo" *Immunology* **85** 8-15

Koebnik R (1999) "Membrane assembly of the Escherichia coli outer membrane protein OmpA: exploring sequence constraints on transmembrane beta-strands" *J. Mol. Biol.* **285** 1801-1810

Koehler JE, Birkelund S, and Stephens RS (1992) "Overexpression and surface localization of the Chlamydia trachomatis major outer membrane protein in Escherichia coli" *Mol. Microbiol.* **6** 1087-1094

Kubo A and Stephens RS (2000) "Characterization and functional analysis of PorB, a Chlamydia porin and neutralizing target" *Mol. Microbiol.* **38** 772-780

Kubo A and Stephens RS (2001) "Substrate-specific diffusion of select dicarboxylates through Chlamydia trachomatis PorB" *Microbiology* **147** 3135-3140

Kuo CC, Puolakkainen M, Lin TM, Witte M, and Campbell LA (2002) "Mannose-receptor positive and negative mouse macrophages differ in their susceptibility to infection by Chlamydia species" *Microb. Pathog.* **32** 43-48

Lambden PR, Everson JS, Ward ME, and Clarke IN (1990) "Sulfur-rich proteins of Chlamydia trachomatis: developmentally regulated transcription of polycistronic mRNA from tandem promoters" *Gene* **87** 105-112

Lobley A, Whitmore L, and Wallace BA (2002) "DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra" *Bioinformatics.* **18** 211-212

Manavalan P and Johnson WC, Jr. (1987) "Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra" *Anal. Biochem.* **167** 76-85

Manning DS and Stewart SJ (1993) "Expression of the major outer membrane protein of Chlamydia trachomatis in Escherichia coli" *Infect. Immun.* **61** 4093-4098

Markwell MA, Haas SM, Bieber LL, and Tolbert NE (1978) "A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples" *Anal. Biochem.* **87** 206-210

McCafferty MC, Herring AJ, Andersen AA, and Jones GE (1995) "Electrophoretic analysis of the major outer membrane protein of Chlamydia psittaci reveals multimers which are recognized by protective monoclonal antibodies" *Infect. Immun.* **63** 2387-2389

Minetti CASA, Tai JY, Blake MS, Pullen JK, Liang SM, and Remeta DP (1997) "Structural and functional characterization of a recombinant PorB class 2 protein from Neisseria meningitidis. Conformational stability and porin activity" *J. Biol. Chem.* **272** 10710-10720

Mygind P, Christiansen G, and Birkelund S (1998) "Topological analysis of Chlamydia trachomatis L2 outer membrane protein 2" *J. Bacteriol.* **180** 5784-5787

New England Biolabs Inc. (2002) "2002-03 Catalog and Technical Reference"

Newhall WJ (1987) "Biosynthesis and disulfide cross-linking of outer membrane components during the growth cycle of Chlamydia trachomatis" *Infect. Immun.* **55** 162-168

Newhall WJ and Jones RB (1983) "Disulfide-linked oligomers of the major outer membrane protein of chlamydiae" *J. Bacteriol.* **154** 998-1001

Paetzel M, Dalbey RE, and Strynadka NC (2000) "The structure and mechanism of bacterial type I signal peptidases. A novel antibiotic target" *Pharmacol. Ther.* **87** 27-49

Pal S, Theodor I, Peterson EM, and de la Maza LM (2001) "Immunization with the Chlamydia trachomatis mouse pneumonitis major outer membrane protein can elicit a protective immune response against a genital challenge" *Infect. Immun.* **69** 6240-6247

Penttila T, Vuola JM, Puurula V, Anttila M, Sarvas M, Rautonen N, Makela PH, and Puolakkainen M (2000) "Immunity to Chlamydia pneumoniae induced by vaccination with DNA vectors expressing a cytoplasmic protein (Hsp60) or outer membrane proteins (MOMP and Omp2)" *Vaccine* **19** 1256-1265

Preston GM, Haubold B, and Rainey PB (1998) "Bacterial genomics and adaptation to life on plants: implications for the evolution of pathogenicity and symbiosis" *Curr. Opin. Microbiol.* **1** 589-597

Prilipov A, Phale PS, Van Gelder P, Rosenbusch JP, and Koebnik R (1998) "Coupling site-directed mutagenesis with high-level expression: large scale production of mutant porins from E. coli" *FEMS Microbiol. Lett.* **163** 65-72

Provencher SW and Glockner J (1981) "Estimation of globular protein secondary structure from circular dichroism" *Biochemistry* **20** 33-37

Rodriguez-Maranon MJ, Bush RM, Peterson EM, Schirmer T, and de la Maza LM (2002) "Prediction of the membrane-spanning beta-strands of the major outer membrane protein of Chlamydia" *Protein Sci.* **11** 1854-1861

Salari SH and Ward ME (1981) "Polypeptide composition of Chlamydia trachomatis" *J. Gen. Microbiol.* **123** 197-207

Sanchez-Campillo M, Bini L, Comanducci M, Raggiaschi R, Marzocchi B, Pallini V, and Ratti G (1999) "Identification of immunoreactive proteins of Chlamydia trachomatis by Western blot analysis of a two-dimensional electrophoresis map with patient sera" *Electrophoresis* **20** 2269-2279

Sandbulte J, TerWee J, Wigington K, and Sabara M (1996) "Evaluation of Chlamydia psittaci subfraction and subunit preparations for their protective capacities" *Vet. Microbiol.* **48** 269-282

Schulz GE (2000) "B-barrel membrane proteins" *Curr. Opin. Struct. Biol.* **10** 443-447

Sreerama N, Venyaminov SY, and Woody RW (2000a) "Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis" *Anal. Biochem.* **287** 243-251

Sreerama N and Woody RW (1993) "A self-consistent method for the analysis of protein secondary structure from circular dichroism" *Anal. Biochem.* **209** 32-44

Sreerama N and Woody RW (2000b) "Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set" *Anal. Biochem.* **287** 252-260

Stagg AJ, Elsley WA, Pickett MA, Ward ME, and Knight SC (1993) "Primary human T-cell responses to the major outer membrane protein of Chlamydia trachomatis" *Immunology* **79** 1-9

Stephens RS, Mullenbach G, Sanchez-Pescador R, and Agabian N (1986) "Sequence analysis of the major outer membrane protein gene from Chlamydia trachomatis serovar L2" *J. Bacteriol.* **168** 1277-1282

Su H, Morrison RP, Watkins NG, and Caldwell HD (1990a) "Identification and characterization of T helper cell epitopes of the major outer membrane protein of Chlamydia trachomatis" *J. Exp. Med.* **172** 203-212

Su H, Watkins NG, Zhang YX, and Caldwell HD (1990b) "Chlamydia trachomatis-host cell interactions: role of the chlamydial major outer membrane protein as an adhesin" *Infect. Immun.* **58** 1017-1025

Su H, Zhang YX, Barrera O, Watkins NG, and Caldwell HD (1988) "Differential effect of trypsin on infectivity of Chlamydia trachomatis: loss of infectivity requires cleavage of major outer membrane protein variable domains II and IV" *Infect. Immun.* **56** 2094-2100

Swanson AF, Ezekowitz RA, Lee A, and Kuo CC (1998) "Human mannose-binding protein inhibits infection of HeLa cells by Chlamydia trachomatis" *Infect. Immun.* **66** 1607-1612

Swanson AF and Kuo CC (1994) "Binding of the glycan of the major outer membrane protein of Chlamydia trachomatis to HeLa cells" *Infect. Immun.* **62** 24-28

Taraktchoglou M, Pacey AA, Turnbull JE, and Eley A (2001) "Infectivity of Chlamydia trachomatis serovar LGV but not E is dependent on host cell heparan sulfate" *Infect. Immun.* **69** 968-976

Vanrompay D, Cox E, Mast J, Goddeeris B, and Volckaert G (1998) "High-level expression of Chlamydia psittaci major outer membrane protein in COS cells and in skeletal muscles of turkeys" *Infect. Immun.* **66** 5494-5500

Ward ME (1995) "The immunobiology and immunopathology of chlamydial infections" *APMIS* **103** 769-796

Watson MW, Lambden PR, Everson JS, and Clarke IN (1994) "Immunoreactivity of the 60 kDa cysteine-rich proteins of Chlamydia trachomatis, Chlamydia psittaci and Chlamydia pneumoniae expressed in Escherichia coli" *Microbiology* **140 (Pt 8)** 2003-2011

Watson MW, Lambden PR, Ward ME, and Clarke IN (1989) "Chlamydia trachomatis 60 kDa cysteine rich outer membrane protein: sequence homology between trachoma and LGV biovars" *FEMS Microbiol. Lett.* **53** 293-297

Wuppermann FN, Hegemann JH, and Jantos CA (2001) "Heparan sulfate-like glycosaminoglycan is a cellular receptor for Chlamydia pneumoniae" *J. Infect. Dis.* **184** 181-187

Wyllie S, Ashley RH, Longbottom D, and Herring AJ (1998) "The major outer membrane protein of Chlamydia psittaci functions as a porin-like ion channel" *Infect. Immun.* **66** 5202-5207

Wyllie S, Longbottom D, Herring AJ, and Ashley RH (1999) "Single channel analysis of recombinant major outer membrane protein porins from Chlamydia psittaci and Chlamydia pneumoniae" *FEBS Lett.* **445** 192-196

Yuan Y, Zhang YX, Watkins NG, and Caldwell HD (1989) "Nucleotide and deduced amino acid sequences for the four variable domains of the major outer membrane proteins of the 15 Chlamydia trachomatis serovars" *Infect. Immun.* **57** 1040-1049

Zhang D, Yang X, Berry J, Shen C, McClarty G, and Brunham RC (1997) "DNA vaccination with the major outer-membrane protein gene induces acquired immunity to Chlamydia trachomatis (mouse pneumonitis) infection" *J. Infect. Dis.* **176** 1035-1040

Zhang JP and Stephens RS (1992) "Mechanism of C. trachomatis attachment to eukaryotic host cells" *Cell* **69** 861-869

Zhang YX, Stewart S, Joseph T, Taylor HR, and Caldwell HD (1987) "Protective monoclonal antibodies recognize epitopes located on the major outer membrane protein of Chlamydia trachomatis" *J. Immunol.* **138** 575-581

Appendix A: Secondary structure prediction raw data

A.1 Diederichs topology prediction data

residue #	z coordinate	residue #	z coordinate	residue #	z coordinate	residue #	z coordinate
1	0.135	41	0.369	81	0.666	121	0.417
2	0.141	42	0.352	82	0.652	122	0.444
3	0.168	43	0.365	83	0.665	123	0.415
4	0.152	44	0.353	84	0.551	124	0.414
5	0.160	45	0.308	85	0.594	125	0.459
6	0.209	46	0.332	86	0.673	126	0.400
7	0.293	47	0.325	87	0.711	127	0.377
8	0.590	48	0.342	88	0.742	128	0.415
9	0.584	49	0.321	89	0.658	129	0.339
10	0.449	50	0.354	90	0.613	130	0.252
11	0.344	51	0.390	91	0.560	131	0.248
12	0.319	52	0.438	92	0.546	132	0.370
13	0.285	53	0.388	93	0.492	133	0.335
14	0.232	54	0.469	94	0.510	134	0.358
15	0.228	55	0.466	95	0.617	135	0.381
16	0.264	56	0.492	96	0.581	136	0.403
17	0.258	57	0.475	97	0.576	137	0.480
18	0.274	58	0.438	98	0.490	138	0.447
19	0.363	59	0.500	99	0.523	139	0.520
20	0.439	60	0.499	100	0.474	140	0.476
21	0.548	61	0.425	101	0.341	141	0.638
22	0.586	62	0.503	102	0.349	142	0.696
23	0.610	63	0.453	103	0.357	143	0.709
24	0.703	64	0.568	104	0.360	144	0.749
25	0.725	65	0.628	105	0.382	145	0.758
26	0.832	66	0.641	106	0.409	146	0.706
27	0.860	67	0.703	107	0.302	147	0.563
28	0.859	68	0.728	108	0.385	148	0.504
29	0.863	69	0.747	109	0.355	149	0.454
30	0.857	70	0.760	110	0.256	150	0.419
31	0.843	71	0.709	111	0.297	151	0.437
32	0.800	72	0.699	112	0.280	152	0.417
33	0.706	73	0.606	113	0.327	153	0.454
34	0.568	74	0.560	114	0.262	154	0.448
35	0.495	75	0.580	115	0.313	155	0.429
36	0.371	76	0.602	116	0.424	156	0.387
37	0.341	77	0.565	117	0.311	157	0.402
38	0.424	78	0.685	118	0.337	158	0.376
39	0.399	79	0.672	119	0.392	159	0.452
40	0.461	80	0.617	120	0.394	160	0.400

residue #	z coordinate	residue #	z coordinate	residue #	z coordinate	residue #	z coordinate
161	0.370	201	0.536	241	0.766	281	0.677
162	0.357	202	0.528	242	0.612	282	0.579
163	0.361	203	0.519	243	0.589	283	0.580
164	0.345	204	0.518	244	0.643	284	0.547
165	0.289	205	0.511	245	0.672	285	0.494
166	0.258	206	0.423	246	0.578	286	0.470
167	0.206	207	0.406	247	0.561	287	0.473
168	0.224	208	0.443	248	0.453	288	0.492
169	0.235	209	0.458	249	0.387	289	0.424
170	0.251	210	0.531	250	0.344	290	0.437
171	0.338	211	0.554	251	0.296	291	0.417
172	0.363	212	0.518	252	0.286	292	0.413
173	0.360	213	0.544	253	0.268	293	0.363
174	0.290	214	0.558	254	0.240	294	0.343
175	0.321	215	0.500	255	0.198	295	0.347
176	0.359	216	0.454	256	0.221	296	0.358
177	0.430	217	0.471	257	0.259	297	0.380
178	0.491	218	0.451	258	0.234	298	0.413
179	0.563	219	0.475	259	0.225	299	0.548
180	0.546	220	0.434	260	0.211	300	0.558
181	0.514	221	0.492	261	0.235	301	0.471
182	0.517	222	0.506	262	0.238	302	0.492
183	0.452	223	0.430	263	0.257	303	0.534
184	0.344	224	0.414	264	0.248	304	0.613
185	0.373	225	0.392	265	0.250	305	0.676
186	0.354	226	0.350	266	0.263	306	0.664
187	0.440	227	0.362	267	0.334	307	0.659
188	0.543	228	0.426	268	0.314	308	0.631
189	0.607	229	0.490	269	0.349	309	0.729
190	0.489	230	0.484	270	0.387	310	0.739
191	0.502	231	0.596	271	0.438	311	0.644
192	0.436	232	0.591	272	0.404	312	0.600
193	0.405	233	0.632	273	0.452	313	0.535
194	0.471	234	0.640	274	0.474	314	0.476
195	0.606	235	0.720	275	0.497	315	0.446
196	0.543	236	0.648	276	0.523	316	0.406
197	0.629	237	0.709	277	0.510	317	0.322
198	0.625	238	0.775	278	0.682	318	0.308
199	0.541	239	0.763	279	0.655	319	0.277
200	0.556	240	0.773	280	0.637	320	0.323

residue #	z coordinate	residue #	z coordinate	residue #	z coordinate	residue #	z coordinate
321	0.322	341	0.488	351	0.446	361	0.518
322	0.299	342	0.508	352	0.494	362	0.544
323	0.335	343	0.469	353	0.409	363	0.542
324	0.320	344	0.443	354	0.387	364	0.503
325	0.259	345	0.453	355	0.375	365	0.477
326	0.251	346	0.480	356	0.389	366	0.491
327	0.301	347	0.484	357	0.427	367	0.378
328	0.334	348	0.452	358	0.418	368	0.334
329	0.476	349	0.472	359	0.462	369	0.276
330	0.564	350	0.423	360	0.526	370	0.260
331	0.591					371	0.221
332	0.631						
333	0.632						
334	0.654						
335	0.552						
336	0.532						
337	0.519						
338	0.526						
339	0.510						
340	0.523						

A.2 TMBETA prediction data

Residue #	Amino acid	Probability	Beta strand?	Residue #	Amino acid	Probability	Beta strand?
1	L	0	no	21	F	0.026523	no
2	P	0	no	22	G	0	no
3	V	0	no	23	G	0	no
4	G	0	no	24	D	0	no
5	N	0	no	25	P	0	no
6	P	0	no	26	C	0	no
7	A	0	no	27	D	0	no
8	E	0	no	28	P	0	no
9	P	0.000001	no	29	C	0	no
10	S	0.043363	no	30	A	0	no
11	L	0.701077	yes	31	T	0	no
12	M	0.97126	yes	32	W	0	no
13	I	0.999983	yes	33	C	0	no
14	D	0.999075	yes	34	D	0	no
15	G	0.999579	yes	35	A	0.000001	no
16	I	0.999923	yes	36	I	0.891532	yes
17	L	1	yes	37	S	0.627244	yes
18	W	1	yes	38	M	1	yes
19	E	0.999999	yes	39	R	1	yes
20	G	0.999267	yes	40	V	1	yes

Residue #	Amino acid	Probability	Beta strand?	Residue #	Amino acid	Probability	Beta strand?
41	G	0.999902	yes	91	M	0.999931	yes
42	Y	0.996885	yes	92	Q	0.999981	yes
43	Y	1	yes	93	D	0.273308	no
44	G	0.999999	yes	94	A	0.000603	no
45	D	0.99824	yes	95	E	0.028411	no
46	F	0.996786	yes	96	M	0.745253	yes
47	V	0.999997	yes	97	F	0.996383	yes
48	F	0.999591	yes	98	T	0.987616	yes
49	D	1	yes	99	N	0	no
50	R	0.99999	yes	100	A	0.000098	no
51	V	0.999995	yes	101	A	0	no
52	L	0.977549	yes	102	C	0	no
53	K	0.999989	yes	103	M	0.000001	no
54	T	0.999997	yes	104	A	0.001068	no
55	D	0.662277	yes	105	L	0.999791	yes
56	V	0.000002	no	106	N	0.999967	yes
57	N	0.024001	no	107	I	1	yes
58	K	0.999735	yes	108	W	0.999919	yes
59	E	0.999972	yes	109	D	0.983935	yes
60	F	0.999938	yes	110	R	0.865765	yes
61	Q	0.978604	yes	111	F	0.999919	yes
62	M	0.990672	yes	112	D	0.010468	no
63	G	0.003673	no	113	V	0	no
64	A	0.000001	no	114	F	0.020018	no
65	K	0.000001	no	115	C	0.008333	no
66	P	0.000003	no	116	T	0.000003	no
67	T	0.000002	no	117	L	0.000143	no
68	T	0.000035	no	118	G	0.00046	no
69	D	0	no	119	A	0.338433	no
70	T	0	no	120	T	0.994593	yes
71	G	0	no	121	S	0.991135	yes
72	N	0.000001	no	122	G	0.991119	yes
73	S	0	no	123	Y	0.997532	yes
74	A	0	no	124	L	0.993318	yes
75	A	0	no	125	K	0.994921	yes
76	P	0	no	126	G	0.055222	no
77	S	0.000036	no	127	N	0.000447	no
78	T	0.06596	no	128	S	0.054853	no
79	L	0.09359	no	129	A	0.143232	no
80	T	0.961543	yes	130	S	0.120573	no
81	A	0.379301	no	131	F	0.968284	yes
82	R	0.000004	no	132	N	0.999681	yes
83	E	0	no	133	L	0.999977	yes
84	N	0	no	134	V	1	yes
85	P	0.000004	no	135	G	1	yes
86	A	0.000003	no	136	L	0.999939	yes
87	Y	0.000064	no	137	F	0.997708	yes
88	G	0.999995	yes	138	G	0.112499	no
89	R	1	yes	139	D	0	no
90	H	0.999914	yes	140	N	0.000001	no

Residue #	Amino acid	Probability	Beta strand?	Residue #	Amino acid	Probability	Beta strand?
141	E	0.000009	no	191	F	0.999965	yes
142	N	0.000001	no	192	Q	0.99984	yes
143	Q	0.135274	no	193	Y	0.98215	yes
144	K	0.994528	yes	194	A	0.999092	yes
145	T	0.998401	yes	195	Q	0.000033	no
146	V	0.852921	yes	196	S	0.000018	no
147	K	0.947027	yes	197	K	0.00001	no
148	A	0.991556	yes	198	P	0.000018	no
149	E	0.009412	no	199	K	0.000076	no
150	S	0.000263	no	200	V	0.932613	yes
151	V	0.001216	no	201	E	0.203022	no
152	P	0.00007	no	202	E	0.99887	yes
153	N	0.00001	no	203	L	0.999967	yes
154	M	0.912405	yes	204	N	0.794554	yes
155	S	0.000717	no	205	V	0.205572	no
156	F	0.553716	yes	206	L	0.000007	no
157	D	0.065684	no	207	C	0	no
158	Q	0.0132	no	208	N	0	no
159	S	0.999801	yes	209	A	0	no
160	V	0.999865	yes	210	A	0.000254	no
161	V	1	yes	211	E	0.600644	yes
162	E	1	yes	212	F	0.997304	yes
163	L	0.999998	yes	213	T	0.999775	yes
164	Y	0.999991	yes	214	I	0.000089	no
165	T	0.999748	yes	215	N	0.000431	no
166	D	0.970545	yes	216	K	0.000016	no
167	T	0.051384	no	217	P	0.00007	no
168	T	0.999054	yes	218	K	0.00206	no
169	F	0.999225	yes	219	G	0.376327	no
170	A	1	yes	220	Y	0.769769	yes
171	W	0.999981	yes	221	V	0.998698	yes
172	S	0.999992	yes	222	G	0.999953	yes
173	V	0.999996	yes	223	K	0.718313	yes
174	G	0.995633	yes	224	E	0.951607	yes
175	A	0.091799	no	225	F	0.036892	no
176	R	0.999725	yes	226	P	0.088799	no
177	A	1	yes	227	L	0.003714	no
178	A	0.997113	yes	228	D	0.118682	no
179	L	0.969085	yes	229	L	0.000087	no
180	W	0.256853	no	230	T	0.978269	yes
181	E	0	no	231	A	0.001874	no
182	C	0	no	232	G	0.00367	no
183	G	0	no	233	T	0.000397	no
184	C	0	no	234	D	0.000084	no
185	A	0	no	235	A	0.000005	no
186	T	0.000062	no	236	A	0.000004	no
187	L	0.000533	no	237	T	0.001412	no
188	G	0.959657	yes	238	G	0.00045	no
189	A	0.998387	yes	239	T	0.022668	no
190	S	0.999348	yes	240	K	0.00203	no

Residue #	Amino acid	Probability	Beta strand?	Residue #	Amino acid	Probability	Beta strand?
241	D	0.029414	no	291	I	0.953313	yes
242	A	0.000004	no	292	F	0.108673	no
243	S	0.411113	yes	293	D	0.114839	no
244	I	0.096225	no	294	T	0.080032	no
245	D	0.999239	yes	295	T	0.998553	yes
246	Y	0.999798	yes	296	T	0.00001	no
247	H	1	yes	297	L	0.000001	no
248	E	1	yes	298	N	0.000004	no
249	W	1	yes	299	P	0.002574	no
250	Q	0.99976	yes	300	T	0.008378	no
251	A	0.99983	yes	301	I	0.012496	no
252	S	0.991764	yes	302	A	0	no
253	L	0.991586	yes	303	G	0.001463	no
254	A	0.999944	yes	304	A	0.000295	no
255	L	0.999999	yes	305	G	0.006524	no
256	S	0.999999	yes	306	D	0.01631	no
257	Y	1	yes	307	V	0.010385	no
258	R	1	yes	308	K	0.941148	yes
259	L	0.999999	yes	309	T	0.612431	yes
260	N	0.998561	yes	310	G	0.254807	no
261	M	0.226026	no	311	A	0.000682	no
262	F	0.99703	yes	312	E	0.066741	no
263	T	0.993002	yes	313	G	0.957084	yes
264	P	0.041754	no	314	Q	0.08721	no
265	Y	0.991069	yes	315	L	0.12233	no
266	I	0.998817	yes	316	G	0.293475	no
267	G	0.999516	yes	317	D	0.325395	no
268	V	0.999999	yes	318	T	0.194999	no
269	K	1	yes	319	M	0.99997	yes
270	W	1	yes	320	Q	0.99999	yes
271	S	0.997002	yes	321	I	1	yes
272	R	0.999967	yes	322	V	0.999964	yes
273	A	0.996683	yes	323	S	0.999996	yes
274	S	0.00094	no	324	L	0.999999	yes
275	F	0.001971	no	325	Q	0.999993	yes
276	D	0.000923	no	326	L	0.998562	yes
277	A	0	no	327	N	0.999753	yes
278	D	0.352213	no	328	K	0.658104	yes
279	T	0.047103	no	329	M	0.999978	yes
280	I	0.998673	yes	330	K	0.999848	yes
281	R	0.992073	yes	331	S	0.950136	yes
282	I	0.852299	yes	332	R	0.000009	no
283	A	0.000093	no	333	K	0.000078	no
284	Q	0.000001	no	334	S	0	no
285	P	0.000001	no	335	C	0	no
286	K	0	no	336	G	0	no
287	S	0.000015	no	337	I	0.000299	no
288	A	0.00003	no	338	A	0.105824	no
289	T	0.999841	yes	339	V	0.986488	yes
290	A	0.97466	yes	340	G	0.965498	yes

Residue #	Amino acid	Probability	Beta strand?	Residue #	Amino acid	Probability	Beta strand?
341	T	0.999994	yes	356	R	0.999985	yes
342	T	0.999999	yes	357	L	0.999756	yes
343	I	0.999983	yes	358	I	0.995516	yes
344	V	0.166668	no	359	D	0.081346	no
345	D	0.393379	no	360	E	0.021016	no
346	A	0.000314	no	361	R	0.000591	no
347	D	0.007414	no	362	A	0.999998	yes
348	K	0.001813	no	363	A	0.99516	yes
349	Y	0.999689	yes	364	H	0.99883	yes
350	A	0.999982	yes	365	V	0.997226	yes
351	V	0.999998	yes	366	N	1	yes
352	T	0.999983	yes	367	A	0.999974	yes
353	V	0.999999	yes	368	Q	0.999792	yes
354	E	0.999999	yes	369	F	0.999992	yes
355	T	1	yes	370	R	0.979395	yes
				371	F	0.006537	no

Predicted segments:

Segment 1 : L11 to G20
Segment 2 : I36 to D55
Segment 3 : L105 to F111
Segment 4 : T120 to K125
Segment 5 : F131 to F137
Segment 6 : S159 to D166
Segment 7 : T168 to G174
Segment 8 : G188 to A194
Segment 9 : D245 to N260
Segment 10 : Y265 to A273
Segment 11 : M319 to S331
Segment 12 : Y349 to I358
Segment 13 : A362 to R370

A.3 B2TMPRED prediction data

Predicted segments:

	Starts	-	Ends
TM (1)	38	-	45
TM (2)	47	-	54
TM (3)	57	-	62
TM (4)	89	-	94

TM (5)	120	-	125
TM (6)	128	-	137
TM (7)	159	-	165
TM (8)	167	-	177
TM (9)	188	-	194
TM (10)	247	-	262
TM (11)	266	-	275
TM (12)	319	-	330
TM (13)	348	-	357
TM (14)	361	-	370

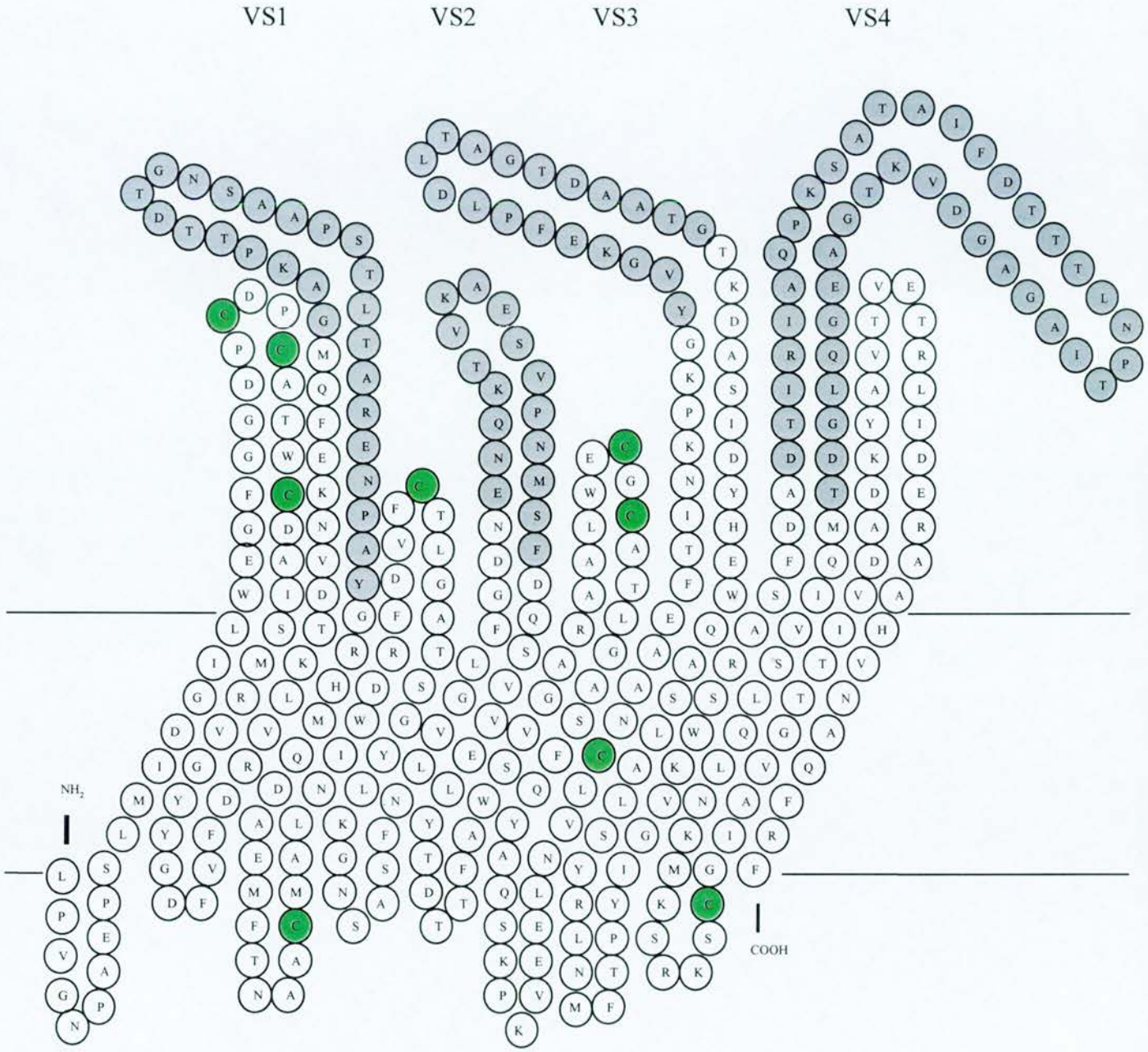
Appendix B: PCR primers and conditions

Primer name	Sequence	Melting temp(°C)	Restriction Site	Annealing temp(°C)
TR 5's	ACG AAT TCA GAT GTC **CAT ATG** CTG CCT GTG GGG AAT C	71.7	*Nde* I	65
TR 5'as	GAT TCC CCA CAG GCA G**CA T AT G**GA CAT CTG AAT TCG T	71.7	"	"
TR r-nde s	TGC TTA CGG CCG ACA CAT GCA GGA TGC TGA G	72.1	remove *Nde* I	65
TR r-nde as	CTC AGC ATC CTG CAT GTG TCG GCC GTA AGC A	72.1	"	"
MU 5's	CTC CTT GCA T**CA TAT G**CC TGT GGG GAA TC	68.1	*Nde* I	57
MU 3'as	CGA TCG AAA CC**G GAT CC**A GAA ACA CAG	66.5	*Bam* H-I	60
MU r-nde s	CGC TTA TGG CAA GCA CAT GCA AGA TGC AG	68.1	remove *Nde* I	60
MU r-nde as	CTG CAT CTT GCA TGT GCT TGC CAT AAG CG	68.1	"	57
			for self-priming s tep 8 @ 60, 25 @ 57	
PB 5's	CGT ATC TAC ATC ATT AGA C**CA TAT** GCC TGC	65.4	*Nde* I	60
PB 3'as	GCT ATG CAT CAG **CCA TGG** AGT ATA GTC	65.0	*Nco* I	"
TR C26A s	TTC GGC GGA GAT CCT **GCC** GAT CCT TGC GCC ACT	75.7	-	65
TR C26A as	AGT GGC GCA AGG ATC GGC AGG ATC TCC GCC GAA	75.7	-	"
TR C29A s	GAT CCT GCC GAT CCT **GCC** GCC ACT TGG TGT GAC	75.7	-	65
TR C29A as	GTC ACA CCA AGT GGC GGC AGG ATC GGC AGG ATC	75.7	-	"
TR C33A s	CCT GCC GCC ACT TGG **GCT** GAC GCT ATC AGC ATG	75.7	-	65
TR C33A as	CAT GCT GAT AGC GTC AGC CCA AGT GGC GGC AGG	75.7	-	"
TR C102A s	TTT ACA AAT GCC GCT **GCG** ATG GCA TTG AAT ATT	65.8	-	55
TR C102A as	AAT ATT CAA TGC CAT CGC AGC GGC ATT TGT AAA	65.8	-	"
TR C115A s	CGT TTT GAT GTA TTC **GCT** ACA TTA GGA GCC ACC	68.2	-	57
TR C115A as	GGT GGC TCC TAA TGT AGC GAA TAC ATC AAA ACG	68.2	-	"
TR C182A s	GCA GCT TTG TGG GAA **GCT** GGA TGT GCA ACT TTA	69.5	-	57
TR C182A as	TAA AGT TGC ACA TCC AGC TTC CCA CAA AGC TGC	69.5	-	"
TR C184A s	TTG TGG GAA GCT GGA **GCT** GCA ACT TTA GGA GCT	70.7	-	65
TR C184A as	AGC TCC TAA AGT TGC AGC TCC AGC TTC CCA CAA	70.7	-	"
TR C207A s	GAA TTA AAC GTT CTC **GCC** AAT GCA GCA GAG TTT	67.0	-	60
TR C207A as	AAA CTC TGC TGC ATT GGC GAG AAC GTT TAA TTC	67.0	-	"
TR C335A s	AAA TCT AGA AAA TCT **GCC** GGT ATT GCA GTA GGA	65.8	-	60
TR C335A as	TCC TAC TGC AAT ACC GGC AGA TTT TCT AGA TTT	65.8	-	"

Primer name	Sequence	Melting temp(°C)	Restriction Site	Annealing temp(°C)
nl-TR *S*s1	GAG TTC TGC TTC CTC CTT GCA AGC TCT GCC TGT GGG GAA TCC TG	77.8	-	65
nl-TR *3*as	GAC GCG TTA CGT ATC GGA TCC AGA ATT CGT GAT **CCATG**	72.7	*Nco* I	
nl-TR *S*s2	GAA ATC GGT ATT AGT ATT TGC CGC TTC GAG TTC TGC TTC CTC CTT G	73.9	-	5 @ 50, 13 @ 60
nl-TR *S*s3	CAG AGG TA**C ATA TG**A AAA AAC TCT TGA AAT CGG TAT TAG	66.3	*Nde* I	5 @ 52, 13 @ 57
ot-TR *S*s1	TAT CGC GAT CAG CTC TTT TGC GAT GCT GCC TGT GGG GAA TCC	76.3	-	65
ot-TR *3*as	GAC GCG TTA CGT ATC GGA TCC AGA ATT CGT GAT **CCATG**	72.7	*Nco* I	
ot-TR *S*s2	CTG GGA ATA GTC CTG ACA ACC CCT ATC GCG ATC AGC TC	74.9	-	5 @ 50, 13 @ 60
ot-TR *S*s3	GAT CGA **CAT ATG** CGG GCG AAA CTC CTG GGA ATA GTC C	73.9	*Nde* I	5 @ 50, 13 @ 57
nl-PB *S*s	GCC CTT GAC **CAT ATG** AGC AGC AAG CTA G	68.1	*Nde* I	60
nl-PB *3*as	GCT ATG CAT CAG **CCATGG** AGT ATA GTC	65.0	*Nco* I	"
ot-PB *S*'s1	TAT CGC GAT CAG CTC TTT TGC GAT GCC TGC AGG GAA TCC G	75.6	-	60
ot-PB 3'as	GCT ATG CAT CAG **CCATGG** AGT ATA GTC	65.0	*Nco* I	"
nl-MU *S*s	GAG GTA **CAT ATG** AAA AAA CTC TTG AAA TCG G	62.9	*Nde* I	57
nl-MU 3'as	CGA TCG AAA CC**G GAT CC**A GAA ACA CAG	66.5	*BamH* I	"
nl-AB *S*s1	CGG GTT CCG CTC TCT CCT TTA CAA GCC TTG CCT GTA GGG AAC CCA G	79.5	-	55
nl-AB *3*as	CAT G**CC ATG GC**A GAG ATT CCT AGG TTC TGA TAG CGG GAC AA	74.4	*Nco* I	
nl-AB *S*s2	GAA ATC GGC ATT ATT GTT TGC CGC TAC GGG TTC CGC TCT CTC C	76.1	-	55
nl-AB *S*s3	GCG AGG TG**C ATA TG**A AAA AAC TCT TGA AAT CGG CAT TAT TGT TTG C	71.2	*Nde* I	5 @ 52, 13 @ 57
ot-AB *S*s1	TAT CGC GAT CAG CTC TTT TGC GTT GCC TGT GGG GAA CCC AGC	77.3	-	60
ot-AB *3*as	GCG AGG TG**C ATA TG**A AAA AAC TCT TGA AAT CGG CAT TAT TGT TTG C	71.2	*Nco* I	

Primer name	Sequence	Melting temp(°C)	Restriction Site	Annealing temp(°C)
TRvs1 *S*s	ATA AAG AAT TTC AG**A CCG GT**G CCA AGC CTA CA	66.9	*Age*-I	60
TRvs1 *S*as	TGT AGG CTT GGC **ACC GGT** CTG AAA TTC TTT AT	66.9	"	"
TRvs1 *3*s1	AGA GAG AAT CCT GCT ACC GGC CGA CAC ATG CAG	72.0	-	65
TRvs1 *3*as1	CTG CAT GTG TCG GCC GGT AGC AGG ATT CTC TCT	72.0	-	"
TRvs1 *3*s2	GAG AAT CCT GCT **ACC GGT** CGA CAT ATG CAG GAT G	71.9	*Age*-I	65
TRvs1 *3*as2	CAT GCT GCA TAT GTC G**AC CGG T**AG CAG GAT TCT C	71.9	"	"
TRvs2 *3*s	GTT TGG AGA TAA **TGA TCA** TCA AAA AAC GGT CA	63.1	*Bcl*-I	58
TRvs2 *3*as	TGA CCG TTT TTT GA**T GAT CA**T TAT CTC CAA AC	63.1	"	"
TRvs3 *S*s1	ATT AAT AAA CCT AAA GAC TAT GTA GGT AAG GAG	62.0	-	55
TRvs3 *S*as1	CTC CTT ACC TAC ATA GTC TTT AGG TTT ATT AAT	62.0	-	"
TRvs3 *S*s2	AAT AAA CCT AAA **GAC GTC** GTA GGT AAG GAG TTT CC	67.1	*Asr*-II	60
TRvs3 *S*as2	GGA AAC TCC TTA CCT AC**G ACG TC**T TTA GGT TTA TT	67.1	"	"
TRvs3 *3*s1	ACA GAT GCT GCG ACA GAC ACT AAG GAT GCC TCT	70.7	-	65
TRvs3 *3*as1	AGA GGC ATC CTT AGT GTC TGT CGC AGC ATC TGT	70.7	-	"
TRvs3 *3*s2	GAT GCT GCG ACA **GAC GTC** AAG GAT GCC TCT ATT G	71.9	*Asr*-II	65
TRvs3 *3*as2	CAA TAG AGG CAT CCTT **GA CGT C**TG TCG CAG CAT C	71.9	"	"
TRvs4 *S*s	CGA GCA AGC TTT **GAC GTC** GAT ACG ATT CGT ATA GC	70.6	*Asr*-II	65
TRvs4 *S*as	GCT ATA CGA ATC GTA TC**G ACG TC**A AAG CTT GCT CG	70.6	"	"
TRvs4 *3*s	GGT CAG CTC GGA **GAC GTC** ATG CAA ATC GTT TCC	72.0	*Asr*-II	65
TRvs4 *3*as	GGA AAC GAT TTG CAT **GAC GTC** TCC GAG CTG ACC	72.0	"	"

Appendix C: *C. trachomatis* MOMP Internal mutagenesis



Green circles are cysteine residues. They are mutated to alanine residues in M9-MOMP.

Grey circles are portions of VS domains digested out.

Appendix D: FoldIt reagents and data

D.1 Combinations of reagents in FoldIt screen

	Buffer	Salt	Osmolyte	Chaotrope	Cation chelator	Polar additive	Nonpolar additive
1	50mM Tris pH 8.2	250mM NaCl, 10mM KCl	none	none	1mM EDTA	none	none
2	50mM MES pH 6.5	10mM NaCl, 0.4mM KCl	0.03% PEG 3350	500mM GuHCl	2mM MgCl, 2mM CaCl	none	none
3	50mM MES pH 6.5	10mM NaCl, 0.4mM KCl	none	500mM GuHCl	1mM EDTA	500mM L-arginine	400mM sucrose
4	50mM Tris pH 8.2	250mM NaCl, 10mM KCl	0.03% PEG 3350	none	2mM MgCl, 2mM CaCl	500mM L-arginine	400mM sucrose
5	50mM MES pH 6.5	250mM NaCl, 10mM KCl	0.03% PEG 3350	none	2mM MgCl, 2mM CaCl	none	400mM sucrose
6	50mM Tris pH 8.2	10mM NaCl, 0.4mM KCl	none	500mM GuHCl	1mM EDTA	none	400mM sucrose
7	50mM Tris pH 8.2	10mM NaCl, 0.4mM KCl	0.03% PEG 3350	500mM GuHCl	2mM MgCl, 2mM CaCl	500mM L-arginine	none
8	50mM MES pH 6.5	250mM NaCl, 10mM KCl	none	none	1mM EDTA	500mM L-arginine	none
9	50mM MES pH 6.5	250mM NaCl, 10mM KCl	none	500mM GuHCl	2mM MgCl, 2mM CaCl	none	400mM sucrose
10	50mM Tris pH 8.2	10mM NaCl, 0.4mM KCl	0.03% PEG 3350	none	1mM EDTA	none	400mM sucrose
11	50mM Tris pH 8.2	10mM NaCl, 0.4mM KCl	none	none	2mM MgCl, 2mM CaCl	500mM L-arginine	none
12	50mM MES pH 6.5	250mM NaCl, 10mM KCl	0.03% PEG 3350	500mM GuHCl	1mM EDTA	500mM L-arginine	none
13	50mM Tris pH 8.2	250mM NaCl, 10mM KCl	0.03% PEG 3350	500mM GuHCl	1mM EDTA	none	none
14	50mM MES pH 6.5	10mM NaCl, 0.4mM KCl	none	none	2mM MgCl, 2mM CaCl	none	none
15	50mM MES pH 6.5	10mM NaCl, 0.4mM KCl	0.03% PEG 3350	none	1mM EDTA	500mM L-arginine	400mM sucrose
16	50mM Tris pH 8.2	250mM NaCl, 10mM KCl	none	500mM GuHCl	2mM MgCl, 2mM CaCl	500mM L-arginine	400mM sucrose

D.2 Spectroscopic data

Foldit Reagent Combination	Absorbance at 400nm	
	MP	M9
1	0.031	0.007
2	0.107	0.240
3	0.020	0.070
4	0.008	0.008
5	0.841	0.370
6	0.020	0.012
7	0.017	0.014
8	0.590	0.324
9	0.081	0.100
10	0.022	0.023
11	0.013	0.015
12	0.059	0.096
13	0.033	0.015
14	0.503	0.149
15	0.440	0.195
16	0.019	0.023