IN VITRO AND *IN SILICO* STUDIES OF A NON-COMMENSAL SOIL BACTERIUM AND ITS INTERACTIONS WITH A MAMMALIAN HOST

stine anita indrelid

Norwegian University of Life Sciences
Department of Chemistry, Biotechnology and Food Science (IKBM)
Master Thesis 60 credits 2012

Sammendrag

Methylococcus capsulatus (Bath) er en Gram-negativ, metanotrof bakterie. Til tross for at den aldri har vært beskrevet å være en del av tarmflora har *M. capsulatus* (Bath) vist seg å ha positive effekter ved inflammatoriske tilstander i vertebrater. *M. capsulatus* (Bath) hindrer utvikling av soya-indusert enteritt i Atlanterhavslaks og har vist seg å ha profylaktisk effekt ved dekstran natriumsulfat-indusert kolitt hos mus. *In vitro* studier har vist at *M. capsulatus* (Bath) kan binde seg spesifikt til humane $CD14^+$ monocyttderiverte dendrittiske celler (DC). Denne ikke-kommensale, jordlevende bakterien ser altså ut til å kunne interagere med humane immunceller, men også å ha immunmodulerende effekter in vivo. Mekanismene bak disse effektene er imidlertid enda ikke kjent. Målet for dette arbeidet har vært å identifisere strukturer og egenskaper ved *M. capsulatus* (Bath) som gjør den i stand til å interagere med en vert og påvirke immunresponser. Proteomet til *M. capsulatus* (Bath) ble analysert ved bioinformatiske metoder. Subcellulær lokalitet for samtlige proteiner ble predikert og bakteriens sekretom ble definert. Interaksjon med DC ble studert og utskilte proteiner ble identifisert ved eksperimentelle metoder. Arbeidet ledet til identifisering av adhesiner med putative bindingsseter for adhesjonsmolekyler på vertebratimmunceller, proteiner med mulige roller i invasjon av vertsceller og proteiner med potensiell effekt på viktige immunregulatoriske signalveier i vertebrater.

Abstract

Methylococcus capsulatus (Bath) is a Gram-negative, methanotroph bacterium. Although it has never been described to be part of vertebrate intestinal microbiota, *M. capsulatus* (Bath) has been demonstrated to have a positive effect on inflammatory conditions in vertebrates. *M. capsulatus* (Bath) abolishes soya induced enteritis in Atlantic salmon and to have prophylactic effect on dextran sodium sulphate induced colitis in mice. *In vitro* studies have shown that *M. capsulatus* (Bath) adheres specifically to human $CD14^+$ monocyte derived dendritic cells (DC). This soil dwelling, non-commensal bacteria thus appear to be able not only to interact with human key immune cells, but also to have immunomodulatory effects *in vivo*. The mechanisms behind the seen effects, however has not been determined so far.

The objective of this study has been to identify structures and properties of *M. capsulatus*(Bath) that can elucidate the relationship of *M. capsulatus* (Bath) with mammalian hosts, and the mechanisms by which it interacts with host cells and affects inflammatory response.

The sub cellular location for all *M. capsulatus* (Bath) proteins has been predicted and the secretome of *M. capsulatus* (Bath) defined by *in silico* analysis. Released proteins and interactions of *M. capsulatus* (Bath) with DC were examined by experimental approaches. This work has identified adhesins with putative binding sites for vertebrate cell surface adhesion molecules, and possible roles in host interactions and proteins with potential effect on key immune regulatory pathways

Acknowledgments

Working with this thesis has been great fun but also challenging at times. I would like to use the opportunity to thank all those who have been there to support me.
First of all I would like to thank Tor Lea who has read an unknown number of versions of my thesis with great patience always providing valuable advice. I don´t know how many times you have had to changed M. capsulatus to *M. capsulatus* (Bath) after my cut and paste episodes. There are 206 occurrences of the name in this document, and I think you have made yourself acquainted with most of them at some point. Thanks to Charlotte Kleiveland for reading and commenting on my even earlier versions and for doing so in the kindest possible manner. Your advice here and in the lab has always been of great help and highly appreciated. Thanks to Geir Mathiesen for always sharing your time and expertise when needed the most and for showing enthusiasm over bioinformatic peculiarities that strangely few others seem to value. Thanks to Lene Olsen Hult and Trine Eker Christoffersen for all help and support, and Charlotte, Lene and Trine for being great academic role models for me.
Thanks to Ellen Karlstrøm and Katarzyna Kuczkowska for excellent technical support and for making the lab a nice place to work. Thanks also to Morten Jacobsen for always taking an interest in my work and inspiring me to continue and Eirin and everyone else at the lab for making lab work enjoyable. I would like to acknowledge the invaluable help received from Elin Ørmen at the Imaging Center for taking the SEM pictures for my thesis and Magnus Arntzen at the Biotechnology center of Oslo for performing the LC-MS analysis.
Last, but not least I would like to thank Ask and Alma for being patient with me during late hours writing and Lars Martin for always making me look forward to coming home.

Table of Contents

Abbreviations

BioProteinTM	BP
Dextran sodium sulfate	DSS
CD14+ monocyte-derived dendritic cell	DC
Chaperone-usher pili	CU
Type IV pili	T4P
Inner membrane	IM
Outer membrane	OM
Lipopolysaccharide	LPS
Outer membrane protein A	OmpA
Filamentous hemagglutinin	FHA
Lipopolysaccharides	LPS
The type I secretion system	(T1SS)
The type II secretion system	(T2SS)
The type III secretion system	(T3SS)
The type V secretion system	(T5SS)
Signal peptide	SP
Signal peptidase	SPase
Signal peptidase I	SPaseI
Signal peptidase II	SPaseII
Twin-arginine translocation	Tat
Trans membrane helix	TMH
Outer membrane β-barrel	OM BB
Lipoprotein	LP
Macrophage migration inhibitory factor	MIF
IL-17 receptor	IL-17R
SEF/IL-17 receptor domain	SEFIR
‘Signaling molecule that associates with mouse pelle-like kinase domain’	SIMPL
Mammalian cell entry protein	Mce

1 Introduction

1.1 Background

Animal proteins have traditionally been used in fodder for fish, pets and livestock, but this is a costly resource and alternatives are being sought. Legumes, such as peas and soybean, have been tested as potential substitutes in salmon but show some detrimental health effects limiting their use. Another alternative to animal protein-based fodder is using microorganisms as protein source. Norferm produces BioProtein™ (BP), a single cell protein product in which a Gram-negative methanotroph bacterium, *Methylococcus capsulatus* (Bath) is the main ingredient. In 2011Romarheim et al. compared the effects of feeding Atlantic salmon (*Salmo salar*) solvent-extracted soybean meal (SBM), traditional fish meal, BP or a soybean meal combined with BP. The results were surprising: While salmon fed SBM developed what is known as soy bean-induced enteritis, inflammation was abolished in those fed a soybean meal with inclusion of BP. The inflammatory response and morphological changes seen in soybean-induced enteritis resemble those seen in human inflammatory bowel disease. These findings raised questions of whether *M. capsulatus* (Bath) may have a similar effect on inflammatory intestinal conditions in higher vertebrates.

Aiming to investigate if BP could have potential for therapeutic use in inflammatory bowel disease in mammals, mice were fed BP or a control diet. At day 8 dextran sodium sulfate (DSS) was delivered in the drinking water inducing progressive colitis in the mice. Mice fed the BP meal showed improved colitis-associated parameters compared to mice given the control diet as measured by body weight, colon length and epithelia integrity and reduced signs of inflammation as measured by levels of acute phase reactants, and neutrophil infiltration when compared to mice fed the control diet (Kleiveland et al. 2012b) In a control experiment BP was substituted with a *M. capsulatus* (Bath) bacterial meal without the supplementary bacteria found in BP. Mice fed the bacterial meal showed increase in bodyweight comparable with the control group. Histological examination showed the colon of these animals exhibited a typical normal structure while DSS colitis mice fed the control meal showed substantial signs of tissue damage. The results demonstrate that *M. capsulatus* (Bath) is the active anti-inflammatory principle in the BP preparation

In vitro studies later showed that when *M. capsulatus* (Bath) was co-cultivated with human $CD14^+$ monocyte-derived dendritic cells (DC), the bacteria displayed a striking ability to adhere to the DCs. Further investigations showed that bacteria were internalized in DCs upon

binding and triggered release of several inflammatory cytokines (Kleiveland et al., manuscript in preparation).

Taken together these results show that orally introduced *M. capsulatus* (Bath) affects the intestinal homeostasis of vertebrates *in vivo* and interacts with and stimulates immune cells *in vitro*. In an attempt to better understand these interactions this master thesis examines surface associated and possible immune active proteins of *M. capsulatus* (Bath). Two different approaches have been used: The secretome of *M. capsulatus* (Bath) has been defined and characterized using bioinformatics tools. Released proteins and interactions with dendritic cells have been examined by experimental proteomics.

1.2 Microbe host interactions

During evolution microbes have adapted to live in close association with multicellular hosts. The nature of the relationship between a host and colonizing microbes varies depending on type of microbe and immunological properties of the host. The association may be beneficial to both partners, detrimental to one partner or have potential for both harmful and beneficial effects for both partners. The host and the microbe must respond to each other accordingly.

1.2.1 Intestinal homeostasis, a dynamic balance between commensals, the epithelium and immune cells of the mucosa

Like the skin and epithelia of the respiratory tract, the surface of the gastrointestinal tract represents an interface between the organism and the environment and provides a possible entry point for microbes into the body. At the same time the intestinal tract is colonized by a startling 10^{13}-10^{14} bacteria many of which aids the host by breaking down indigestible food components to absorbable nutrients. The immune system must respond to and protect the intestines from invasion by pathogenic microorganisms yet remain tolerant to the commensal microbiota that is necessary for the gut to maintain its normal functions. The balance between these contrasting needs depends on maintaining mucosal and epithelial integrity and regulating pro-inflammatory signaling, innate and adaptive immune responses in an appropriate manner. Homeostasis is retained through a dynamic interaction and communication between the intestinal microbiota, the epithelium, and the mucosal immune cells. A failure of any of these partners in fulfilling its role can lead to chronic inflammatory pathology.

1.2.1.1 The intestinal epithelia and the mucus layer

The gastrointestinal tract is lined with a single layer of epithelial cells covered by a mucus layer. In line with the many contrasting requirements, the epithelial cell layer consists of many different cell types with different levels of specialization. The epithelium provides a physical barrier preventing entrance of pathogens and detrimental substances, and transport across the layer must be tightly regulated. At the same time the intestinal epithelium regulates nutrient uptake, water and electrolyte balance of the body and must therefore allow transport of such substances across the barrier. Tight junctions between epithelial cells are crucial regulators of paracellular transport. These protein complexes form continuous intercellular barriers between the cells. In addition to providing a physical barrier, epithelial cells secrete and transport several types of antimicrobial peptides (AMPs), enzymes and antibodies (e.g sIgA) from the basolateral to the luminal side of the intestine (Kunisawa & Kiyono 2012). The mucus layer also contributes to the barrier function of the gut. In the large bowel it consists of a dense, thin inner layer that is firmly attached to the epithelium and a looser, thick outer layer that is continuously shed and renewed. This outer layer provides a scaffold for antimicrobial peptides (AMPs) and immunoglobulins, particularly secretory IgA (sIgA) (McGuckin et al. 2011). Antigen-specific sIgA binds microbes and efficiently prevents intrusion to the inner mucus layer and epithelium by immune exclusion.

1.2.1.2 Commensals as regulators of homeostasis

The importance of maintaining a balanced community of commensals is demonstrated by the fact that germ-free mice do not develop normal lymphoid structures, are highly susceptible to infections and show altered immune as reviewed by (Arrieta & Finlay 2012). Another line of evidence is provided by studies of gut colonization in mice treated with antibiotics. Following treatment the mice are susceptible to infection by harmful bacteria and parasites that normally are not able to colonize the intestines in numbers large enough to pose any threat. The commensal microbiota not only provides colonization resistance, microbiota components appear to be able to stimulate protective immune responses, and following antibiotic treatment reintroduction of normal microbiota is capable of re-establishing pathogen clearance (Jarchum & Pamer 2011)

Microbiota-driven host defense mechanisms are typically activated by microbial components (e.g. lipopolysaccharide (LPS), peptidoglycans, flagellin) interacting with pattern recognizing receptors (PRRs) on epithelial cells and immune cells of the intestines. Interactions trigger

signaling events leading to various protective responses such as up-regulated production of AMPs, increased neutrophil bacterial killing, reduced apoptosis and proliferation of epithelial cells (Jarchum & Pamer 2011).

Gut microbiota may also affect the players and tools of the adaptive immune response. Ivanov et al. (2008) showed that commensal microbiota was necessary to induce Th17 cell differentiation in the small intestine lamina propria. Specifically, Gram-negative Cytophaga-Flavobacter-Bacteroidetes appeared to be correlated with the differentiation of Th17 cells. Th17cells develops from $CD4^+$ T cells as does regulatory T cells (Tregs). The developmental programs of the two T cell subsets are linked and the lamina propria Treg population is increased in germfree mice lacking the Th17 cells. Commensal bacteria may thus shape the composition of T cell subsets present in the gut and affect the balance between potentially proinflammatory T effector cells and anti-inflammatory regulatory T cells.

1.2.2 Adherence and host colonization

Bacteria that live in close relation with a host be it as commensals or pathogens have evolved unique strategies and structures specialized for such lifestyle. In order for the bacterium to colonize or infect a host, it must first overcome a multitude of barriers. Blinking and peristalsis, cell shedding and sneezing all contribute to epithelial colonization by microorganisms. The skin and the epithelial cells of the respiratory, urogenital and gastrointestinal tract provide physical obstacles. High acidity in the stomach, the mucous layers, and electrostatic repulsion may further enhance the defensive properties of such barriers. The strategies used to overcome these challenges varies from species to species and may be highly specialized, but some common themes exists (Finlay & Falkow 1997).

Adherence to host cells is commonly the initial step in colonization. By adhering to host cells the bacteria can avoid many of the above mentioned challenges. It is better prepared to resist mechanical forces and is properly positioned for entry into the host cell (internalization) or tissue. Equally important: Forming a close spatial interaction with the host allows signaling events between the bacteria and the host allowing the microorganism to suppress or modulate host responses. Adherence is mediated by structures on the bacterial surface, adhesins, interacting with host receptors.

An adherence factor is any factor that directly or indirectly contributes to cell binding, while an adhesin is the specific bacterial ligand that binds a host cell molecule or structure.

Typically adhesins bind host cell receptors like integrins or selectins, but other host cell elements, like extracellular matrix component glycoproteins or glycolipids, are also common targets for microbial adhesion. Adhesins are usually classified in two groups based on overall appearance: Fimbrial (filamentous) adhesins and afimbrial adhesins.

1.2.2.1 Fimbrial adhesins

By coming in close contact with host cells, bacteria may be detected by cell membrane receptors initiating host defensive mechanisms. Cell surfaces of both eukaryotic host cells and Gram-negative bacteria are typically negatively charged causing a mutual electrical repulsion. For these reasons bacteria have developed protruding surface structures that enable them to interact with host cells at a safe distance

Pili or fimbria are filamentous protein structures found on Gram-negative and less frequently on Gram-positive prokaryotes. Several types of pilus structures exist (Figure 1.1) and mediate an array of functions, like motility, DNA uptake and conjugative transfer, microfilm formation, signaling, secretion and adhesion. The term fimbria is used to describe pili involved in adhesion in order to distinguish these from pili involved in conjugation.

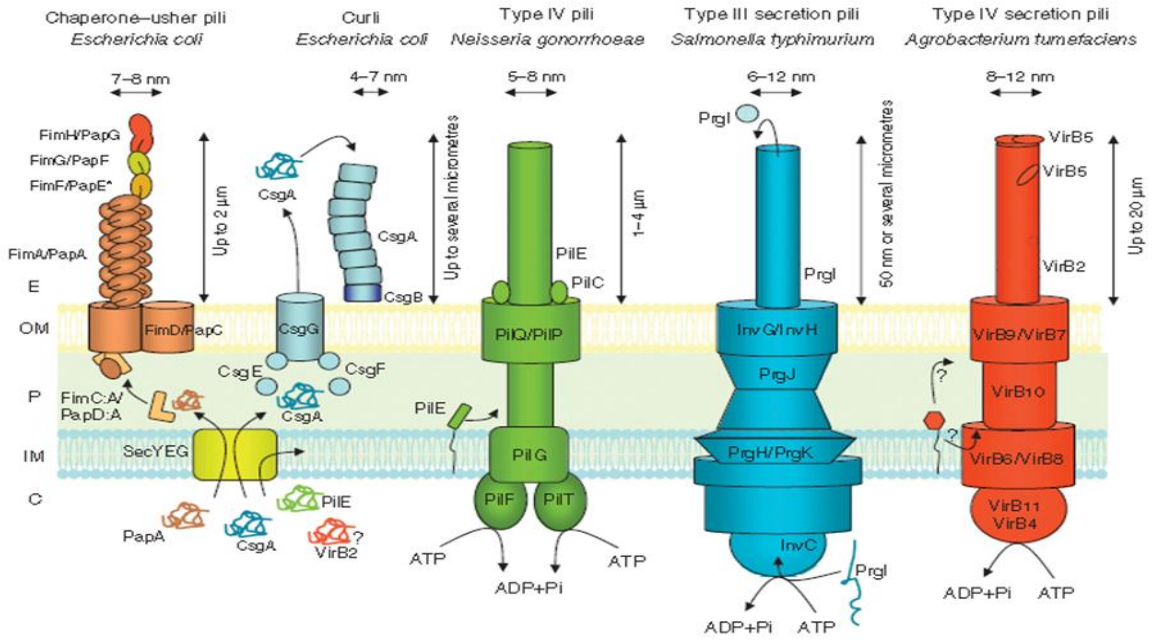


Figure 1.1 Gram-negative pili and their assembly systems (Fronzes et al. 2008). Chaperone-usher pili, curli and type IV pili are frequently involved in adhesion. The type III secretion apparatus is involved in secretion of toxins and effectors over the IM and OM of Gram-negative bacteria and injection into target cells. The type IV secretion apparatus exports transfer-DNA from *Agrobacterium tumefaciens* into a plant cell during transfection. Pili systems relevant to *M. capsulatus* are described in the text

Chaperone-usher type pili

Chaperon-usher pathway of pilus genesis is the most widespread of the pili pathways and gives rise to linear, unbranched pilus structures collectively known as CU pili. 6 clades (α-,β-,γ-, κ-, π- and σ-fimbriae) of related CU pili have been defined based on phylogenetic analysis. No structural information is available for σ- and β-fimbriae, and the following description is based on the well characterized rod-like pili found in the κ, γ and π clades. Uropathogenic *Escherichia coli* (UPEC) type 1-pili and pyelonephritis-associated (P) pilus are perhaps the best-characterized pilus systems and has become the prototypical examples of CU pili systems. Fronzes et al. (2008) and Waksman and Hultgren (2009) has reviewed the architectures and biogenesis of these particularly well known CU pili. Prepilin subunits are secreted across the inner membrane (IM) and into the periplasm. Central to the transport and assembly is a periplasmic chaperone (PapD/ FimC) responsible for the correct folding of the pilin and an outer membrane usher protein (PapC/FimD). The usher protein recruits the chaperone in complex with the pilin subunits to the outer membrane (OM), catalyses polymerization and transport across the OM, and function as an assembly platform for the pilin subunits. The fully assembled pilus is made of several types and hundreds to thousands of pilin subunits that together form a rigid pilus rod (PapA/FimA) with a flexible tip (PapE/FimF) with the adhesin distally.

CU pili are often important virulence factors. P-pili are associated with pyelonephritis and are assumed to be necessary for *E. coli* colonization of the urinary tract as part of the disease development. Type I pili are found in both commensal and uropathogenic *E.coli*, but it appears like the adhesins of pathogenic strains have a higher affinity for glycoprotein-receptor monomannose residues, prevalent in the urinary tract, while adhesins of commensal strains have a higher affinity of trimannose as reviewed by Pizarro-Cerda and Cossart (2006).

Type IV pili

Type IV pili (T4P) are complex structures expressed by several Gram-negative pathogens and are considered important determinants for bacterial virulence and host colonization.

All Type IV pilus systems share some common components as described by Craig and Li (2008). The pilin subunit is the main structural unit of the pili, and the assembled organelle is composed of homopolymers of this subunit. Assembly of the pilin polymer is a rapid process that requires ATP, and an assembly ATPase is necessary to provide energy for the assembly of the structure. An inner membrane protein recruits the ATPase from the cytoplasm. The final structure is stabilized by a lipoprotein, and the pilus can be retracted through the activity

of a retraction ATPase. The adhesin component of the organelle is present on the tip of the structure. Other minor pilins may be present, forming a base for the structure. The fully assembled complex is a homopolymer of thousands of pili subunits protruding like a rod from the cell surface.

Neisseria gonorrhoeae and *N. meningitidis* are human pathogens causing the diseases gonorrhea and meningitis. Both pathogens use T4P in the initial binding to epithelial cells priming the host cell for further binding. Other OM adhesins, Opas and OmpA, are involved later in the process. In the pathogenic *Neisseria ssp.* the tip adhesin has been identified as PilC. Both *N. gonorrhoeae* and *N. meningitidis* have two alleles for PilC, PilC1 and PilC2. In *N. gonorrhoeae* both alleles are adhesive, while in *N. meningitidis* only the PilC1 confers adherence (Morand et al. 2001).

1.2.2.2 Afimbrial adhesins

Although fimbrial adhesins may have received the most attention, a number of afimbrial adhesins are recognized. The *Escherichia coli* intimin and the *Y. pseudotuberculosis* invasins are structurally similar OM proteins. These adhesins are both rod shaped and composed of immunoglobulin-like domains with a tip containing an incomplete lectin-like domain engaging the host receptor (Donnenberg 2000).

Outer membrane protein A (OmpA) are proteins with important structural roles in a number of Gram-negative bacteria. These multifaceted proteins are found to function as adhesins and invasins across several enterobacteria (Smith, Mahon et al. 2007; Serino, Nesta et al. 2007) conferring binding and invasion to a range of cell types .

In *E.coli K1*, the cause of meningitis in neonates, OmpA contributes to binding to macrophages, phagocytosis and intracellular survival within the phagocyte (Sukumaran et al. 2003). An OmpA-like protein found in *N gonorrhea* was also shown to important for adhesion and invasion into human cervical carcinoma and endometrial cells and to be required for entry into macrophages (Serino et al. 2007).

Bordetella pertussis is a pathogen causing whooping cough. *B. pertussis* adheres to epithelial cells and leukocytes by filamentous hemagglutinin (FHA), a surface-associated protein (Ishibashi et al. 1994; van den Berg et al. 1999). FHA is an adhesin with broad binding ability. At least three different attachment activities have been identified: A carbohydrate

binding site used for attachment to ciliated cells of the upper respiratory tract, a heparin binding site, possibly involved in adhesion to epithelial cells and extracellular matrixes, and a integrin binding motif involved adhesion to macrophages (Locht et al. 1993).

1.2.3 Manipulating host membrane or cytoskeleton

As a part of an invasion or dissemination process several pathogenic bacterial species are capable of manipulating host cell membranes and cytoskeleton by targeting key components of host cell pathways (Bhavsar et al. 2007)

N. meningitidis and *N. gonorrhoeae* T4P mediated adhesion trigger rearrangement of host-cell surfaces and cytoskeleton. This initial adhesion event induces formation of cortical plaques enriched in specific subsets of membrane proteins (Merz, Enns et al. 1999). Among these are receptors for other *Neisseria* adhesins (CD44v3), proteins involved in migration and activation of immune cells (CD44 and ICAM-1) and proteins associated with the cortical cytoskeleton (EGFR). Merz, Enns et al. (1999) demonstrated that formation of cortical plaques was T4P dependent, and not influenced by other *Neisseria* adhesion factors. Initial adhesion thus primes the host cell for subsequent adhesion events and may influence the inflammatory response of the host. Merz, Enns et al. (1999) suggested that the pilus either exerted its effect by mechanical forces or by exporting secreted effector molecules.

Enteropathogenic *E. coli* (EPEC) and intracellular *Salmonella enterica* both harbor specialized secretion systems capable of injecting effectors directly into host cells and can manipulate the host cytoskeleton (Donnenberg 2000). *E. coli* uses the secretion system to transfer a protein, translocated intimin receptor (Tir), into host epithelia cell. Tir is then expressed on the host cell surface where it functions as a receptor for another *E.coli* protein, the adhesin intimin. On the cytoplasmic face of the host membrane Tir is tyrosine-

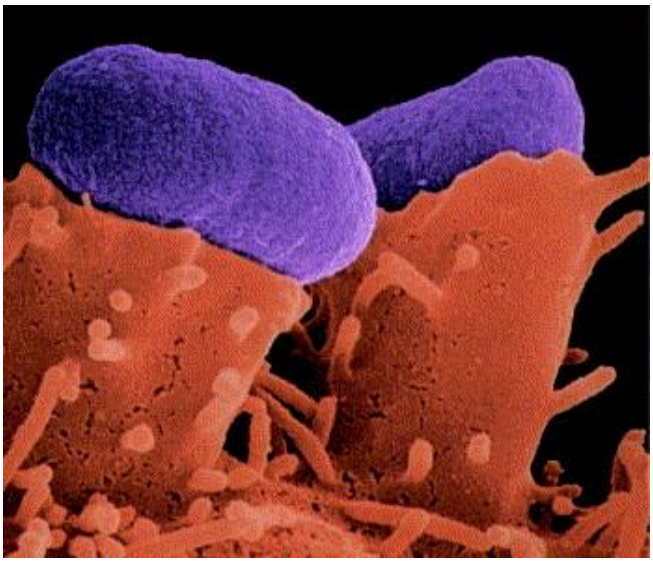


Figure 1.2. Enteropathogenic E.coli manipulate host-cell actin pathways resulting in formation of a pedestal beneath the bacterium (http://www.finlaylab.msl.ubc.ca/research_projects/E.coli.html)

phosphorylated leading to recruitment of the host adapter protein Nck which in turn recruits other proteins that mediate polymerization of actin filaments at the site of bacterial attachment. The result is formation of an elongated host cell structure, a pedestal, beneath the bacterium (Figure 1.2).

Bacteria of the *Shigella* genus may also take advantage of host actin pathways by polymerizing actin filaments on one pole of the bacterium thereby propelling the bacterium into a neighbouring cell (Donnenberg 2000).

1.2.4 Host invasion and colonizing the phagocyte.

Some microorganisms are able to enter into phagocytic or non-phagocytic host-cells .By invading host cells the pathogen may avoid defensive host immune responses and competition from other microbes, but the strategy requires specialization and poses several challenges.

First of all the bacterium must find a way to penetrate the host membrane. Intracellular microbes may enter the cells actively, by binding to host receptors and manipulating the actin cytoskeleton, or it may take advantage of host internalization mechanisms for example by utilizing clathrin-mediated endocytosis or passively by allowing itself to be engulfed by phagocytic cells (Bhavsar, Guttman et al. 2007).

Secondly, the bacterium must survive within the host cell. Intentionally allowing itself to be engulfed by phagocytic cells may appear as risky behavior since most phagocytosed microorganisms are rapidly degraded within the phagocyte. Immediately after internalization the microorganism is contained within a host-cell membrane-derived vesicle. The microbe may choose to stay in the vacuole and replicate here, or it may escape from the vacuole and replicate in the cytosol. Escaping from the vacuole typically involves secretion of a pore-forming protein and/or one or more phospholipases (Hybiske & Stephens 2008).

Finally, the pathogens must find a way to escape the host-cell. Escape strategies include host cell lysis; protrusion into neighboring cells; extrusion into membrane-bound compartments, expulsion by exocytosis or induction of either apoptotic or proinflammatory cell death (Figure 1.3) (Hybiske & Stephens 2008)

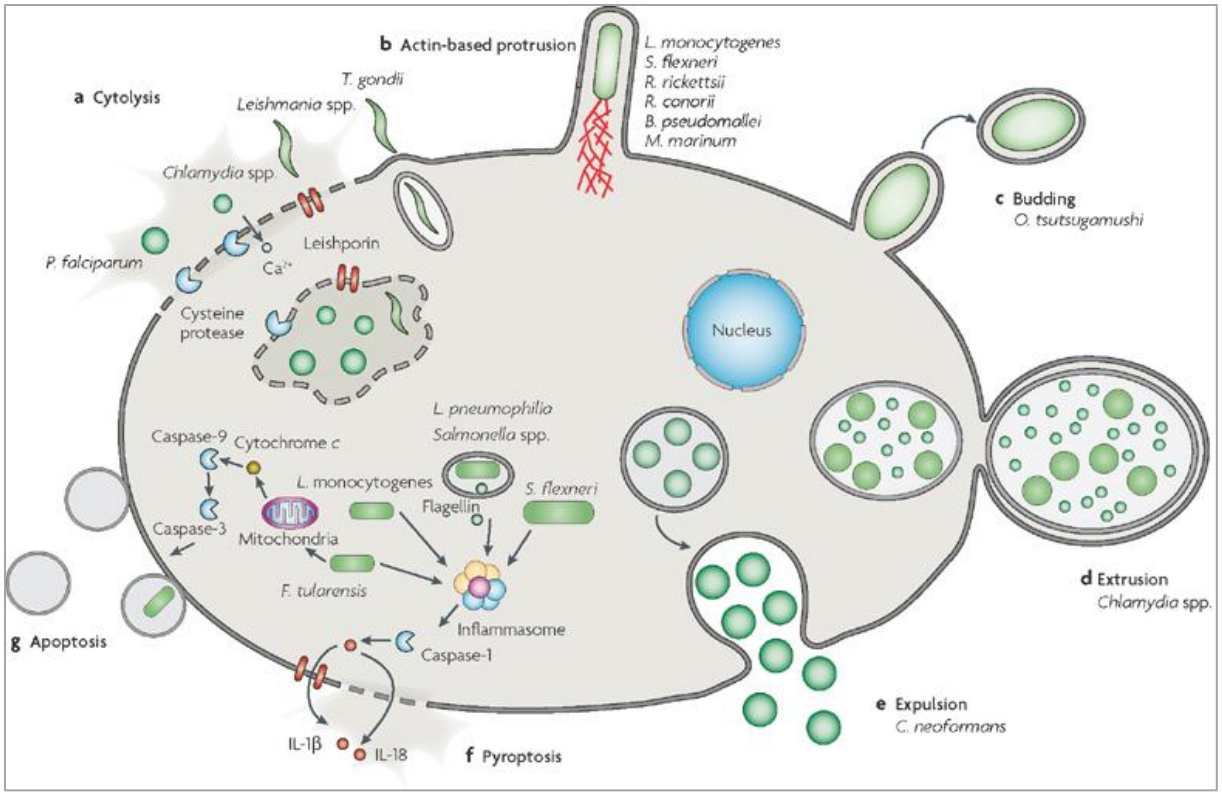


Figure 1.3. (Hybiske and Stephens 2008) Strategies for escaping host cell destruction. A) Cytolysis, disruption of host cell membrane induced by proteases, pore-forming proteins or undefined mechanism. B) Actin filament -driven protrusion from the cell membrane and into a neighbouring cell. C) Budding out to extracellular space in a membrane/cytoplasm surrounded vacuole. D) Extrusion of a double vacuole. E) Expulsion by exocytosis. F) Proinflammatory host cell death triggered by host sensing of bacterial molecules, activation of caspase-1 and cytokine production. G) Apoptosis by intrinsic pathway activation triggered by unknown bacterial molecule.

1.2.5 Evading or modulating host immuneresponces

Not surprisingly, pathogenic bacteria have evolved a number of mechanisms to evade host defence mechanisms. Hiding within phagocytic or non-phagocytic host cells may be an efficient way of avoiding the host "radar", but extracellular pathogens must find other ways to cope with the defense mechanisms of the host.

The cells of the innate immune system rely on recognition of structural patterns, microbe-associated molecular patterns (MAMPs), not present in the host, but common to many bacteria rather than species specific properties. Innate immune cells express a wealth of PRRs, in particular Toll like receptors (TLRs), nucleotide oligomerization domain receptors (NODs) and NOD-like receptors (NLRs). One important effect of stimulation of immune cells through PRRs is the expression of proinflammatory cytokines, under the control of the transcription factor NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells**)**. Upon receptor stimulation the IκB kinase (IKK) phosphorylates an NF-κB inhibitor, IκB resulting in ubiquitination and proteolytic degradation and subsequent activation of the

transcription factor. Given its central role in proinflammatory responses it is not surprising that several pathogens target the NF-κB activation pathway and have developed strategies to circumvent it (Bhavsar et al. 2007).

The complement system is one of the main effector systems used against pathogenic microorganisms. Complement activation initiates a cascade of events leading to formation of opsonins, (e.g. C3b) that enhances phagocytosis of bacteria, release of anafylatoxins (e.g. C3a and C5a) and formation of a terminal complement complex contributing to lysis of the microbe. As the complement cascade is a powerful and potentially detrimental defense system it is tightly regulated by the host. Some bacteria have learned to take advantage of the host regulatory proteins thereby avoiding activation of the complement cascade. *Neisseria* T4P and OmpA of *Escherichia coli K1,* contributes to serum resistance by binding the C4 binding protein,C4bp, a regulator of the complement cascade (Prasadarao et al. 2002). C4bp remains functional after binding resulting in significantly lower deposition levels of C3b, C5b, and the terminal complement complex on $OmpA^+$ *E. coli* than on $OmpA^-$ *E. coli* (Sukumaran et al. 2003).

1.3 The bacterial secretome

Proteins are structurally and functionally diverse and are involved in all the vital processes of bacterial cells. Surface-exposed proteins are of particular importance as they are present on the interface between the bacteria and its environment and allow the microbe to sense and respond to its surroundings. In a context of microbe-host interactions secreted proteins are important virulence factors: Surface-attached proteins play roles in chemotaxis, host cell recognition, attachment and invasion, immune evasion and modulation. Secreted and released proteins are important effectors during pathogenesis (Bhavsar et al. 2007; Finlay & Falkow 1997).

1.3.1 OM-associated proteins, the lipoproteins and the β-barrels

Two types of proteins are found attached to the OM of Gram-negative bacteria, lipoproteins and β-barrels. Lipoproteins (LPs) associate with the inner or outer leaflet of the OM via a lipid anchor. Many LPs are important virulence factors with roles in adhesion, colonization and immune modulation (Kovacs-Simon et al. 2011), or envelope proteins involved in protein sorting and transport of LPS, drugs and proteins (Tokuda 2009).

In contrast to IM proteins whose membrane integral regions show α-helical secondary structure, the integral proteins of the OM consist of antiparallel β-sheets folded into a barrel-shaped structure. These proteins called β-barrels have important roles in transport of proteins, ions and virulence factors. Some are proteases or have defensive functions binding to foreign proteins (Schulz 2002).

1.3.2 The secretome and the surfactome

Distinguishing between surface-attached and released proteins can be difficult for several reasons: Both types of proteins are exported to the exterior through the same mechanism, and identification by bioinformatics methods is based on searching for motifs common to both groups. Predicting subcellular location is further complicated by the fact that not all exported proteins carry secretion motifs, and not all proteins containing such motifs end up on the surface. In bacteria carrying two membranes they end up in the periplasmic compartment by default.

When using experimental approaches, contamination of secreted proteins by cytoplasmic proteins is a common problem and may obscure identification of true surface proteins. In addition, some proteins may appear both as attached proteins and, under different conditions, may be released to the environment. One of the most abundant OM proteins in *M. capsulatus* (Bath), 'Methylococcus outer membrane protein' (MopE) indeed behaves in this manner. Under low copper conditions MopE is upregulated, and a truncated variant, copper-binding MopE* is released to the medium as reviewed by Karlsen, Larsen et al. (2011). Such examples show that the distinction between released and surface-expressed proteins is not always clearcut.

In the broadest sense an organism's secretome is defined as all proteins exported by that organism (Greenbaum, Luscombe et al. 2001). According to that definition, both proteins attached to the bacterial surface, proteins released to the environment and subunits of surface appendages are part of the secretome as they are all moved from the cell's interior to its exterior. Several authors indeed use the term by that definition (Desvaux, Hebraud et al. 2009). In the following, the definitions suggested by Desvaux, Hebraud et al. (2009) will be used: Export is the active transport of substances from the cytoplasmic space, translocation is movement across a lipid bilayer and secretion is transport from an interior cellular compartment to the exterior of the cell. A cell's secretome defines all surface-attached

proteins and secreted proteins released to the media under normal growth conditions while the term “surfactome” will be reserved for referring to proteins when these are believed to be attached to the bacterial surface. Exoproteins or extracellular proteins are proteins found in the extracellular environment, and released will be used when discussing proteins released to the extracellular milieu after secretion.

1.3.3 Gram-positive and Gram-negative bacteria and their cell walls

Bacteria have traditionally been separated in two groups based on their ability to retain crystal violet after Gram-staining and alcohol/acetone destaining. Gram-positive bacteria carry a dense outer peptidoglycan wall that retain crystal violet and make the bacteria appear purple after staining, while the much thinner cell wall of Gram-negative bacteria allow crystal violet to be washed out by destaining, leaving the bacterium with a red appearance. Gram staining reflects structural differences, but Gram-variable taxa exist, and some bacteria stain positive, yet have two bilayers. “Monoderm” (referring to bacteria with one membrane) and “diderm” (referring to bacteria with two membranes) is therefore sometimes preferred when discussing protein transport as it refers to the structural differences that are responsible for the final fate of exported proteins and not to the ability to retain stain. The build-up of a typical monoderm and diderm prokaryotic cell wall is illustrated in figure 1.4)

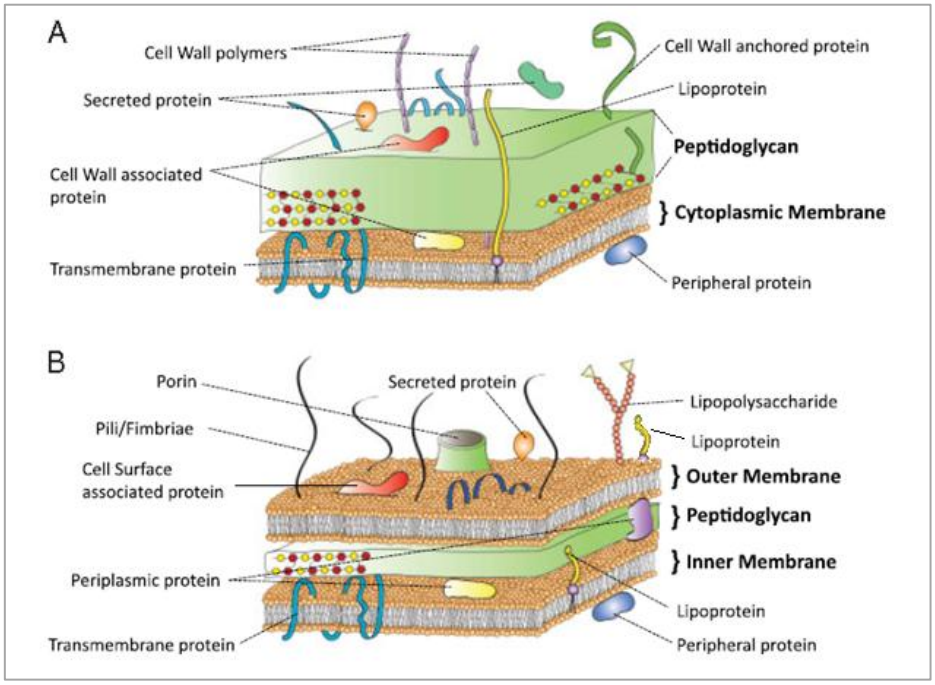


Figure 1.4. Modified from Solis and Cordwell (2011). Prokaryote Cell walls

(A) Gram-positive bacteria with a cytomplasma membrane and a dens outer peptidoglycan layer with associated cell wall polymers

(B) Gram-negative cell with an inner membrane and outer membrane separated by a thin peptidoglycan layer. Integral outer membrane proteins lipoproteins and pili structures are associated with the outermembrane.

Methylococcus capsulatus (Bath) has a classical three-layered cell wall structure characteristic of Gram-negative cells and belongs to gammaproteobacteria, a group of Gram-negative bacteria. A closer description of a Gram-negative cell wall follows.

1.3.4 The Gram-negative cell wall

The cell wall is a highly specialized structure that must accommodate often contrasting demands: It is a barrier protecting the cell from harmful substances, yet it must allow nutrient uptake. It must prevent leakage of cellular components, but at the same time allow excretion of waste products. It should support the structure of the cell, but must be flexible enough to allow change in cell size. Structures of Gram-negative cell walls have been reviewed by Beveridge (1999).

Gram-negative bacteria carry two distinct membranes. The plasma membrane, often referred to as the IM, is a semipermeable phospholipid bilayer that encloses the cytoplasma. It is not considered a part of the cell wall, but is part of the cell envelope. Embedded in the membrane, or sticking out from the membrane, are membrane proteins containing one or more alpha helices. Surrounding the plasma membrane is a thin layer of peptidoglycan, a polymer of glycans cross-linked by peptide side chains that provides structural support to the cell. Between the peptidoglycan and the OM is a gelatinous matrix called the periplasm. Proteins and other macromolecules aimed for release or transport to the OM must pass this region, making the periplasm an ever changing environment reflecting the cell's metabolic state. OM is a lipid-protein bilayer and the outermost layer of the cell wall. Noticeably, lipids are not symmetrically distributed over the inner and outer faces of the membrane. The inner face of the membrane contains most of its phospholipids, while the outer face consists mainly of lipopolysaccharides (LPS). These provide an overall negative charge to the cell surface, providing opportunity for interactions with cations in the cell's environment. Proteins associated with the OM are of two different types: Lipoproteins can be found anchored to the OM through lipid anchors, acyl chains that are attached to an aminoterminal cysteine residue on the protein (Tokuda 2009). The other type of proteins is membrane integral proteins consisting of β-sheets wrapped into cylindrical shapes, the β-barrel proteins.

1.3.5 Sec and Tat translocation pathways and fate of exported proteins in Gram-positive and Gram negative bacteria

Whether a bacterium has one or two membranes have important implications for how proteins are transported in that cell and where those proteins end up. In monoderm bacteria most proteins aimed for secretion need only be translocated across one membrane. Most such proteins are exported in an unfolded state through the Sec translocation pathway or in folded conformation through the twin-arginine translocation (Tat) pathway. A smaller number of proteins are transported through alternative transport pathways. Both systems operate by recognition of an N-terminal sequence, a signal peptide that directs the protein for translocation across the IM and is cleaved off or remains anchored to the IM after the export. In diderm bacteria Sec and Tat transport in Gram- negative cells only mediates translocation over the IM. Specialized systems are required for the subsequent translocation of proteins across or insertion of proteins into the OM. Predicting subcellular protein location therefore requires knowledge of the structures of the cell as well as an understanding of the transport systems used by the cell. The fate of protein after export, translocation and secretion in monoderm vs. diderm bacteria is illustrated in (
Figure 1.5)

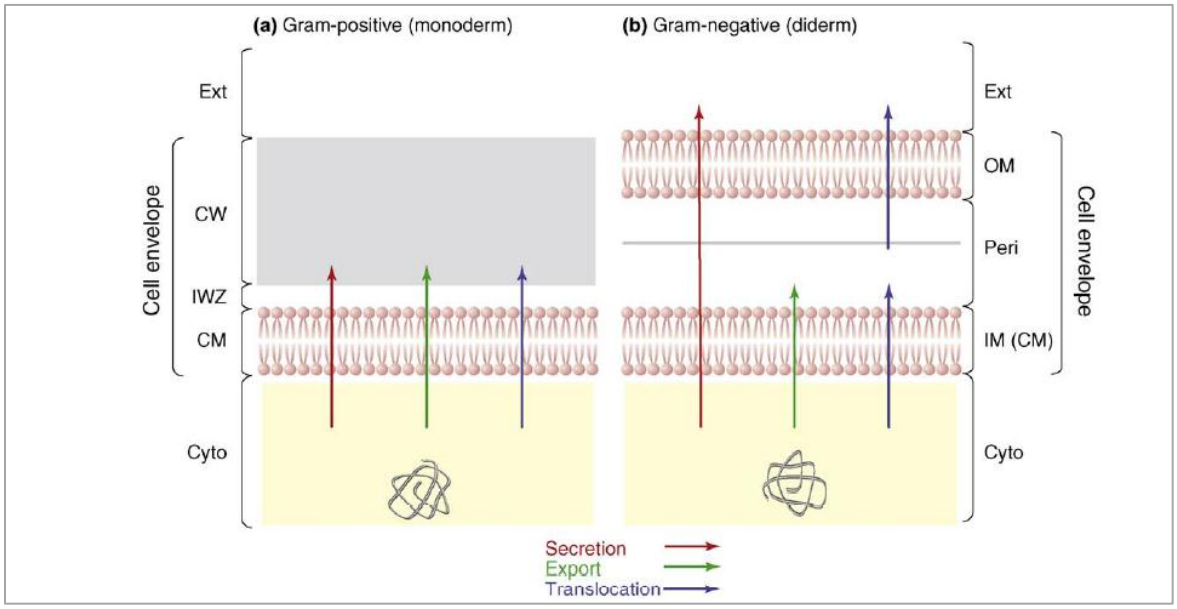


Figure 1.5. (Desvaux, Hebraud et al. 2009) Fate of proteins after secretion in a) Monoderm and b) Diderm bacteria. Ext: extracellular milieu; CW: cell wall, IWZ, inner wall zone; CM, cytoplasmic membrane; Cyto, cytoplasm; OM, outer membrane; Peri: periplasm; IM, inner membrane.

1.3.6 Protein secretion across the OM of diderm bacteria

Translocation of proteins across the OM of diderm bacteria requires specialized structures. Six major protein secretion systems have been defined so far (

Figure 1.**6**) (some authors suggest eight) and can be further classified according to phylogenetic differences. (Thanassi and Hultgren 2000),(Stathopoulos, Hendrixson et al. 2000),(Korotkov, Sandkvist et al. 2012).

These systems are essential for bacterial interactions with host cells. Adhesins and other virulence determinants are important substrates, and the repertoire of secretion systems that a bacterium accommodates affects the kinds of interactions it may have with its host. Knowing which systems a bacterium harbors may therefore give important information about the relationship between the microbe and its host.

In the following the secretion systems of Gram negative bacteria relevant to *M. capsulatus* (Bath) are described and examples of their roles in pathogenesis or host interaction are given.

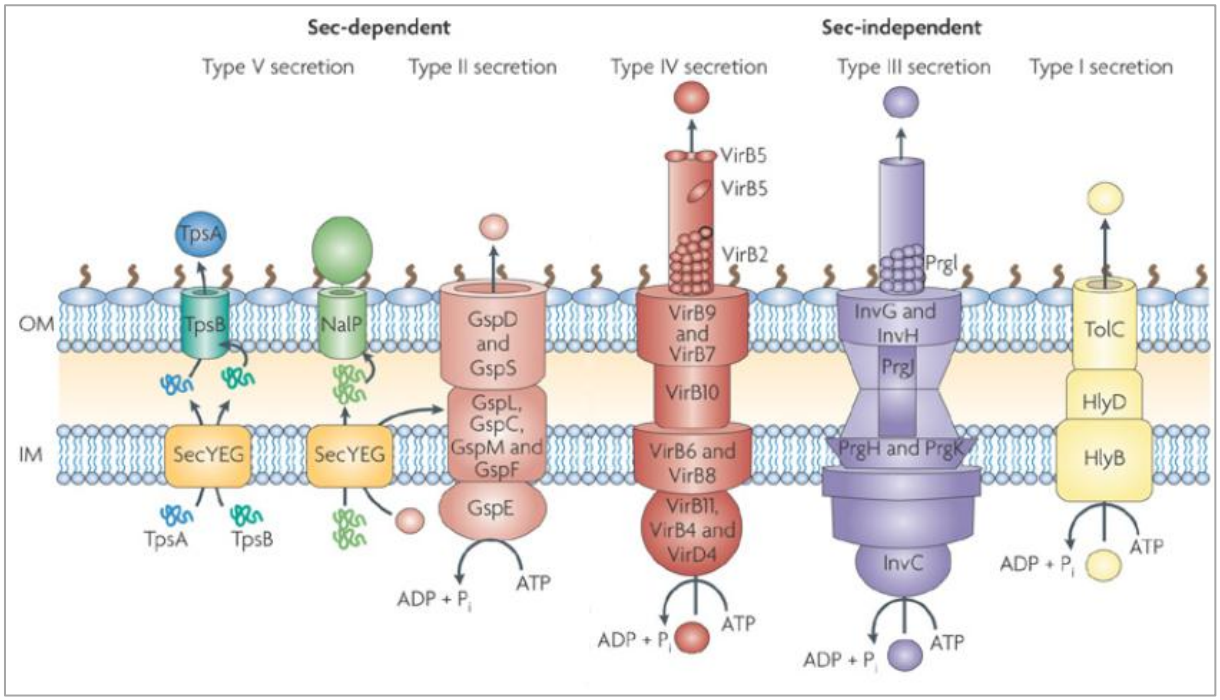


Figure 1.6. Secretion systems of Gram-negative bacteria. The Sec translocation dependent type II and type V, and the Sec independent type IV, III and I secretion systems are show. The poorly characterized type VI secretion system is not depicted.

1.3.6.1 Type I secretory system, ABC-transporter dependent secretion

The type I secretion system (T1SS) spans the IM, periplasm and OM, and thereby surpasses the Sec/Tat machineries. The system consists of three components: An ABC transporter with an ATP-binding cassette, a membrane fusion protein and an outer membrane factor. The outer membrane factor forms a channel through the OM and extends into the periplasmic space. The membrane fusion protein spans the periplasmic compartment, connecting the OMF with the IM ABC transporter. The energy for the secretion process is provided by ATP-hydrolysis. Two major groups of ABC exporters exist: one specific for transport of large proteins, the other for export of peptides and small proteins (Tseng, Tyler et al. 2009).

1.3.6.2 The type two secretory system, a Sec translocation-dependent secretion system

The type two secretion system (T2SS) is a structure exclusively found in Gram-negative bacteria that function in the secretion of folded enzymes and toxins into the extracellular environment. Unlike the rather simple structure of T1SS the T2SS is a large complex consisting of as many as 12-15 different proteins (Figure 1.6). Most of these proteins are found associated with the IM, but are in fact not involved in protein translocation across the IM. Instead proteins aimed for secretion via T2SS is translocated via the Sec or Tat translocation pathways.

Although the fully assembled type II secretion structure has not been visualized, current knowledge suggest that the apparatus spans both the inner and OM and consists of four different subassemblies: The pseudopilus spanning the periplasm, the OM pore, the IM platform and the secretion ATPase (Korotkov, Sandkvist et al. 2012).

The *Vibrio cholera* cholera toxin is an example of a protein exported via this pathway.

1.3.6.3 Type three secretion system, the injectisome

Type three secretion systems (T3SS) and other Gram-negative transport systems are reviewed by Thanassi and Hultgren (2000) and Tseng et al. (2009).The type three transport system (T3SS) is a highly specialized system found in some host-interacting bacteria. This secretion system is sometimes called an injectisome because it is capability of injecting secreted effector molecules directly into a host cell cytosol. This secretion system is Sec independent and forms a large complex of up to 25 proteins spanning the IM, the periplasm, the OM and, during injection, presumably the host cell membrane. Proteins referred to in the text are examples from *Yersinia*. The components of the complex form a basal body similar to a

flagellar basal body, spanning the IM and a central channel spanning the periplasm. Translocation is probably powered by proteins at the cytoplasmic side of the basal body. An OM secretin provides a channel possibly gated by YopN. Two proteins (YopB and YopD) are thought to form a secretin through the host cell plasma membrane allowing secretion directly into the cytosol of the host. *E. coli* (EPEC) uses a T3SS to inject its Tir receptor into host cells (Donnenberg 2000)

1.3.6.4 Type five secretion systems, autotransporters

Proteins utilizing the type five secretion system (T5SS), like T2SS transported proteins, are dependent on Sec translocation over the IM. Autotransporters are defined by their ability to form an OM protein pore, the translocator. The domain being transported is referred to as the passenger domain. Several subclasses of TVSS have been defined. The two types of T5SS described here are shown in figure 1.7

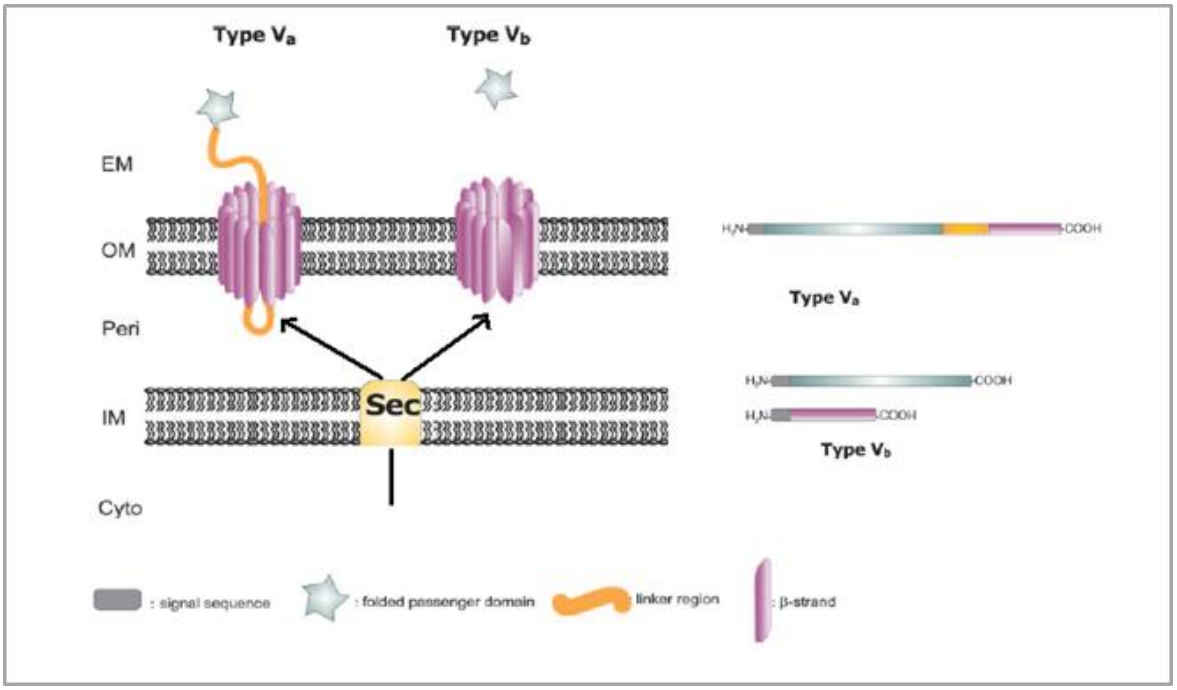


Figure 1.7 Modified from Henderson et al.(2004). Left: Type Va and Vb autotransporter systems schematically depicted. Right: Protein with functional domains; the signal peptide, the passenger domain, the linker region and the β-domain.

The classical autotransporters, type Va secretion system, carries N-terminal signal peptides and are translocated to the periplasm by the Sec translocation system. The traditional view is that autotransported proteins, as indicated by the name, carry all components needed for transport in their own amino acid sequence. However an OM protein Omp85 (BamA) have been shown to be required for insertion of the OM pore in the membrane. After signal peptide-directed Sec translocation a C-terminal translocation unit inserts into the OM in the form of a β-barrel structure through which the passenger domain can be secreted. Some, but not all autotransporter are autoproteolytic and contains a domain capable of releasing the passenger domain after secretion. Autotransported proteins are often important in pathogenesis. Both extracellular proteases and adhesins are found within this group.
Two-partner translocation, Type Vb secretion system, resembles AT translocation, but the passenger and translocator domains are located in separate polypeptide chains. The translocator proteins usually contain periplasmic polypeptide-transport-associated (POTRA) domains. The translocator and passenger is usually located together in the genome. Filamentous hemagglutinin of *B. pertussis* is a well-studied example of a type Vb secreted protein.

1.4 Predicting subcellular location by bioinformatics

The amino acid sequence of a protein contains all the information necessary for a bacterium to know how to process and place that protein, with or without the help of other proteins. This wealth of information can also be used to predict subcellular location *in silico.* Computer programs using advanced algorithms are usually needed to read the information that the cell understands "intuitively", and this information must then be interpreted in the light of cell biology. Information gathered can be based on a conserved sequence or physicochemical characteristics of the amino acid sequence, like amino acid charge, hydrophobicity or polarity.

N-terminal signal sequences, twin arginine motifs and special membrane anchoring motifs are examples of sequence features that are associated with particular secretion pathways and therefore can be used to predict the fate of proteins carrying them. Stretches of hydrophobic amino acids are usually buried among the hydrophobic tails of a lipid layer and can be used to predict membrane proteins while some amino acids are more commonly found in particular secondary structures than others. Such information can be used to predict tertiary structures with implications for subcellular localization.

1.4.1 Identifying Sec or Tat secreted proteins, SPaseI cleaved signal peptides

Proteins destined for export through the Sec or Tat translocation pathway are synthesized as precursor proteins carrying telltale signs of their travels, namely signal peptides (SP). Signal peptides are short N-terminal peptides that direct proteins for transport across a membrane (Figure 1.8). After transport the SP is cleaved off by a signal peptidase (SPase I) releasing the mature protein to the periplasm, or remains attached to the protein anchoring it to a membrane.

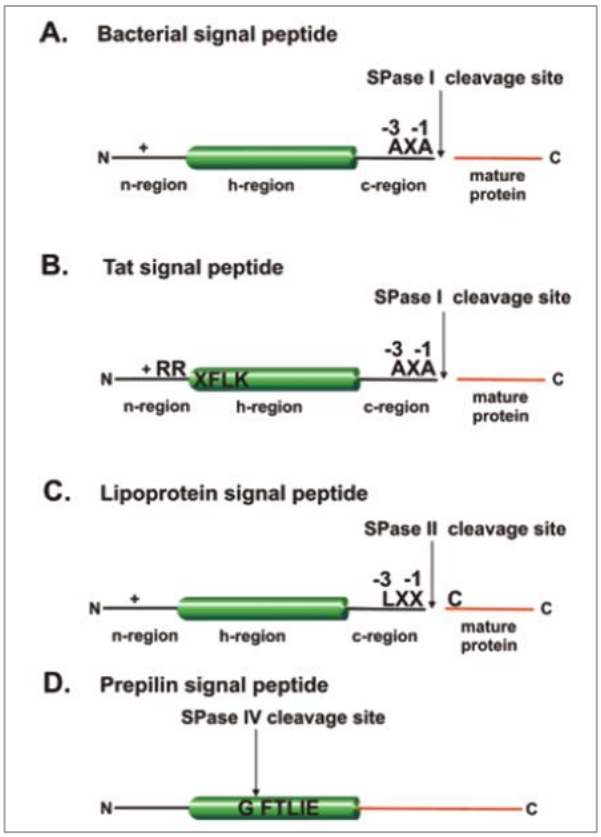


Figure1.8 Bacterial signal peptides (Paetzel, Karla et al. 2002).
A) Sec signal peptid B)Tat signal peptide with twin-arginine motif C)Lipoprotein signal peptid with lipobo (D) Prepilin like SPs found in type IV pili proteins and prepilin like proteins involved in type II secretion.

SPs vary, but typically show a tripartite structure with a N-terminal region (1-5 residues long) with positively charged residues, a mid-region (7-15 residues long) with a stretch of hydrophobic amino acids and a C-terminal region (3-7 residues) of more polar residues (von Heijne 1990). This structure, as well as the cleavage site for the SPase, is a feature that can be used to identify proteins transported over the IM.

Tat type SP can be distinguished from sec SP by containing a highly conserved twin-arginine motif and a less hydrophobic h-region than Sec SP. They are also generally longer. Some proteins contain translocation signals that are not cleaved by SPases. Such proteins remain attached to the cytoplasmic membrane after translocation. The uncleaved SP is then called a signal anchor, and the protein referred to as a type II membrane protein.

1.4.2 Identifying and predicting subcellular location of lipoprotein

Lipoproteins are associated with the IM or OM. Lipoproteins are transported over the IM by the Sec translocation pathway, and carry N-terminal SPs similar to Sec and Tat SPs. The hydrophobic region in lipoprotein SP, however, is shorter, the c-region is less polar and the SP is cleaved by a different SPase (SPaseII). Lipoproteins (LPs) can be identified by a highly conserved four amino acid lipobox around the SPaseII cleavage site with an invariable cysteine in the +1 position relative to the cutting site. Based on consensus sequences and additional rules, several prediction programs have been developed that can identify LPs from amino acid sequence. Targeting of LPs to the IM or OM requires the lipoprotein localization machinery (Lol). This consists of a periplasmic chaperone (LolA), the transmembrane complex (LolCDE) and the OM receptor (LolB) (Tokuda 2009) An aspartate in the +2 position is thought to inhibit recognition by the LolCDE complex (Tommassen 2010). The +2 rule is therefore used as a discriminator between OM and IM LPs. Later studies have shown that proteins containing other +2 residues can also be found located at the IM (Seydel et al. 1999).

1.4.3 Predicting transmembrane proteins

As mentioned, integral membrane proteins of the inner and outer membrane of Gram-negative bacteria are structurally different and different algorithms are therefor used to predict inner and outer membrane proteins. IM proteins contain one or more α-helices separated by cytoplasmic or periplasmic loops (Figure 1.9A). The inside of a lipid bilayer is hydrophobic, and amino acids embedded in this layer should therefore not have hydrophilic side chains. Some amino acids are commonly found in alpha helices. They possess high helix-forming propensity while other amino acids are so called helix breakers. Both hydrophobicity and helix forming propensity can be used to identify stretches of amino acids involved in helixes. Cytoplasmic parts of the sequence often contain positively charged residues, another feature that can be used in the prediction. By combining searches for helical regions with the search for the overall “grammar” of a protein with alternating cytoplasmic and non-cytoplasmic loops, the prediction can be improved (Krogh et al. 2001).

In Gram-negative bacteria a second type of integral membrane proteins is found in the OM (Figure B). Algorithms for prediction of outer membrane β-barrel proteins have been developed. The same way as some amino acids are more likely than other to occur in an alpha

helix, some amino acids are more likely to be involved in beta structures than others. Stretches of amino acid sequences can therefore be predicted

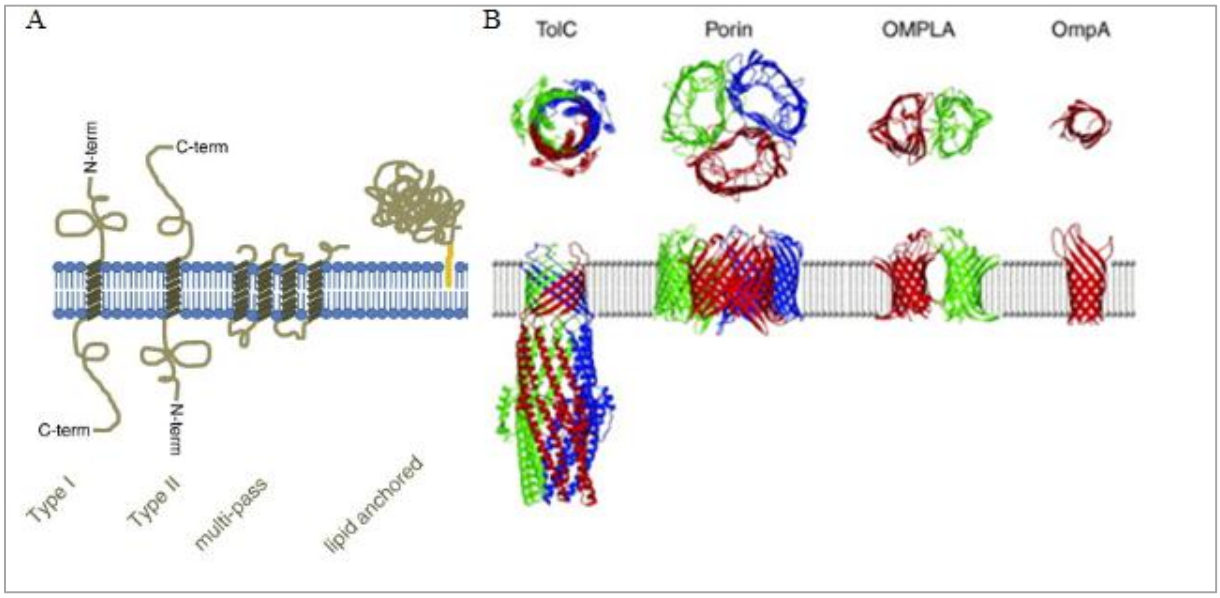


Figure 1.9. Shows the two types of membrane proteins found in Gram-negative bacteria. A) Inner mebrane proteins consisting of one or more α helices separated by cytoplasmic and periplasmic loops. Two single pass and one multipass IM proteins are shown. An IM lipid-anchored protein is also depicted. Modified from (Helbig, Heck et al. 2010). B) Examples of some beta barrel proteins, OM proteins found in Gram-negative bacteria. Top view and side view. Modified from (Koronakis, Andersen et al. 2001)

Insertion of BB into the OM requires the Bam complex in which BamA (sometimes called Omp85) is essential and a number of variable accessory proteins (Bam B-E) may be involved dependent on the bacterium studied. Components of the of the Bam complex recognizes a C-terminal signature (Tommassen 2010). Searching for C-terminal signatures can be implemented in the prediction program strengthening the prediction (Berven et al. 2004).

1.4.4 Non classical secretion

Although most secreted proteins contain SPs, there are examples of proteins with known extracellular location, but with no recognizable SP. Such proteins will not be identified by motif-searching algorithms. However, extracellular proteins, regardless of secretory pathway, share some properties that are a result of the amino acid composition but not directly sequence-dependent. Such features may be hydrophobicity, amino acid composition, charged amino acids, or other physicochemical properties. By combining features with discrimatory value in a predicatory network one can distinguish secreted proteins from proteins residing in the cell without relying on conserved sequences (Bendtsen et al. 2005).

1.4.5 Conserved domains

Protein domains are shorter parts of proteins that are often conserved across otherwise dissimilar proteins. The conservation of protein domains reflects that they are functional units. Several databases of curated families of protein domains exist. If a domain of known function in other proteins is found in a protein of which little is known, this information can give clues about the function of the less described protein. Some domains are typical for proteins found in a particular subcellular location and can for example be used as an indicator for the protein being extracellular.

1.4.6 Genome-wide sub cellular location prediction

Although a number of useful prediction programs exist that can handle genome-wide prediction, all programs have their shortcomings, and results must be interpreted with care. Different programs constructed to predict the same features use different algorithms and may give conflicting results. Taking into account the *a priori* assumptions made by each algorithm and evaluating the strength of predictions is very important to avoid erroneous predictions. In this study the proteome of *Methylococcus capsulatus* (Bath) has been examined and subcellular localization predictions made for all proteins using bioinformatics methods, like those described above. Extracellular proteins from exponential growth phase have been identified by LC-MS and proteins analyzed for possible roles in microbe-host interactions.

2 Materials and methods

2.1 Materials

Nitrate Mineral Salts Medium (NMS) (Whittenbury et al. 1970) 50ml NMS salt solution (KNO_3 (0.1g/ml), $MgSO_4x6H_2O$(0.1g/ml), $CaCl_2$ (0.1g/ml)), 0,5ml $Na_2MoO_4x2H_2O$ (0.26g/l), 0.5 ml trace elements ($CuSO_4x5H_2O$ (200g/l), $FeSO_4x/H_2O$ (500g/l), $ZnSO_4x7H_2O$ (400g/l), $CoCl_2X6H_2O$ (0.05g/l), EDTA disodium salt (250g/l), $MnCl_2x4H_2O$ (0.02 g/l), $NiCl_2x6H_2O$ (0.01g/l)) and 0.05ml FeEDTA (0.038g/ml) in dH_2O to a final volume of 500ml adjusted to pH 6.8 with phosphate buffer ($Na_2HPO_4x12H_2O$ (71.6 g/l) and KH_2PO_4 (26.0g/l)).

Modified RPMI1640 cell culture medium. 500 ml RPMI 1640 medium (PAA, Austria) was enriched with 10% heat-inactivated fetal calf serum (obtained from PAA), 5 ml 110mM sodium pyruvate, 5ml 100x non-essential amino acids (PAA), 25µl 1M monothioglycerol and 50 mg/ml gentamycin.

Buffy coats. Buffy coats from normal healthy volunteers were obtained from Østfold Hospital Trust after informed consent.

Antibodies. Mouse anti-human CD14-coupled magnetic beads were provided by Miltenyi Biotec, Germany.

Cytokines and growth factors. GM-CSF (100000ng/ml) and IL-4 (100000ng/ml) was provided by (ImmunoTools, Germany.)

2.2 Bioinformatic tools and databases

ß-barrel outer membrane protein predictor (BOMP) (Berven, Flikka et al. 2004) predicts outer membrane ß-barrels and is available at (http://www.bioinfo.no/tools/bomp).

LipoP predicts lipoprotein signal peptides in Gram-negative Eubacteria(Juncker, Willenbrock et al. 2003) and is available at (http://www.cbs.dtu.dk/services/LipoP/)

Lipo predicts lipoproteins in Gram-negative bacteria (Berven et al. 2006) and is available at (http://www.bioinfo.no/tools/lipo

TatP is a prediction server that predicts twin-arginin signal peptides (Bendtsen, Nielsen et al. 2005) TatP is available at (http://www.cbs.dtu.dk/services/TatP/)

Tatfind is a prediction server that predicts the presence of prokaryotic Twin-Arginine Translocation (Tat) signal peptides (Rose et al. 2002) Tatfind is available at (http://signalfind.org/tatfind.html)

TMHMM is a hidden Markov model that predict membrane protein topology (Krogh, Larsson et al. 2001). TMHMM is available at (http://www.cbs.dtu.dk/services/TMHMM/)

Phobius is a hidden Markov model that combines transmembrane topology and signal peptide predictions (Kall, Krogh et al. 2007). Phobius is available at (http://phobius.cgb.ki.se) and (http://phobius.binf.ku.dk).

SecretomeP is a prediction method for identification of non-classically secreted proteins in both Gram-positive and Gram-negative bacteria (Bendtsen et al. 2005) SecretomeP is available at (http://www.cbs.dtu.dk/services/SecretomeP/).

NCBI's Conserved Domain Database (CDD) is a conserved domain database that can be searched using protein queries (Marchler-Bauer et al. 2011) CDD is available at (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)

Pfam is a collection of protein families (Finn et al. 2010) that can be searched using a protein query. Pfam is available at (http://pfam.sanger.ac.uk/)

The Basic Local Alignment Search Tools (Blast) are tools for searching protein and DNA databases for sequence similarities (Altschul et al. 1997) using a nucleotide, translated nucleotide or protein query. Blast is available at http://blast.ncbi.nlm.nih.gov/Blast.cgi

The comprehensive microbial resources (CMR) genome properties tool is a tool for displaying information on the characteristics of organisms derived from genomic data and literature sources (http://www.tigr.org/Genome_Properties)(Davidsen et al. 2010)

EMBOSS Transeq translates nucleic acid sequences to their corresponding peptide sequences. Transeq is available at (http://www.ebi.ac.uk/Tools/st/emboss_transeq).

Emboss Needle is a tool for conducting global pairwise sequence alignments. Emboss is available at (http://www.ebi.ac.uk/Tools/psa/emboss_needle).

SwissModel is a protein structure homology-modeling server available at (http://swissmodel.expasy.org/)

PyMol is a molecular visualization system. PyMol was downloaded from (http://www.pymol.org/).

Uniprot is a catalog of protein informatation available at (http://www.uniprot.org/),

National Center for Biotechnology Information, (NCBI) is a database of biotechnology information available at (http://www.ncbi.nlm.nih.gov/).

The Comprehensive Microbial Resource (CMR) is a database displaying publicly available, complete prokaryotic genomes. CMR is available at (http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi)

The Virulence Factors of Pathogenic Bacteria database (VFPB) is a database of prokaryote virulence factors available at (http://www.mgc.ac.cn/VFs/).

2.3 Methods

2.3.1 *Methylococcus capsulatus* (Bath) culture, biotin labeling and trypsin shaving

Continuous culture of Methylococcus capsulatus (Bath)

A continuous culture of *M. capsulatus* (Bath) was used for all experiments conducted in this study. *M. capsulatus* (Bath) was grown on NMS medium in 100-300ml flasks with Suba-Seal

red rubber closures with a headspace of air:CH_4:CO_2 75:23.75:1.25. The cultures were incubated at 45°C with 200 rpm.

Proteolytic shaving of M. capsulatus (Bath)

Trypsin-shaving was performed according the protocol used by Bohle, Riaz et al. (2011) Cells were grown ON and then resuspended in 10mM HEPES pH 7.4 to OD_{440nm} = 0.4 to wash the cells before trypsin treatment. Cells from a 50 ml suspension were pelleted by centrifugation (7000 rpm, 10 min, 4°C) and resuspended in 1.4 ml 10mM HEPES, pH 7.4. To control was added 100µl trypsin buffer, no trypsin and 5mM DTT. To samples were added 100µl trypsin buffer with 20ng trypsin (Promega) and 5mM DTT. Samples and control was incubated at 37°C for 2 hours. The reaction was stopped by transferring tubes to ice. Bacteria were then centrifuged at (8000 rpm, 4°C, 10 min), washed and resuspended to OD440nm = 0.27 in enriched cell medium 5µg/ml chloramphenicol.

Deglycosylation

20ml ON culture was pelleted by centrifugation (7000 rpm, 10 min, 4°C) and washed in 10 ml of PBS. The pellet was resuspended in PBS to OD_{440nm} = 0.4. 1ml cell suspension was pelleted by centrifugation (7000 rpm, 10 min, 4°C). Two different treatments and a negative control were prepared in a 50µl total volume and 10xG7 buffer: 1) 10µl protein deglycosylation mixture (New England Biolabs), 2) 15µl protein deglycosylation mixture, 3) Negative control with dH2O instead of enzyme. Samples and control was incubated at 37°C for 5 hours. Bacteria were pelleted by centrifugation (7000 rpm, 10 min, 4°C), washed two times in PBS and resuspended in enriched RPMI1640 cell medium

Testing membrane integrity after trypsin treatment

Methylococcus capsulatus (Bath) was treated with trypsin as described above and incubated for 0.5 and 2 hours. An experimental control was incubated for 2 hours with buffer only. 5µl pre- and post-treatment samples were diluted in PBS and stained with 4µl propidium iodide. Samples were incubated at room temperature for 5 minutes before testing membrane integrity by flow cytometry. Stained heat killed cells and unstained untreated cells were included as positive and negative staining controls.

Biotin labeling of cell surface proteins

M. capsulatus (Bath) culture (10 ml) from early logarithmic growth phase (OD_{440nm} = 0.4) was harvested by centrifuging at 4000 rpm for 10 min. Pellets were washed three times in ice- cold PBS and then re-suspended in 1 ml ice-cold 10mM HEPES containing 2mM NHS-PEG4 biotin. Samples were incubated on ice for 30 minutes, and the biotin-labeling reaction was terminated by adding 0.5M pH 6.8 Tris-HCl to 50mM final concentration. Bacteria were washed three more times using PBS to remove unbound biotin and then resuspended in 300µl enriched cell medium

Trypsin digestion of bovine serum albumin (BSA)

BSA was diluted 1:10 in four different media (10mM HEPES pH 7.4, Ringers solution pH 6.5, 40% sucrose in PBS pH 6.9 and NMS pH 6.8) to a concentration of 1µg/µl. Before trypsin treatment 20 µl 100mM DTT was added to a 120 µl BSA solution to reduce disulfide bonds and stabilize the proteins during digestion. Trypsin was added in ratio 1:40 (w/w), enzyme to protein. The sample was divided in three and incubated at 30°C for 0.5, 2 and 24 hours, respectively.

2.3.2 SDS-PAGE and Western blotting

SDS-PAGE

All samples were boiled at 100°C for 5 minutes before loading on the gels. SDS-PAGE was performed with a 10% (w/v) running gel and a 4% (w/v) stacking gel. Precision Plus Protein Dual Color Stain was used as a molecular weight standard. Proteins were stained by Coomassie Brilliant Blue R250.

Western blot and enzymatic visualization

A PVDF membrane was briefly activated in methanol then rinsed in distilled water. Gel and membrane was equilibrated in transfer buffer before electrophoretic transfer of proteins to the membrane. The membrane was blocked in TBS (diluted 1:10 in distilled water) containing 5% skimmed milk powder (w/v) and 0.1% Tween 20. Membrane was washed four times in TBS with 0.1% Tween 20 and then incubated in 10 ml blocking buffer with 2.5µl streptavidin-HRP for 1 hour. After incubation the membrane was washed another four times 10 min in TBS. TBS was removed and chemiluminescent HRP-substrate was added.

Biotinylated proteins on the blot were visualized by exposing the blot to Kodak X-Omat film for 30 seconds.

2.3.3 Isolation of cells and differentiation of $CD14^+$ monocytes to dendritic cells

Isolation of peripheral blood lymphocytes from buffy coats

Mononuclear cells (MNC) were isolated by density gradient centrifugation on a Lymphoprep™ (Axis-Shield) gradient according to standard protocols. MNCs were collected with a glass pipette and washed with PBS before centrifuging again for 15 minutes at 700 rpm to get rid of platelets. The MNCs were washed two more times and re-suspended in cell medium.

Positive selection of CD14+ monocytes from PBL

Cells were counted, washed in 2mM EDTA and 0.5 % (w/ v) BSA in sterile PBS, labeled with mouse anti-human CD14 coupled magnetic beads and incubated on ice for 15 min. Labeled cells were isolated using a MACS system (Miltenyi Biotech), washed three times with washing buffer and eluted by adding buffer under positive pressure from a plunger. CD14+ cells were counted in a particle counter then centrifuged at 1300 rpm for 10 minutes. 7000 000 cells were re-suspended in 3 ml of enriched RPMI1640.

Cultivation of cells and differentiation to dendritic cells

Approximately 680 000 CD14+ monocytes in enriched RPMI1640 with 50 ng GC-CSF and 25 ng IL-4 per ml medium, were seeded in cell culture inserts in a 24 well tissue culture tray. Cells were incubated at 37°C in humified atmosphere with 5% CO_2. Medium was changed after four days. After six days cells had differentiated to DCs and were ready for further treatment.

2.3.4 Adherence assays

Cocultivation of trypsin-shaved and deglycosylated <u>M. capsulatus</u> (Bath) *with DCs*

Medium was removed from all wells. Fresh cell medium was added beneath inserts. 300µl untreated, trypsinated or deglycosylated *M. capsulatus*(Bath) resuspended in cell medium was added to inserts and bacteria were co-cultivated with DCs for 3 hours in humidified atmosphere with 5% CO_2

Preparing samples for scanning electron microscope, SEM

Samples were fixed with 4% PFA and 2.5% glutardialdehyde (1:1) for 20 minutes at room temperature and then washed 3x 5min with PBS. Samples were dehydrated in a series of increasing ethanol concentrations: 50%, 70%, 80% and 96% for 15 minutes each, then 3x in 100% ethanol for five minutes followed by critical-point dehydration in a CPD 030 (Bal-Tec) drier. Samples were coated with ~500 Å Pt in a SC7640 sputter-coater and analyzed by an EVO-50 microscope (Zeiss).

SEM and scoring of binding

For each sample five random microscopic fields were captured, each covering approximately 5-6 DC. Images were randomized before binding was scored by two independent viewers.

2.3.5 Identification of secretive proteins in the growth medium of *M. capsulatus*

Growth curve, M. capsulatus (Bath)

Three parallel cultures were grown. 100 ml NMS medium in a 300 ml flask was inoculated with an ON culture to an OD_{440nm} of approximately 0.2. Flasks were sealed with Suba-Seal red rubber closures. 2.5 ml CO_2 and 47.5ml methane was injected through the closure. Cultures were then incubated at 45°C at 200 rpm. Every hour 1 ml culture was withdrawn by a syringe through the closure. Optical density at 440nm was measured and cells were counted by flow cytometry. Average OD_{440nm} from the three cultures was plotted against time to make the growth curve.

Harvesting culture supernatant from early- mid- and late logarithmic/early stationary growth phase

An *M. capsulatus* (Bath) ON culture was pelleted by centrifugation (3000xg at 4°C 10 min). Supernatant was discarded and the pellet resolved in 3 ml fresh medium. The resuspended ON culture was added to 150ml NMS to an OD_{440nm} of 0.2(+/- 0.01). Flask was sealed with Suba-seal closure and 47.5 ml methane and 2.5ml CO2 was added through the closure. Culture was incubated at 45°C and 200rpm.

OD_{440nm} of the culture was monitored and bacteria harvested at OD_{440nm} 0.43, 0.64 and 0.73 (+/- 0.05), corresponding to early logarithmic phase, mid logarithmic phase, and late logarithmic/ early stationary phase. For one parallel another sample was harvested after 24 hours of growth at an OD_{440nm} of 0.75. Harvesting was done by injecting a syringe through the rubber closure withdrawing 25 ml culture. Bacteria were spun down at 3500x g and 4°C for

10 min. 20 ml supernatant was filtered through a 0.2µm low protein-binding filter and transferred to a fresh Nunc tube. 1µl/ml 100mM PMSF was added to the supernatant. Samples were stored at -20°C until TCA precipitation. Three parallel cultures were sampled and analyzed for each growth stage except for lag phase.

TCA precipitation of proteins from supernatants

0.2 mg/ml sodium deoxycholate was added to each sample to disrupt protein interactions and samples were incubated on ice for 30 min. 100% (w/v) TCA was added to 16% end concentration. Samples were vortexed and incubated on ice for one hour. Samples were centrifuged at 16500 rpm at 4°C for 5 minutes. Centrifugation was repeated, and precipitates from 10 ml of supernatant were collected in the same tube for each sample. The pellet was washed with 1 ml of ice cold acetone and spun down at 16500rpm, 4°C for 20 minutes between washes. Pellet was washed with 200 ml ice-cold acetone and spun down again. Acetone was removed, pellet dried at 37°C for 15 minutes and then resuspended in 30µl sample buffer. Proteins were separated by SDS-PAGE.

In-gel digestion and protein extraction

Gel was rinsed in dH_2O prior to cutting. Gel lanes were sliced in five pieces as indicated in (Figure 5.2) and each piece was cut in small cubes before the destaining procedure. Destaing, in-gel digestion and protein extraction was carried out as described by (Shevchenko, Tomas et al. 2006). Prior to saturating the gel pieces with trypsin, 500µl acetonitrile was added to each sample. Gel pieces were covered with 10 mM ammonium bicarbonate, 10% (v/v) acetonitrile buffer containing 13ng/µl trypsin and were incubated for 30 minutes on ice. Fresh trypsin was added and samples incubated for 90 more minutes before adding 20µl of ammonium bicarbonate buffer to each tube. Samples were then digested over night at 37°C and 400 rpm. To one volume of digest were added two volumes extraction buffer (1:2 (v/v) 5 % formic acid: acetonitrile). Peptides were extracted for 15 min at 37°C at 400 rpm and supernatants transferred to fresh tubes. All samples were dried in a vacuum centrifuge, then resolved in 10µl 0.1 % trifluoroacetic acid; purified and desalted using ZipTipC18 (Millipore) and then eluted in 50%ACN/ 0,1% TFA.

LC-MS and Database searching

Tandem mass spectra were extracted. Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK;

version Mascot). Mascot was set up to search the M_capsulatus_bath database (unknown version, 2925 entries) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 10.0 PPM. S-carbamoyl-methylcysteine cyclization of the N-terminus, deamidation of asparagine and glutamine, oxidation of methionine, acetylation of the N-terminus and acrylamide adducts of cysteine were specified in Mascot as variable modifications.

Criteria for protein identification

Scaffold (version Scaffold_3.3.3, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 80.0% probability as specified by the Peptide Prophet algorithm (Keller, Nesvizhskii et al. 2002). Keller, A et al. Anal. Chem. 2002; 74(20):5383-92). Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, Keller et al. 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

2.3.6 Genome-wide *in silico* analysis of *M. capsulatus* (Bath) subcellular protein localization

Proteome sequence and prediction strategy

Amino acid FASTA sequences from 2956 proteins were acquired by downloading all proteins linked to the *M.capsulatus* (Bath) genome from the NCBI protein database.

Amino acid sequences were analysed by online available prediction servers in a strategy similar to the one used by Romine (2011)

Identifying Phage proteins

Ward et al recognized two putative prophages in the *M. capsulatus* (Bath) genome, one spanning from MCA2632 to MCA2689 and the other, a Mu-prophage like element, spanning from MCA2900 to MCA2959. Proteins encoded in prophage loci were annotated “prophage” or “Mu- like prophage”.

Identifying secretion system components in M. capsulatus (Bath)

The JCVI CMR Genome Properties Tool was used to search the *M. capsulatus* (Bath) genome for: "sec system preprotein translocase", " Tat(Sec-independent) protein export", "Lipoprotein localization system lolABCl", "Type I secretion", "Type II secretion", "Type III secretion", "Type IV secretion", "Type VI secretion", "Type VII" and "signal recognition system, bacterial" and "outer membrane protein assembly complex".

Identified secretion system components were assigned cellular location with support from relevant literature (Stathopoulos, Hendrixson et al. 2000; Thanassi and Hultgren 2000; Paetzel, Karla et al. 2002; Lee, Tullman-Ercek et al. 2006; Tseng, Tyler et al. 2009; Tommassen 2010; Korotkov, Sandkvist et al. 2012)

Identifying outer membrane beta barrels, BOMP

All proteins in the *M. capsulatus* (Bath) proteome were analysed by BOMP. All proteins predicted by BOMP to be β-barrels (BBs) were searched for Pfam or TIGRFAM domains. If a beta barrel domain was found, the protein was predicted to be an outer membrane β-barrel protein (OM BB).

Identifying lipid anchored proteins, LipoP and Lipo

FASTA sequences of all *M. capsulatus* proteins were analyzed by LipoP and Lipo. All proteins predicted by LipoP or Lipo to contain a lipoprotein signal peptide (LP SP) were searched for Pfam/TIGRFAM domains associated with lipoproteins (LPs). If a) a lipoprotein domain was found in a protein predicted by either program to be a LP, or b) both prediction programs predicted a protein to be a LP and no OMBB domain was found the protein was defined as a LP without further analysis.

A regular expression pattern was constructed based on lipobox motifs from *M.capsulatus* (Bath) LPs. A regular expression was constructed to allow any residue found in any of the *M. capsulatus* (Bath) lipoproteins in position -3 to +1 after the cleavage site. All proteins not defined as LP after criterium a) or b), but annotated as LP; predicted to be LP by only one program or containing a LP domain, was searched with the regular expression. If the regular expression was found the protein was defined as a lipoprotein.

All proteins defined as LPs were manually checked: If an aspartic acid was found in position +2 after the cleavage site of a lipoprotein, it was assigned to the inner membrane (IM), if not it was assigned an OM location.

Identifying proteins secreted by the Twin-arginine translocation pathway

M. capsulatus (Bath) proteins were screened for twin-arginine signal peptides (Tat SPs) using the TatP and Tatfind prediction servers and by searching for proteins with the TIGRFAM1409 domain. All proteins predicted to contain a Tat SP were manually screened: If the cut-site was found N-terminal from the twin-arginin motif the prediction was considered a false positive. If the cut-site was found within the twin-arginin motif the prediction was also considered false positive.

Predicting integral membrane proteins, TMHMM and Phobius

TMHMM and Phobius was used to search for transmembrane helices (TMH) in the *M. capsulatus* (Bath) proteome. If a) Phobius predicted more than one TMH, and TMHMM predicted more than two TMH, or b) two TMH were predicted and less than 10 of the amino acids found in helices were found within the first 60 amino acids, the protein was assigned to the IM.

If programs found one TMH, less than one TMH or did not agree, the proteins were analyzed further.

Predicting Sec translocation signal peptides SignalP 3.0, Phobius

Gram-negative bacteria were chosen as an organism group, and both neural network and hidden Markov model selected. All proteins that did not contain a Tat SP, a LP SP or more than one TMH were searched for signal peptides by the SignalP 3.0 and Phobius prediction tools.

If SignalP 3.0 NN algorithm and HMM algorithm both predicted the presence of a signal peptide, or if one SignalP 3.0 algorithm and Phobius both predicted a SP, the protein was defined as secreted. All secreted proteins were screened for domains indicating OM location or periplasmic location and was assigned localization accordinglyProteins not containing an OM domain and not found to contain one TMH by Phobius, was assigned to the periplasmic space.

If only one SignalP 3.0 algorithm or Phobius predicted a signal peptide the protein was searched for OM or periplasmic domains. If such a domain was found, the protein was considered to be secreted. If no OM or periplasm domain was found, the protein was not regarded as a secreted protein.

Membrane anchored proteins

Secreted proteins predicted by Phobius to hold one TMH were screened for IM domains. Proteins containing an IM domain were predicted to be IM anchored. If not containing an IM domain, proteins were predicted to be periplasmic.

Non-classical secretion

All proteins were analyzed by SecretomeP. Proteins predicted to be secreted, that had not previously been predicted to contain SP by LipoP, TatP, SignalP or Phobius were regarded candidates for non-classical secretion. These were all screened for domains indicating IM, periplasmic or OM location and assigned SCL accordingly. If no such domain was found, the protein was predicted not to be secreted and was assigned to the cytoplasm.

2.3.7 Identification of virulence and adhesion proteins in the *M.capsulatus* proteome

Conserved domain and homology search

All predicted OM BB, lipoproteins and extracellular proteins were searched for domains, homologs and annotation information indicating involvement in host interactions. Conserved Domain (CD) Blast), Pfam and the comprehensive microbial resources database (CMR) was used to identify conserved domains. Blast was used for sequence-based database searches. E value < 0.01 was considered statistically significant. Blastp was used to find informative homologs of *M. capsulatus* (Bath) proteins.

Uniprot, National Center for Biotechnology Information, (NCBI) and CMR was used to find general information of *M. capsulatus* (Bath) proteins. The Virulence Factors of Pathogenic Bacteria (VFPB) database and literature was searched for proteins involved in virulence or host interactions, and the *M. capsulatus* (Bath) proteome searched for homologs of, or domains associated with such proteins.

2.3.8 Texas/Bath pili protein comparison

Pairwise alignment of M. capsulatus Texas and Bath amino acid sequences

The newly annotated genome of *M. capsulatus* (Texas) (Kleiveland et al. 2012a) was examined for type IV pili genes. Nucleic acid sequences were translated into amino acid sequences by EMBOSS. The closest protein homolog in the Bath strain was found by a Blastx search and a pairwise alignment was conducted by Emboss Needle using the BLOSSUM90 scoring matrix, a matrix appropriate for sequences with >90% sequence similarities.

Searching for proteins using a translated nucleotide sequence and identifying templates for 3D-modeling of proteins.

Blastx was used to search protein databases for proteins from the *Methylococcus capsulatus* (Bath) genome using a translated nucleotide sequence when trying to identify the closest homologue of Texas-encoded proteins in the Bath strain. Blastp was used to search the PDB database for templates for modeling 3D-structure of proteins.

Swiss model

The putative tip-adhesin of the T4P complex of *M. capsulatus* Texas and Bath was aligned with a template and cropped to the length of the alignment. The resulting sequence was modeled in SwissModel using automated mode. The resulting model was viewed in PyMole.

3 Results

3.1 *Methylococcus capsulatus* (Bath) interacts with human $CD14^+$ monocyte-derived dendritic cells

Preliminary studies have shown that *Methylococcus capsulatus* (Bath) has the ability to adhere to human $CD14^+$ monocyte-derived dendritic cells (DC) but not $CD14^+$ monocytes in vitro (Kleiveland et al., manuscript in preparation). In order to confirm these results and study the pattern of binding, mononuclear leukocytes were first isolated from human peripheral blood. Then CD14+ monocytes isolated by immunomagnetic cell sorting, seeded on filters and cultivated in the presence of 25ng/ml IL-4 and 50ng/ml GM-CSF to stimulate the cells to differentiate towards dendritic cells. After 6 days when the cells had developed a dendritic morphology, they were incubated with $3x10^7$ *M. capsulatus* (Bath) per 700.000 DC for 3 hours.

As previously observed, *M. capsulatus* (Bath) adhered to DC (Figure 3.1 D-F). In earlier studies incubation with bacteria was carried out for approximately 2 hours (Figure 3.1 A-C). A comparison of images from these experiments with images taken during this study shows a slightly different binding pattern. Although some cells appeared with bacteria evenly distributed over the cell surface, there was a tendency for the bacteria to adhere in clusters on specific areas of the DCs. This binding patter was in contrast to a more dispersed distribution of *M. capsulatus* (Bath) on the cell surface in experiments with shorter incubation times. On some of the images membran protrusions of the DC is seen assosiated with adhering bacteria (Figure 3.1 C,F)

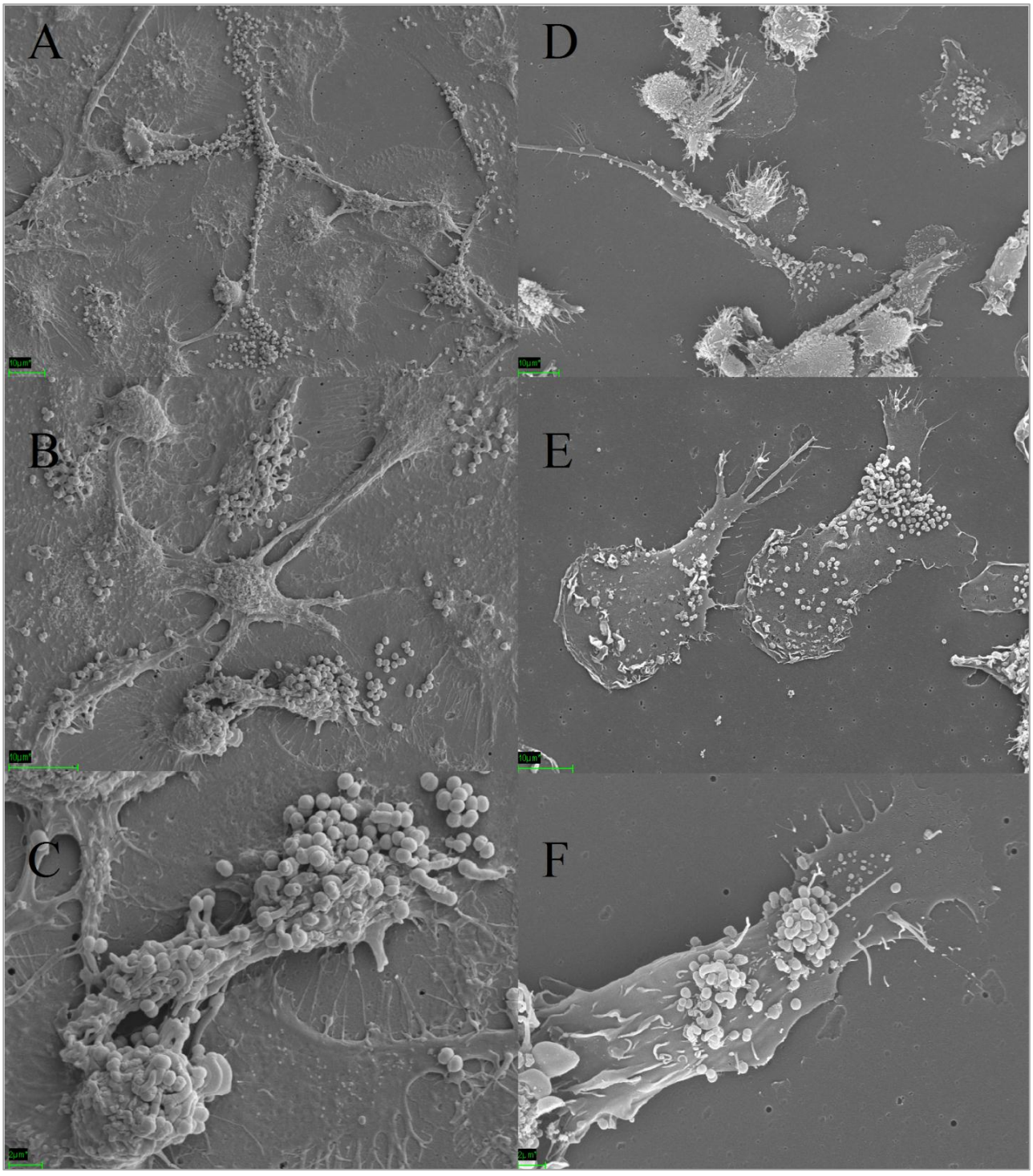


Figure 3.1 Scanning electron microscope images of *M. capsulatus* (Bath) adhering to $CD14^+$ monocyte-derived dendritic cells after A-C) 2 hours of co-incubation and D-F) 3 hours co-incubation. In image C and F protrusions of the DCs are seen beneath single bacteria.

3.1.1 Trypsin shaving and deglycosylation of the *M capsulatus* (Bath) surface

Bacterial adhesins may in principle be any type of organic molecule, protein, lipid or carbohydrate. Trypsin is a serine protease that cleaves polypeptides on the C-terminal side of lysine and arginine aa residues if no proline is found C-terminal to the cleavage site. Trypsin can therefore be used to "shave" proteins off the outer surface of the bacteria before incubation with DC to investigate whether surface proteins of the bacteria are involved in the interaction.

3.1.2 Trypsin is most efficient in HEPES buffer

Different buffers have been used in other studies and the choice of buffer has been shown to greatly influence the rate of cell lysis after trypsination in Gram-negative bacteria (Walters & Mobley 2009). In order to find an appropriate buffer for trypsin-shaving of *M. capsulatus* (Bath), trypsin efficiency was tested in three different buffers commonly used for bacteria and in the *M.capsulatus* (Bath) growth medium, NMS. To examine the efficiency of trypsin in different buffers a tryptic digest of BSA was prepared in three different buffers: 40% sucrose in PBS (pH 6.9), 10mM HEPES (pH 7.4) and Ringer´s solution (pH 6.5). Trypsin efficiency was also controlled in the *M. capsulatus* (Bath) growth medium, NMS (pH 6.8).

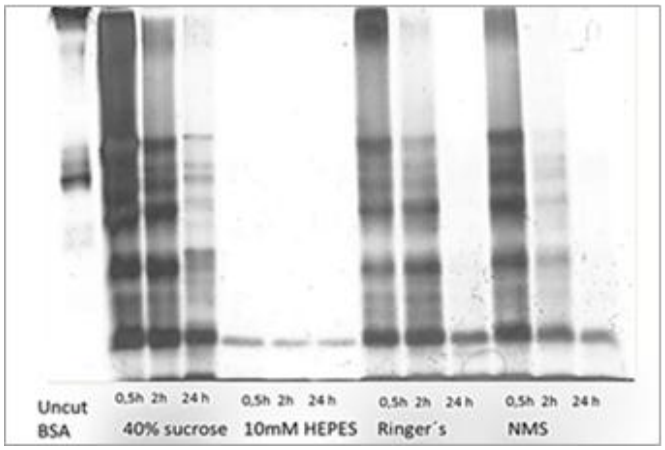


Figure 3.2. Tryptic digest of BSA in 40% sucrose, 10 mM Hepes, Ringer´s solution and nitrate mineral salts medium. Uncut BSA and BSA digestion reactions after 0.5 hour, 2 hours and 24 hours are shown.

Trypsin was active in all buffers tested (Figure 3.2), but required longer incubation times in sucrose, Ringers solution and NMS compared to 10mM HEPES buffer. Longer incubation times will most likely result in increased bacterial lysis, and HEPES was therefore used in the shaving experiments.

3.1.3 Trypsin shaves proteins of the cell surface

Before testing adherence of trypsinated bacteria to DCs it was important to establish that proteins were in fact removed from the bacterial surface. To test the trypsin shaving efficiency, intact bacteria were labeled with membrane-impermeable biotin before trypsin treatment. 0.5 and 2 hours trypsin-treatment was tested. Labeled cell surface proteins were separated on SDS-PAGE, transferred to a PVDF membrane by Western-blotting and detected with streptavidin-HRP and chemiluminescence. Biotin labeled proteins were shown to be removed from the surface of the cell (Figure 3.3), however a subset of proteins appeared to be resistant to trypsinshaving.

Figure 3.3 Biotin labeled cell-pellets after 0.5 and 2 hours of trypsin treatment. An untrypsinated control was included.as control

3.1.4 *Methylococcus capsulatus* remains intact after shaving

Gram-negative bacteria have a thinner cell wall than Gram-positive bacteria and has proven to be difficult to trypsinate without lysing the bacterial cells. In order to establish that membranes were intact after treatment, trypsin-treated bacteria were stained with Propidium iodide (PI), a fluorescent DNA-intercalating molecule that is impermeate to intact membranes. An experimental control was included. Heat killed and untreated bacteria were included as positive and negative staing control. Flow cytometry showed that trypsin-treated ot bacteria did not stain with PI after treatment, demonstrating that the bacterial membrane is not compromised (Figure 3.4 and 3.5)

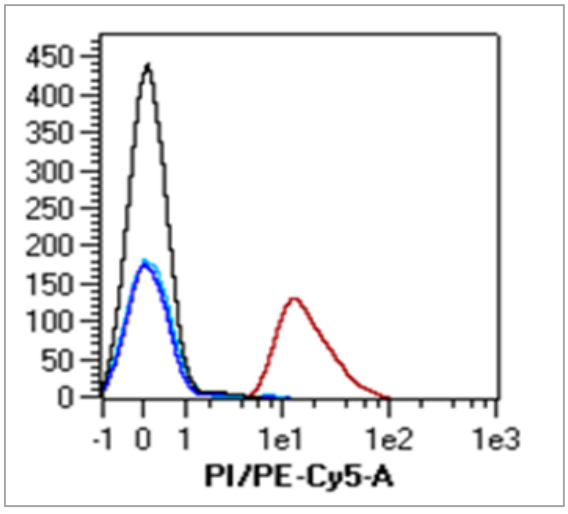


Figure 3.4 Effect of 0.5 hour (light blue), 2 hours (dark blue) incubation with trypsin on viability of M. capsulatus (Bath) as indicated by PI staining. Untreated cells (black) and cells boiled for 1 hour (red) is included as negative and positive controls, respectively.

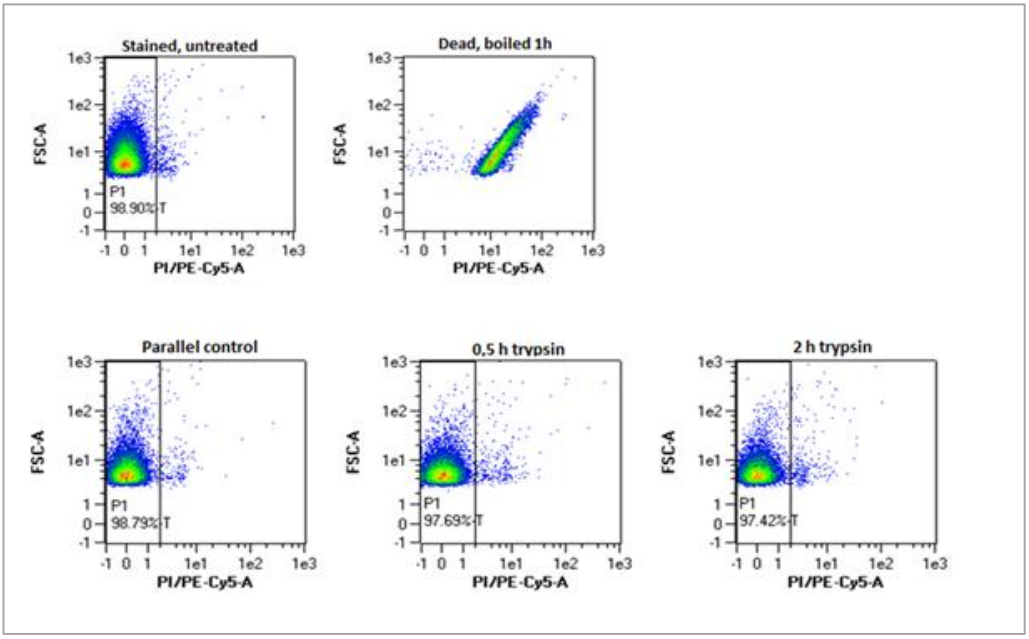
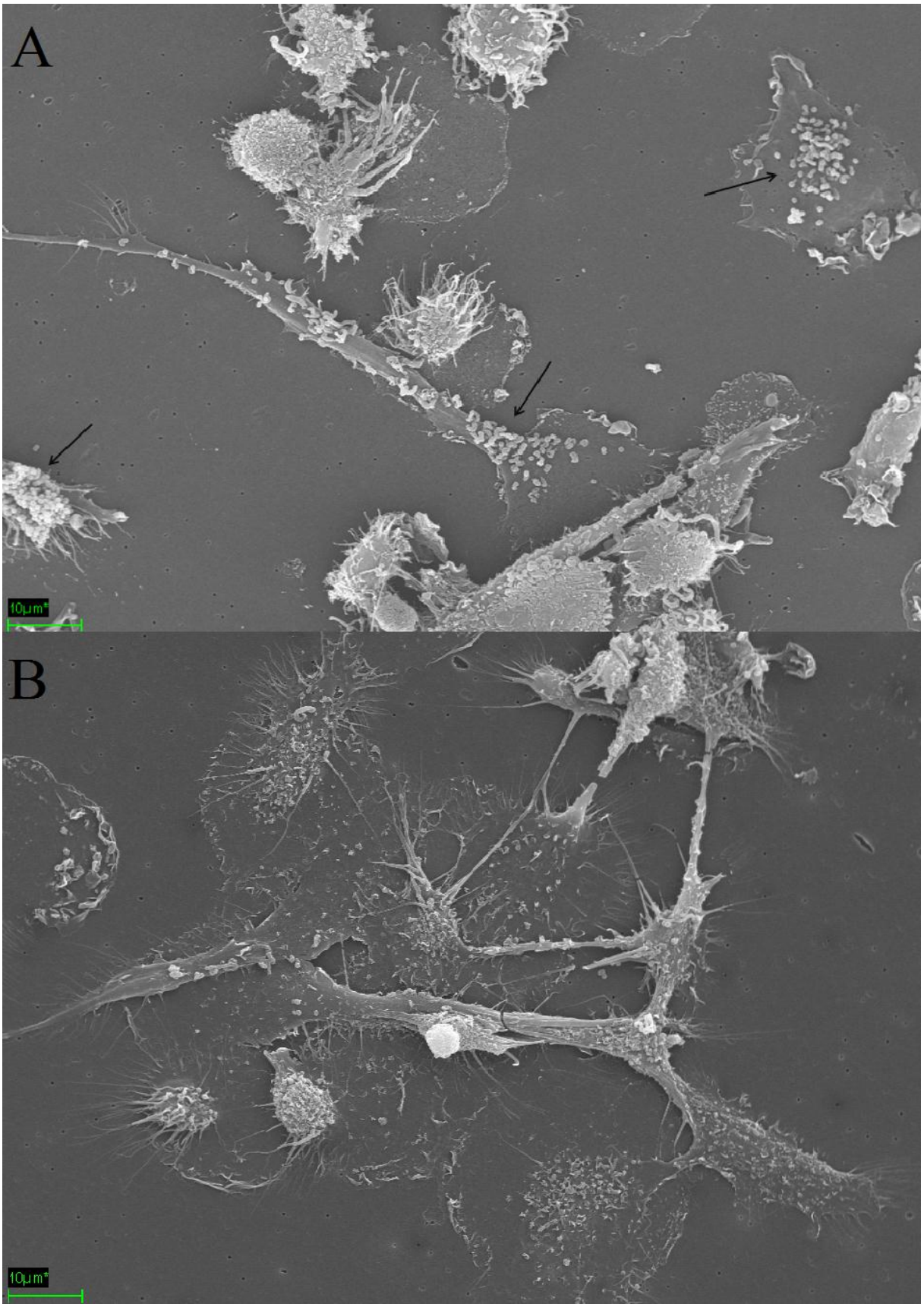


Figure 3.5 Dotplot of *Methylococcus capsulatus* (Bath) stained with PI after A) no treatment (living), B) boiled for 1h (dead) and after C) 0,.5 and D) 2 hours of incubation with trypsin. Heat -killed bacteria are stained by PI while the majority of trypsin-treated bacteria remains unstained and are found within the same area as untreated cells after both incubation lengths.

3.1.5 Surface proteins mediates adhesion of *M. capsulatus* (Bath) to dendritic cells

Theoretically, bacterial adhesins may be any kind of surface molecule. To test whether proteins are involved in the adhesion of *M. capsulatus* (Bath) to DC, surface proteins were shaved off the surface of bacteria prior to co-cultivation with the DC. Trypsin-treated and untreated bacteria were co-cultivated with DC in the presence of chloramphenicol. Chloramphenicol was included to inhibit *de novo* bacterial protein synthesis and thereby reappearance of bacterial surface proteins during co-cultivation. Images of five randomly selected microscopic fields were captured. Binding was blind-scored by two independent viewers. Trypsin-treated *M. capsulatus* showed an impaired ability to bind DC (Figure 3.6) indicating that surface protein(s) are involved in adherence.



Figur 3.6, SEM images of *M. capsulatus* (Bath) co-incubated with DC. A) Untreated bacteria adhering to DC. B) Trypsin shaved bacteria show an impared ability to adhere to DC.

3.1.6 *M. capsulatus* (Bath) treated with deglycosylation enzyme mix shows enhanced adherence to DC

Glycans, carbohydrate moieties attached to proteins, lipids or other organic components adds to the structural diversity of bacterial surfaces and have the potential to define or mask interacting sites. To test whether glycans is involved in adhesion, intact *M. capsulatus* was treated with a deglycosylation enzyme mixture before incubation with DCs to test if glycans, carbohydrates attached to proteins, lipids or other molecules are involved in adhesion *M. capsulatus* showed enhanced adherence after deglycosylation, and the effect was strongest for bacteria treated with the highest concentration of enzyme mixture (Figure 3.7)

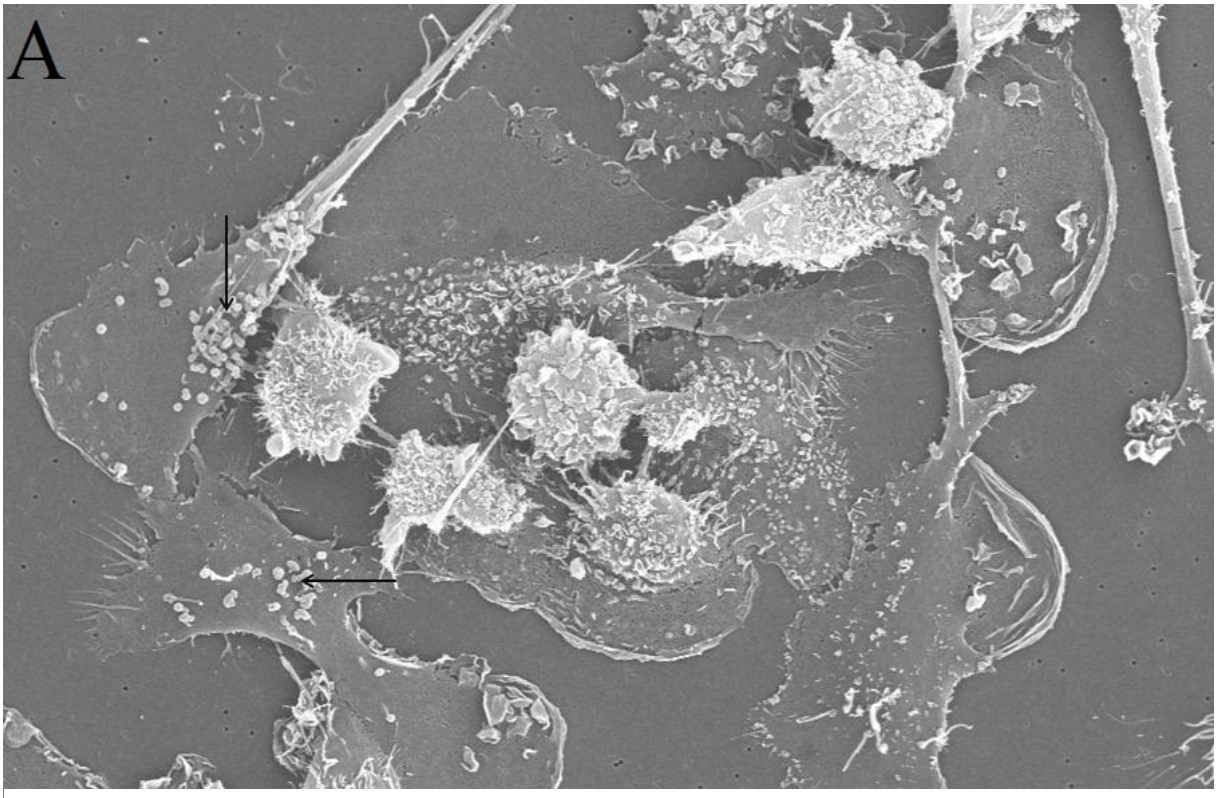


Figure 3.7 A) SEM images of M. capsulatus (Bath) incubated with DCs after different treatments: Negative control

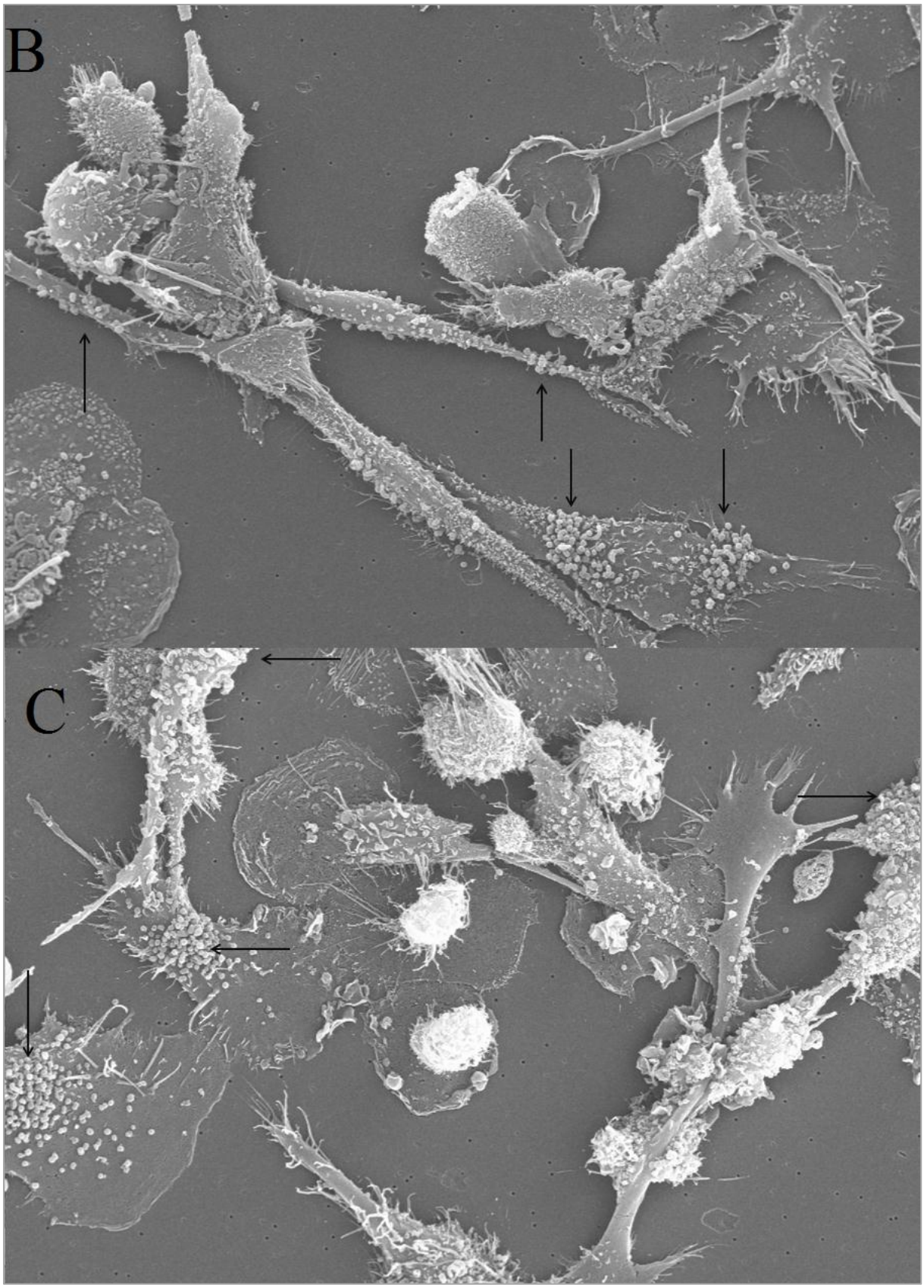


Figure 3.7 B) and C) SEM images of *M. capsulatus* (Bath) incubated with DCs after different treatments: B)Treated with 10µl deglycosylation mix, C) Treated with 15µl deglycosylation mix.
Bacteria showed improved adherence after deglycosylation

3.2 Genome-wide sub-cellular location prediction

Surface-associated proteins, outer membrane (OM) proteins, lipoproteins and extracellular proteins often play important roles in adherence, modulation of host immune responses and transport of virulence factors. To establish which proteins are present on the bacterial surface, the proteome of *M. capsulatus* (Bath) was analysed *in silico*, and subcellular location of proteins were predicted by the strategy shown in (Figure 3.8). To understand their function, all proteins predicted to be surface-related, were examined for domains, annotation and homology to proteins of known function. Subcellular location prediction for all 2956 M. capsulatus (Bath) proteins can be found on a compact disc provided with the thesis

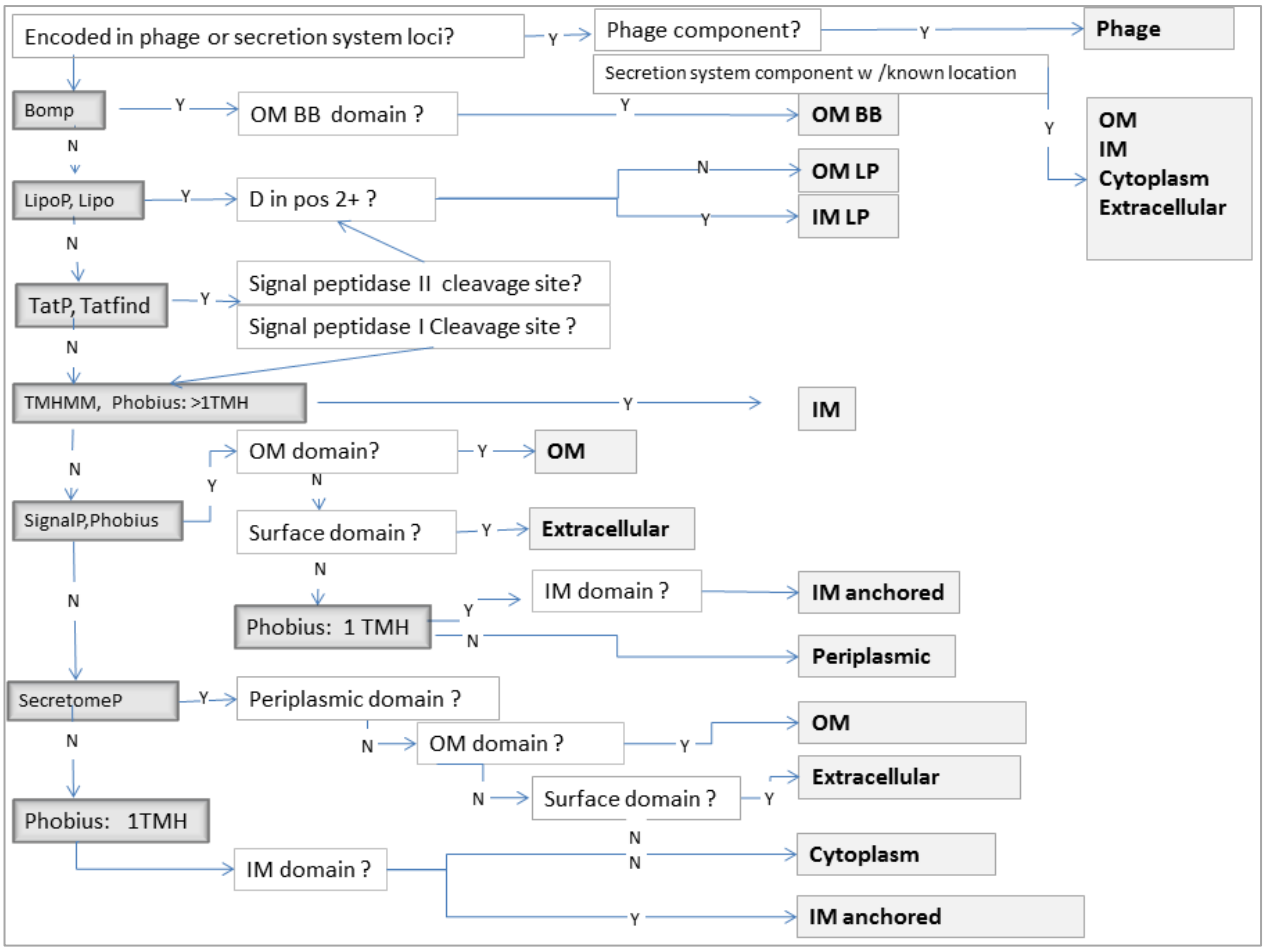


Figure 3.8 Decision chart for predicting subcellular location of *Methylococcus capsulatus* (Bath) proteins

Proteins predicted to be integral to the OM

OM β-barrels (OM BB) are important surface structures, antigens and may be involved in transport and adhesion. 39 proteins were predicted to be integral OM proteins (Table 3.1). Several of the identified OM proteins are involved in major OM protein secretion systems or the OM protein insertion system: TolC, the OM pore of the T1SS, GspD, the secretin of T2SS and Omp85 are central components in the OMP assembly machinery. MCA0304, a protein containing a POTRA domain showing homology to ShlB type polypeptide transport protein, is likely to be a T5SS transporter.

Two proteins identified are related to pili assembly or transmembrane transport: MCA0329, a fimbrial assembly protein and PilQ, which is a T4P Secretin. MCA0304 fimbrial biogenesis protein contains an usher domain and a PapC C-terminal domain and showed homology to CU pili usher proteins.

Twelve proteins, including TolC, contained OM efflux domains and are therefore expected to belong to the OM efflux protein family, a family of proteins that form trimeric channels that allows export of a variety of substrates in Gram-negative bacteria. Six proteins were found to be porins, or putative porins, and believed to be involved in diffusion of hydrophilic molecules (cations/anions).

Some of the integral membrane proteins are involved in signaling or response to organic substances: Six proteins were found to be TonB-dependent receptors involved in uptake of large substrates and signaling of substrate availability to the interior of the cell. One protein was a tolerance protein likely conferring resistance to organic solvents.

Other proteins found are believed to have structural roles: OmpA-like protein: a major structural protein in the OM of Gram-negative bacteria with adhesive properties in some species, MopB protein containing an OmpA domain and OmpH, another major structural protein, were all found in the dataset.

Table 3.1 39 Proteins predicted by *in silico* analysis to be integral to the outer membrane of *M. capsulatus* (Bath)

Locus	Protein name	
MCA1277	Outer membrane protein TolC	T1SS OM pore
MCA1123	General secretion pathway protein D	T2SS Secretin
MCA2226	Hypothetical protein MCA2226	T5SS Vb/ Vctransporter
MCA2447	OMP85 family Outer membrane protein	OM protein insertion
MCA2446	Outer membrane protein OmpH	OM protein transport (chaperone)
MCA0329	Fimbrial assembly protein PilQ	Type IV pilus Secretin
MCA0304	Fimbrial bioGenesis protein	Chaperone-Usher pili usher
MCA1230	OmpA family protein	OM integrity (OmpA domain)
MCA3103	Outer membrane protein MopB	OM integrity (OmpA domain)
MCA2074	TonB domain-containing protein	TonB dependent receptor
MCA2751	TonB domain-containing protein	TonB dependent receptor
MCA1957	TonB-dependent receptor	TonB dependent receptor
MCA2321	TonB-dependent receptor	TonB dependent receptor
MCA2180	TonB-dependent receptor	TonB dependent receptor
MCA0440	TonB-dependent receptor	TonB dependent receptor
MCA0900	Hypothetical protein MCA0900	Porin (O or P)
MCA0947	Hypothetical protein MCA0947	Porin (O or P)
MCA1375	Hypothetical protein MCA1375	Porin (O or P)
MCA0601	Organic solvent tolerance protein	Organic solvent tolerance
MCA2232	Outer membrane efflux family protein	Outer membrane efflux protein
MCA2816	Outer membrane efflux family protein	Outer membrane efflux protein
MCA0284	Outer membrane efflux family protein	Outer membrane efflux protein
MCA1937	Outer membrane efflux protein	Outer membrane efflux protein
MCA1723	Outer membrane efflux protein	Outer membrane efflux protein
MCA2110	Outer membrane efflux protein	Outer membrane efflux protein
MCA1768	Outer membrane heavy metal efflux protein	Outer membrane efflux protein
MCA2262	Outer membrane heavy metal efflux protein	Outer membrane efflux protein
MCA2154	Outer membrane protein CyaE	Outer membrane efflux protein
MCA1362	Outer membrane efflux family protein	Outer membrane efflux protein
MCA1595	Multidrug efflux system Outer membrane subunit	Outer membrane efflux protein
MCA0491	Hypothetical protein MCA0491	Possible porin
MCA1701	Hypothetical protein MCA1701	Possible porin
MCA2112	Hypothetical protein MCA2112	Possible porin
MCA0951	Hypothetical protein MCA0951	
MCA2108	Hypothetical protein MCA2108	
MCA2575	Hypothetical protein MCA2575	
MCA0145	Hypothetical protein MCA0145	
MCA0945	Hypothetical protein MCA0945	
MCA1174	Hypothetical protein MCA1174	

a Putative function as indicated by conserved domains, annotation information and homology to proteins of known function

3.2.1 Proteins predicted to be lipid-anchored to the inner or outer face of the OM

69 Proteins were predicted to be lipoproteins (LPs) anchored to the OM. LPs were found to be involved in transport of various substrates, cell wall and OM assembly and integrity maintenance, protein fate, pili biogenesis, electron transport and metabolic processes (table 3.2).

Table 3.2 31 proteins predicted to be OM lipoproteins with putative function.

Locus	Protein name	[a]Putative function
MCA0328	type 4 fimbrial biogenesis protein PilP	Pili biogenesis
MCA2888	type IV pilus biogenesis protein PilF	Pili biogenesis
MCA0523	peptidyl-prolyl cis-trans isomerase, FKBP-type	Protein folding
MCA1054	molecular chaperone LolB	Protein transport
MCA1643	peptidyl-prolyl cis-trans isomerase	Protein transport
MCA1766	heavy metal efflux system protein	Transmembrane transport
MCA1596	HlyD family multidrug efflux protein	Transmembrane transport
MCA1297	RND family efflux transporter MFP subunit	Transmembrane transport
MCA2814	RND family efflux transporter MFP subunit	Transmembrane transport
MCA0558	hypothetical protein MCA0558	Chaperone for OM pore component
MCA1264	peptide ABC transporter periplasmic peptide-binding protein	Transport
MCA1430	capsular polysaccharide export protein	Polysaccharide transport
MCA1452	rare lipoprotein B	OM assembly
MCA2891	lipoprotein	BB insertion(BamB)
MCA2622	small protein A	BB insertion(BamE)
MCA0971	competence lipoprotein ComL	BB insertion(ComL/BamD)
MCA1819	OmpA domain-containing protein	OM integrity
MCA1231	hypothetical protein MCA1231	OM integrity
MCA0637	hypothetical protein MCA0637	Peptidoglycan crosslinking
MCA1403	membrane-bound lytic murein transglycosylase	Peptidoglycan metabolic process
MCA2114	murein hydrolase B	Peptidoglycan metabolic process
MCA2028	membrane-bound lytic murein transglycosylase	Peptidoglycan metabolic process
MCA2403	membrane-bound lytic murein transglycosylase	Peptidoglycan metabolic process
MCA1826	hypothetical protein MCA1826	Peptidoglycan metabolic process
MCA2076	vacJ lipoprotein	Structural integrity
MCA0129	thiamin biosynthesis lipoprotein ApbE	Thiamine synthesis
MCA1585	fatty acid cis/trans isomerase	fatty acid metabolic process
MCA0049	cellulose-binding domain-containing protein	Carbohydrate metabolic process
MCA2221	acid phosphatase	Metabolic process
MCA2185	lipoprotein	Electron transport
MCA0948	cytochrome c family protein	Electron transport

[a] Putative function as indicated by conserved domains, annotation information and homology to proteins of known location

38 LPs proteins did not contain conserved domains, informative domains or annotation information, and the function of those proteins therefore remains unknown (table 3.3).

Table 3.3 38 proteins predicted to be lipoproteins with unknown functions

Locus	**Protein name**	**Locus**	Protein name
MCA2789	NLP/P60 family protein	MCA2409	hypothetical protein MCA2409
MCA0747	rare lipoprotein A	MCA0185	lipoprotein
MCA2184	Slp family outer membrane lipoprotein	MCA0311	lipoprotein
MCA1714	surface protein-like protein	MCA0672	lipoprotein
MCA1992	TPR domain-containing protein	MCA0717	lipoprotein
MCA0188	hypothetical protein MCA0188	MCA1839	lipoprotein
MCA0410	hypothetical protein MCA0410	MCA1946	lipoprotein
MCA0414	hypothetical protein MCA0414	MCA2763	lipoprotein
MCA0454	hypothetical protein MCA0454	MCA0171	lipoprotein
MCA0469	hypothetical protein MCA0469	MCA0356	lipoprotein
MCA0944	hypothetical protein MCA0944	MCA0364	lipoprotein
MCA1015	hypothetical protein MCA1015	MCA0406	lipoprotein
MCA1029	hypothetical protein MCA1029	MCA0605	lipoprotein
MCA1577	hypothetical protein MCA1577	MCA0981	lipoprotein
MCA1693	hypothetical protein MCA1693	MCA1904	lipoprotein
MCA1881	hypothetical protein MCA1881	MCA2144	lipoprotein
MCA1948	hypothetical protein MCA1948	MCA2179	lipoprotein
MCA2026	hypothetical protein MCA2026	MCA2510	lipoprotein
MCA2245	hypothetical protein MCA2245	MCA2764	lipoprotein

3.2.2 Proteins predicted to be extracellular

18 proteins were predicted to have extracellular location (Table 3.4). Four proteins (MCA0086, MCA1510 and MCA0087) were found to contain pili-related domains. A protein BLAST search showed these to be homologs of pilins or tip-adhesins of a type IV pili system (T4P). Two proteins (MCA0303, MCA0306) contained 'spore coat protein u domains', a domain found in spore coat proteins and chaperone-usher pili assembly systems. A BLAST search revealed these to be homologous to a sigma-fimbria pilin and a sigma-fimbria tip adhesion, respectively. One protein contained domains typical of hemagglutinins and was found to be a homolog of filamentous hemagglutinin (FHA). MCA2328 contained a domain of unknown function. Homologs of this protein were annotated as outer membrane adhesin-like proteins.

MCA2795 was homolog to macrophage migration inhibitory factor (MIF) and contained a MIF domain. In a BLAST search against the human genome the *M. capsulatus* (Bath) MIF showed highly significant homology ($1x10^{-17}$) and 83% query coverage with human MIF. MCA1752 contained four alpha-2-macroglobulin domains and homology to alpha-2-macroglobulin. MCA0312 contained a domain known as 'signaling molecule that associates with mouse pelle-like kinase' (SIMPL). No homologs of known function were found for this protein. Pfam showed MCA0874 to contained 8 Sel1 repeats. This protein is homologous to a sodium-type flagellar motor component and a large number of proteins of unknown function. MCA0553 contained an amylo-alpha-1,6-glucosidase domain and a F5/8 (also known as discoidin or DS domain). Pfam found a receptor for egg jelly (REJ) domain in two proteins (MCA2224, MCA2974). The first of these, MCA2224 was found by a conserved domain (CD) BLAST to contain an early set domain, a domain that may be related to the immunoglobulin and/or fibronectin type III superfamilies. The protein showed significant homology to glycoside hydrolases (chitinases), but only over a limited region of the query MCA0875 sequence. The other REJ- domain protein contained five extracellular repeat domains in addition to the REJ domain and fragmented regions of homology to proteases. Three other exoproteases were found: MCA0875 (containing a peptidase inhibitor I9 domain and peptidase s 8 family domains and a subtilase domain, MCA1217 contains a M36 propeptide and a M36 peptidase (also known as fungalysin) domain and MCA2990, containing a peptidase M22 domain and homology to O-sialoglycoprotein endopeptidases. MCA2589 contained a copper-binding domain and was recognized as MopE by BLAST search.

Table 3.4 18 Proteins predicted to be extracellular

Locus	Protein name	Pfam domain	[a]Domain information
MCA0086	type 4 Fimbrial biogenesis protein PilE	PF07963	Prokaryotic N-terminal methylation motif
MCA1510	Fimbrial protein	PF07963	Type IV pilin N-term methylation site GFxxxE
MCA0087	Hypothetical protein	PF05567	Neisseria PilC beta-propeller domain
MCA0303	Lipoprotein	PF05229	Spore Coat Protein U domain
MCA0306	Spore coat protein, late developmental	PF05229	Spore Coat Protein U domain
MCA2227	Hemagglutinin-like protein	PF05860/PF12545	FHA family outer membrane protein/haemagglutination activity domain
MCA2328	Hypothetical protein	PF13313	Domain of unknown function (DUF4082)
MCA2795	Phenylpyruvate tautomerase	PF01187	Macrophage migration inhibitory factor (MIF)
MCA1752	Hypothetical protein	PF01835/PF01835/PF07703/PF00207	Alpha-2-macroglobulin family N-terminal region/A-2-M family N-terminal region/A-2-M family
MCA0312	Hypothetical protein	PF04402	Protein of unknown function (DUF541)
MCA0874	Hypothetical protein	PF08238	Sel1 repeat (X8)
MCA0553	Discoidin domain-containing protein	PF00754/PF06202	F5/8 type C domain/Amylo-alpha-1,6-glucosidase
MCA2224	PKD domain-containing protein	PF02010/PF00801	REJ domain/PKD
MCA2974	PKD domain-containing protein	PF02010	REJ domain
MCA0875	Serine protease	PF00082/PF05922	Peptidase, S8/S53 family/Subtilisin N-terminal Region
MCA1217	Metalloprotease	PF07504/PF03413/PF02128	Fungalysin/Thermolysin Propeptide Motif/Peptidase propeptide and YPEB domain/Fungalysin metallopeptidase (M36)
MCA2990	O-sialoglycoprotein endopeptidase	PF00814	Glycoprotease family
MCA2589	Surface-associated protein	PF11617	Protein metal binding site

[a] Conserved domains identified by Pfam

3.3 Secreted proteins released by *M. capsulatus* (Bath) during growth

In silico analysis is based on searching for sequence motifs and patterns and may miss biologic information not directly coded in the amino acid sequence. An experimental approach was therefore taken to examine the secretome of *M.capsulatus* (Bath) *in vivo*, and the results were compared with the results of the *in silico* analysis. Fresh cultures of *M. capsulatus* (Bath) were collected in early, mid and late exponential growth phase and during lag phase growth. Proteins released to the growth medium were collected and identified by LC-MS analysis (Table 3.5).

Table 3.5 Proteins identified in the growth medium of *M. capsulatus* during early-, mid- and late exponential growth and during lag phase.

		Early			Mid			Late			[a]Lag	Predicted
Locus ID	Protein name	1	2	3	1	2	3	1	2	3		SCL
MCA0049	Cellulose-binding domain protein				■							OM LP
MCA0155	Putative uncharacterized protein		■		■	■	■	■	■	■	■	Periplasm
MCA0338	Cytochrome c5530 family protein				■	■	■	■	■	■	■	Periplasm
MCA0461	CBS/GGDEF/EAL domain protein O	■										Cytoplasm
MCA0707	60 kDa chaperonin 1				■	■		■	■	■	■	Cytoplasm
MCA0779	Methanol dehydrogenase protein		■	■	■	■	■	■	■	■	■	Periplasm
MCA0875	Serine protease, subtilase family	■	■		■	■	■	■	■	■	■	Extracellular
MCA1082	Putative uncharacterized protein		■									Periplasm
MCA1510	Fimbrial protein	■			■	■	■	■	■	■		Extracellular
MCA1677	Glutamine synthetase				■	■		■	■	■	■	Cytoplasm
MCA1704	60 kDa chaperonin 2			■	■	■		■	■	■	■	Cytoplasm
MCA1796/ MCA2853	Methane monooxygenase	■			■						■	IM
MCA2555	Putative uncharacterized protein				■							Periplasm
MCA2589	Surface-associated protein	■			■			■		■	■	Extracellular
MCA2974	PKD domain-containing protein										■	Extracellular
MCA0087	hypothetical protein MCA0087										■	Extracellular

[a] A large number of cytoplasmicproteins were found in supernatant during lagphase Only proteins identified in the lag phase that were also found in the earlier stages of growth and the two proteins predicted to be extracellular are included in the table. For a complete list see CD provided with the thesis.

150 proteins were identified in the growth medium during lag face, a majority of which were predicted to be cytoplasmic proteins. Only two proteins were predicted to be extracellular among these. Of 14 proteins identified during logarithmic growth, four were predicted to be cytoplasmic. One protein was predicted to be located in the inner membrane and five are predicted to be periplasmic. Three proteins were predicted to be extracellular and one to be lipid-anchored to the OM. LC-MS analysis of proteins harvested from the growth medium confirmed the presence of a type IV protein with homology to PilE and PilA of other Gram-negative bacteria. This protein was secreted by *M. capsulatus* (Bath) after 3, 6 and 9 hours of growth. Another protein (MCA0087), identified by LC-MS in growth medium from a 24 hour culture and annotated `hypothetical protein´, was found to contain a *Neisseria* PilC beta-propeller domain. A BLAST protein search showed this protein to hold significant similarity to type IV fimbrial biogenesis protein PilY1 across a range of Proteobacteria. The other predicted extracellular proteins identified were a peptidase with a peptidase S8 family domain, and the surface associated protein MopE.

3.4 Proteins with possible roles in adhesion, immune modulation or invasion.

Extracellular and cell surface proteins are ideally positioned to participate in microbe-host interactions, and most virulence factors like adhesins, invasins and proteases are found on the bacterial surface or are secreted into the immediate environment (Finlay and Falkow 1997). All proteins predicted to be extracellular, integral to or anchored to the OM were searched for domains or annotations indicating roles in adhesion, invasion or possible immune regulatory effects.

All predicted or experimentally confirmed surface associated and extracellular proteins were therefore examined for adhesion-related domains, or homologous proteins involved in adhesion. Two pili complexes and? Other potential adhesins were found in the *Methylococcus capsulatus* (Bath) proteome.

3.4.1 A chaperone-usher type pili complex found in the *M capsulatus* (Bath) proteome

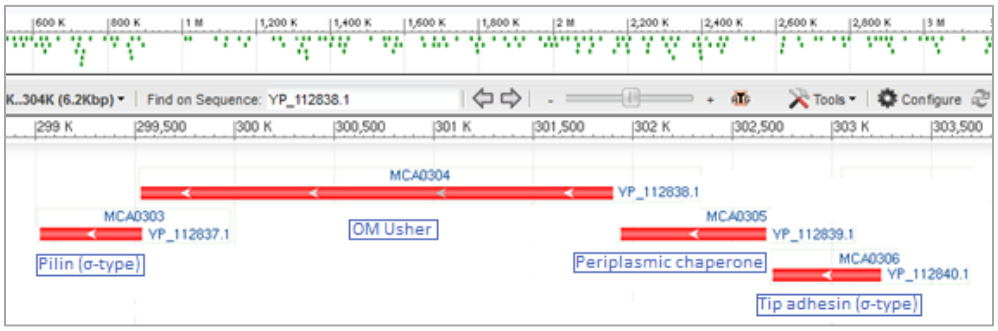
Among the predicted OM BB proteins was found a protein annotated `fimbrial biogenesis protein` (MCA0304). It contained usher domains (PF00577) associated with the secretion of

chaperone-usher (CU) type pili and a C-terminal PapC domain (PF13953) and showed significant homology to PapC porin proteins. Another two proteins MCA0303 and MCA0306 showing significant homology to CU type pili tip adhesins were found among the predicted extracellular proteins.

CU pili require a periplasmic chaperone for assembly and a search for this was conducted. As pili genes are often arranged in clusters the *M. capsulatus* (Bath) genome was scanned for other pili-related genes encoded in the vicinity of MCA0304. A gene encoding a chaperone was found between the predicted usher and one of the predicted tip-adhesins (Figure 3.9). A Pfam search confirmed that the protein contained a Gram-negative pili-assembly chaperone, N-terminal domain. The *in silico* prediction confirmed MCA0304 to be located in the periplasm. The four CU pili proteins are listed in Table 3.6.

Table 3.6. Four chaperone-usher pili proteins with predicted sub cellular location.

Locus	Protein name	Putative function	Informative homologes	Predicted SCL
MCA0303	Lipoprotein	Pilin (σ-type)	Sigma-fimbriae pilin	Extracellular
MCA0304	Fimbrial biogenesis protein	OM usher	OM usher protein	OM
MCA0305	Chaperone protein pmfD	Pili assembly chaperone	Chaperone protein	Periplasmic
MCA0306	Spore coat protein	Tip adhesin (σ-type)	Sigma-fimbriae tip adhesin	Extracellular



Figur 3.9. Four assumed chaperone-usher pili genes in the context of the *M. capsulatus* (Bath)genomic sequence. Putative function is indicated below each locus

3.4.2 A type IV pili complex identified in the *M. capsulatus* (Bath) proteome

Six proteins involved in type IV pili were found among the integral OM proteins, lipid-anchored and extracellular proteins. As assembly of a functional pili complex depends on a number of other proteins, a search for additional type IV pili-related proteins were conducted. In all 17 type IV pili- or pili assembly proteins, and all key components described by Craig and Li (2008) were found in the *M. capsulatus* (Bath) proteome (Table 3.7).

Table 3.7 Pili proteins found in *M. capsulatus* (Bath) and the closest homolog in *N. gonohorrea*

Locus	Protein name	Closest homolog i *N. gonorrhoeae*[a]	Function[b]
MCA2096	type IV pilus assembly protein PilB	pilF	Assembly ATPase
MCA2095	type IV pilin biogenesis protein PilC	PilG	Inner membrane protein
MCA2094	leader peptidase PilD	PilD	Prepilin peptidase
MCA0086	type 4 fimbrial biogenesis protein PilE	T4P modification protein PilV	
MCA1510	fimbrial protein	PilE	Pilin subunit
MCA0325	type 4 fimbrial biogenesis protein PilM	PilM	
MCA0326	type 4 fimbrial biogenesis protein PilN	PilN	
MCA0327	type 4 fimbrial biogenesis protein PilO	hypothetical protein NGO0096	
MCA0328	type 4 fimbrial biogenesis protein PilP	PilP	
MCA0329	fimbrial assembly protein PilQ	PilQ	Secretin
MCA1537	twitching motility protein PilT	PilT	Retraction ATPase
MCA1538	twitching motility protein PilU	Twitching motility protein	
MCA0090	type IV fimbrial biogenesis protein PilV	Put. T4P assembly protein PilV	
MCA0089	pilin-like protein	predicted protein	
MCA0088	hypothetical protein MCA0088	OpaD protein	
MCA0087	hypothetical protein MCA0087	pilC1/PilC2	Tip adhesin
MCA0091	hypothetical protein MCA0091	FimT	

[a]Closest homolog in *Neisseria gonorrhoeae* was determined by a protein Blast against the *Neisseria gonorrhoeae* genome (NCCP11945). [b] Function was assigned based on the function of the closest homolog in *N. gonorrhoeae* according to Craig and Li, 2008.

3.4.3 Potential non-pili adhesins in the *M. capsulatus* (Bath) proteome

MCA2227 was found among the predicted extracellular proteins and was identified as a filamentous hemagglutinin by homology. This protein is known to be exported via a type Vb transport system, and requires a transporter protein for secretion. Type Vb transporters frequently contain a POTRA domain and are predominantly of β-structure. A POTRA domain-containing protein (MCA2226) was found among the predicted β-barrels. The gene for this protein was found immediately next to FHA in the genome.

Another protein found among the predicted extracellular proteins, may have adhesin function: MCA2328 was homologous to proteins annotated as outer membrane adhesin-like proteins and contained a domain of unknown function (DUF) that is described as a parallel beta-helix repeated region positioned between successive cadherin domains. It was unclear however if the cadherin domain was part of the DUF and therefore present in *M. capsulatus* (Bath). Two

proteins (MCA1230 and MCA3103) containing OmpA domains, were found among the predicted BB, and a third one (MCA1819) among predicted lipoproteins. The OmpA domains covered only the C-terminal end in all three proteins. In MCA3103 a BB domain was found N-terminal to the OmpA domain. Manual screening for common integrin-binding sites in the amino acid sequence of these proteins revealed an arginine-glycine-aspartic acid (RGD) motif in MCA3103.

3.4.4 Domains with relevance to invasion, homeostasis or with immune effect

M. capsulatus (Bath) not only interacts with DC *in vitro*, but has been demonstrated to have anti-inflammatory effects in *Salmo salar* with soya-induced enteritis and mice with DSS-induced colitis. In order to explain mechanisms behind the observed effects, a search for proteins involved in host interactions or with possible immune modulatory effects was conducted. Relevant literature and the virulence factor database (VFDB) was used as a starting point for finding microbal proteins involved in host interactions and the *M. capsulatus* (Bath) proteome was searched for protein homologs or domains identifying such proteins. As *M. capsulatus* (Bath) having a prophylactic effect are lysed before being pelleted and fed to the mice, proteins responsible for the beneficial effects are not necessarily extracellular or surface-associated. Intracellular proteins were therefore also examined for any immune-modulatory effects.

Pfam showed MCA3012, a predicted cytoplasmatic protein, to contain a SEFIR domain. These are domains found in IL-17 receptors and their adaptor proteins (Wu, Gong et al. 2012)

MCA2795 phenyl pyruvate tautomerase protein showed homology to macrophage migration inhibitory factor (MIF), an immune-regulatory cytokine (Calandra and Roger 2003). This protein also contained PF01187, a MIF domain.

Three proteins, MCA1967, MCA0172 and MCA0363 showed homology to mammalian cell entry proteins (Mce). This is a family of proteins related to the mammalian cell entry proteins from *M. tubercolosis* shown to be necessary for colonization of and survival in macrophages. All three proteins also contained one or more Mce-related pfam domains.

MCA1752 contained four domains associated with alpha-2-macroglobulin. Prokaryote α2-macroglobulins have been suggested to be colonization factors by (Budd, Blandin et al. 2004). MCA0874, a hypothetical protein was found to contain 8 copies of Sel1, a tetratricopeptide repeat (TPR) motif associated with genes important in infection establishment, host cell trafficking and macrophage entry in *Listeria pneumophila* (Newton et al. 2007). Another protein (MCA0405) contained three copies of the same domain.

The *Pseudomonas aeruginosa* protein MucD is a serine protease described to be necessary for full virulence in *Arabidopsis thaliana*, *C. elegans*, wax moths and mice (Yorgey, Rahme et al. 2001). A protein annotated MucD was found in the *M. capsulatus* (Bath) proteome.

Two proteins, MCA0458 and MCA2498, were found to contain cysteine-rich repeat domains. These repeats are also found in fibroblast growth factor binding protein and E-selectin ligand. 6 proteins contained the von Willebrand factor type A domain. This domain is found in a number of human proteins, among these integrins and proteins involved in adhesion, signaling, and immune defenses. Functions in prokaryotes are unclear.

3.5 Comparison of adhesion proteins in two strains of *Methylococcus capsulatus*

Co-incubation of two different strains of *M. capsulatus*, Bath and Texas, with $CD14^+$ monocyte-derived DCs has in previous experiment shown a different potential of these two strains to adhere to such cells (Kleiveland et al., manuscript in preparation). While *M. capsulatus* (Bath) demonstrates massive binding to DCs, the Texas strain binds more sporadically. A pairwise global alignment of the amino acid sequence of all identified type IV pili proteins was done to examine if differences in amino acid sequences of pili proteins in the Bath and Texas strains could explain their different binding potential. Alignments showed that the majority of pili-related proteins in Texas and Bath share identical or highly similar amino acid sequences (Table 3.8) with conserved or semi conserved amino acid substitutions.

Table 3.8 Parwise alignement of type IV pili proteins in Texas and Bath strains of *M. capsulatus*

Locus	*M. capsulatus* (Bath) annotation	*M. capsulatus Texas* closest homolog[a]	Identity[b]	Similarity[b]
MCA2096	type IV pilus assembly protein PilB	T4P assembly, ATPase PilB	100	100
MCA2095	type IV pilin biogenesis protein PilC	T4P assembly protein PilC	100	100
MCA2094	leader peptidase PilD	Leader peptidase	99.3	99.3
MCA0086	type 4 fimbrial biogenesis protein PilE	T4P biogenesis protein PilE	90.6	91.2
MCA1510	fimbrial protein	T4P biogenesis protein PilE	99,3	100
MCA0325	type 4 fimbrial biogenesis protein PilM	T4P biogenesis protein PilM	100	100
MCA0326	type 4 fimbrial biogenesis protein PilN	T4P biogenesis protein PilN	100	100
MCA0327	type 4 fimbrial biogenesis protein PilO	T4P biogenesis protein PilO	99.5	99.5
MCA0328	type 4 fimbrial biogenesis protein PilP	T4P biogenesis protein PilP	100	100
MCA0329	fimbrial assembly protein PilQ	T4P biogenesis protein PilQ	99.4	99.4
MCA1537	twitching motility protein PilT	twitching motility protein PilT	100	100
MCA1538	twitching motility protein PilU	Twitching motility protein PilT	98.9	99.5
MCA0090	type IV fimbrial biogenesis protein PilV	T4P biogenesis protein PilV	99	99
MCA0089	pilin-like protein	T4P fimbrial biogenesis protein PilW	99.4	99.7
MCA0088	hypothetical protein MCA0088	T4P fimbrial biogenesis protein PilX	99.5	99.5
MCA0087	hypothetical protein MCA0087	T4P biogenesis protein PilY1	99.6	99.7
MCA0091	hypothetical protein MCA0091	T4P biogenesis protein FimT	98.9	98.9

[a] Blastx of nucleotide sequence from type IV pilus related genes in the Texas strain was used to identify protein homologues in the *Methylococcus capsulatus* (Bath) genome. [b] Scores for parvise alignment conducted by EMBOSS Needle (www.ebi.ac.uk/Tools/services/web_emboss_needle)

A gap was found in the alignment between a Bath (MCA0086) and its Texas homolog (Figure 3.10). However, the gap was found N-terminal to a GFXXXE motif recognized and processed by the PilD prepilin peptidase. It is therefore not expected to be part of the mature protein.

Figure 3.10. Pairwise alignment of MCA0086 and the closest Texas homolog showing an N-terminal gap in the alignment. Box indicates the Type IV pilin N-terminal methylation motif, GFxxxE recognized and processed by the prepilin peptidase.

3.5.1 Putative tip-adhesins of Texas and Bath compared

The putative T4P tip-adhesin of the M. capsulatus Texas and Bath strain was manually screened for possible integrin binding sites. Two such motifs were found: A leucine-aspartic acid-valine (LDV) motif starting from position 225 and a leucine-aspartic acid-threonine (LDT) motif starting in position 960. The pairwise sequence alignment between the tip-adhesins showed five aa substitutions between the two strains. None of the substitutions was found in any of the identified putative integrin binding sites. 3HX6_A, a crystal structure that could serve as a template for the C-terminal region of the tip-adhesin was found in the protein data bank through a blast search. This structure was from a Pseudomonas aeruginosa PilY1 and was used to make a model of the C-terminal part of the proteins. Only one of the aa substitutions and one of the integrin binding sites were found in the modeled region, both on the surface of the protein (Figure 3.11).

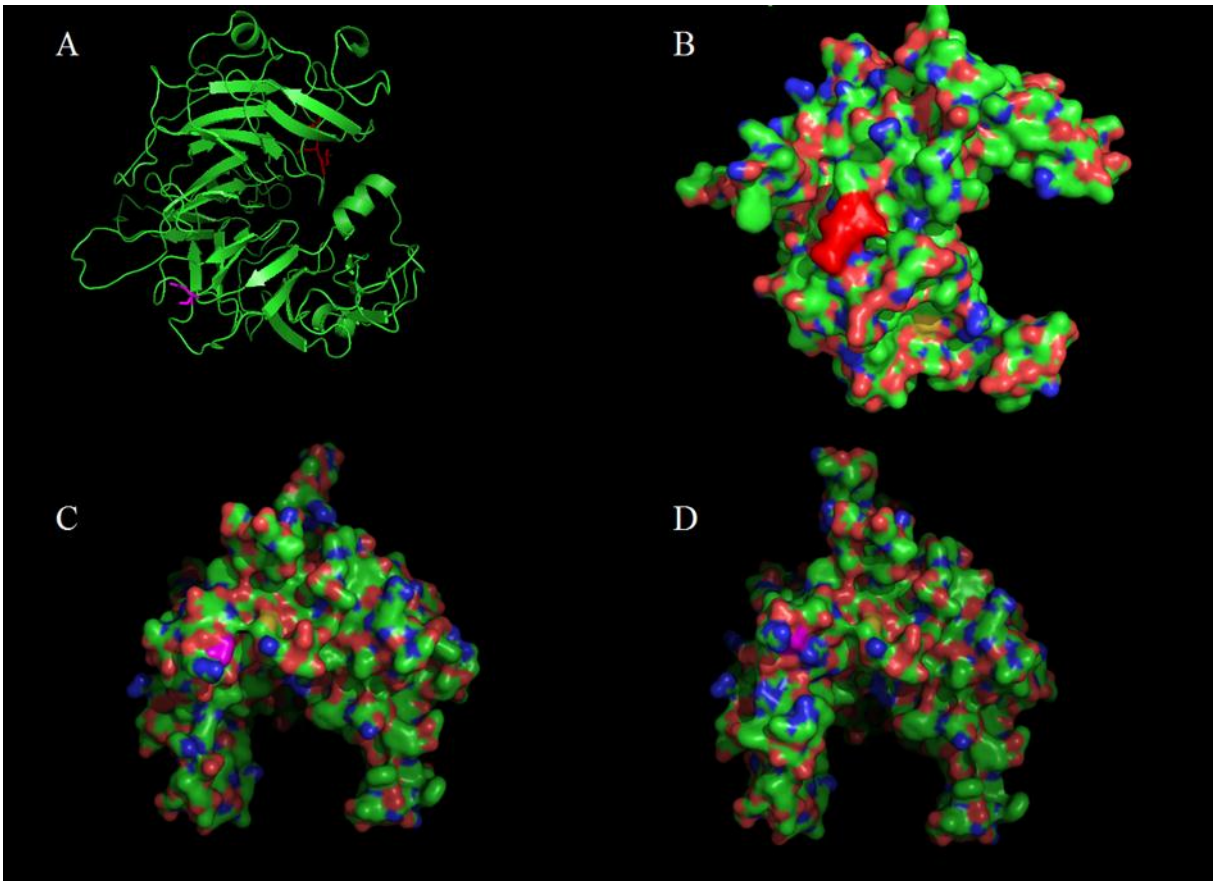


Figure 3.11 The tip-adhesin of *M. capsulatus* modeled after a C-terminal crystal structure of *Pseudomonas aeruginosa* PilY1 tip-adhesin: A-C) *M. capsulatus* (Bath) MCA0087. In A is shown the position of the putative integrin-binding site (red) and the position of the N→D substitution (magenta). B) Shows that a putative integrin-binding site found on the surface of the protein. C and D) Tip-adhesin from *M.capsulatus* Bath and Texas respectively, showing the residue that differs between the two proteins in magenta.

4 Discussion

4.1 *In silico* prediction of subcellular locations of *M. capsulatus* (Bath) proteins

A bacterial meal of *M. capsulatus* (Bath) has been shown to have anti-inflammatory effects in *Salmo salar* with soya-induced enteritis and mice with DSS-induced colitis (Kleiveland et al. 2012b; Romarheim et al. 2011), but the mechanisms behind the observed effects have not yet been identified. Studies of *M. capsulatus* (Bath) and human monocyte-derived DCs *in vitro* have shown that the bacterium adheres to and activates DC (Kleiveland et al., manuscript in preparation). By identifying microbial components involved in the interaction between the microbe and vertebrate cells one might be able to unravel some of the mechanisms involved in the abrogation of inflammatory conditions. In this study both bioinformatics and experimental approaches have been used to identify surface components involved in adhesion to DCs and to search for *M. capsulatus* (Bath) proteins with potential immunomodulatory effects.

To narrow the search for adhesins and determine if adhesion to DCs was mediated by a protein or a carbohydrate component, proteins were removed from bacterial surfaces by trypsin shaving, while carbohydrates were removed with a mixture of deglycosylation enzymes. SEM images of bacteria and DC blind-scored by two independent viewers revealed that trypsin-shaved bacteria demonstrated reduced ability to adhere to DC indicating that one or more proteins are involved in adherence of *M. capsulatus* (Bath) to the DCs. Deglycosylation appeared to increase bacterial adherence. With increasing amounts of deglycosylation enzyme mixture a stronger tendency of adherence was observed. Two important points should be noted: Before treatment the optical density of control and different samples were checked, but for methodological reasons, no quantification was done after treatment, and the amount of viable bacteria added to each well of DC might therefore have differed between samples. A second point of uncertainty relates to the fact that no visualization or quantification of amount of carbohydrates removed from the surface was performed, so that the actual effect of enzymatic treatment on surface carbohydrates has not been determined. The experiments with deglycosylation of surface proteins should therefore be considered as pilot studies. The theoretical implications of these results are interesting, though: If deglycosylation increases the adhesion capacity of *M. capsulatus* (Bath), it

suggests that the *Methylococcus* adhesin may be glycosylated and that removal of carbohydrate moieties uncovers interacting sites.

Genome-wide in silico subcellular location prediction for all M. capsulatus (Bath) proteins

Having established that proteins are involved in interactions between *M. capsulatus* (Bath) and human DCs it was necessary to define the surfactome of this prokaryote since adhesins should be present on the outer face of the cell. Cell envelope proteins in bacteria are typically difficult to characterize due to their low abundance, poor solubility, and problems with the isolation of pure surface fractions (Solis & Cordwell 2011). An *in silico* approach was therefore chosen.

Early in the process it was important to establish by subcellular location prediction what types of translocation systems, secretion systems and protein sorting complexes are used by *M. capsulatus*. This information can be used to look for domains or motifs typical of proteins being secreted through the various translocation and secretion pathways. Components of the Sec and Tat translocation systems; type I secretion system I; type II secretion system; lipoprotein transport system and BB insertion system was identified. A type V secretion system could not be identified by KEGG pathway or CMR genome properties search as the OM channel of this transport system is highly variable. However, a transporter of a type Vb secretion systems could be identified by searching for POTRA domains and the secreted protein by looking for genes encoded in its vicinity with unusually long signal peptides, both typical features of this type of transport system (Leo et al. 2012)

The importance of choosing an appropriate strategy for the location prediction became apparent at an early stage of the *in silico* analysis. Several subcellular prediction programs enabling batch searches are available online, each program constructed for a particular use and each with a set of limitations that must be taken into account when interpreting results. Programs constructed for the same purpose frequently gave different results, and programs predicting different properties gave contrasting results. For example Sec, Tat and lipoprotein signal peptides have many features in common, and signal peptide prediction programs vary in their ability to distinguish between the classes. Distinguishing between lipoproteins and sec transported non-lipoproteins is important to predict the fate of proteins. In this dataset one or both SignalP algorithms were found to predict SPase I-cleaved signal peptides for 55 of the 63 proteins predicted by LipoP to contain lipoprotein signal peptides. However LipoP, in contrast

to SignalP, is constructed to separate between SPaseI- and SPase II-cleaved signal peptides (Juncker et al. 2003) and was indeed shown to be a better discriminator as only five of these proteins were found to be Sec translocated.

Another common problem is the ability to discriminate between transmembrane helices and N-terminal signal peptides as the recognition of both types of peptides are usually based on searching for hydrophobic regions. Transmembrane helices typically have longer hydrophobic stretches of amino acids and do not have cleavage sites, but this pattern is not sufficient to distinguish between the two types of sequences. Therefore conventional transmembrane topology and signal peptide predictors, such as TMHMM and SignalP frequently give overlapping predictions (Kall et al. 2007). The newest version of SignalP, SignalP4.0 was designed to discriminate between signal peptides and transmembrane regions, and this program was initially used in the present study. However, SignalP4.0 was found to miss many proteins that were expected to be secreted. A second approach was attempted, using the older version, SignalP3.0, that has higher cleavage site sensitivity, in combination with Phobius, a hidden Markov model that combines transmembrane topology and signal peptide predictions (Kall et al. 2007) (See methods for details). This approach resulted in recognition of 147 proteins later predicted to be extracellular or OM BB that were not recognized by SignalP4.0.

Finding a good strategy when combining different prediction programs is challenging, and choices made should have a basis in the biology of the organism studied and in the assets and limits of the programs used. Romine (2011) has developed a strategy for genome-wide subcellular location prediction in Gram-negative bacteria, and a similar approach was chosen here. At each step of the process automated prediction was supported by manual evaluation of additional information such as informative domains, genome context, annotation information, homology and relevant literature (see method for details). Suggestion and tips by Emanuelsson et al. (2007) regarding use of different prediction programs proved helpful during analysis.

The chosen strategy and methodology resulted in prediction of subcellular location for all 2956 *M. capsulatus* (Bath) proteins, 129 of which were assigned a surface associated or extracellular location. Complementing automated predictions with manual curation at all steps of analysis was time consuming, but should result in more accurate results. However, relying on external data also has some pitfalls. As the criterion for predicting OM/extracellular location of proteins containing sec or tat peptides was either: 1) OM domain or extracellular domains 2) OM or extracellular homologs or 3) annotation data supporting OM or

extracellular location, some OM/extracellular proteins lacking such domains may erroneously have been predicted to be periplasmic. This is especially problematic when it comes to hypothetical proteins with no close homologs, as the function of secreted "unique" and potentially interesting *M. capsulatus* (Bath) proteins may escape further attention.

4.2 Identification of proteins released to the growth medium during different stages of growth

As mentioned above, the possibility exists that some extracellular proteins are not predicted to be secreted due to lack of conserved domains or homologs of known location. Also, *in silico* analysis depends on sequence information and may miss biological information not directly encoded in the protein sequence. In order to compare predicted extracellular proteins with proteins actually secreted, proteins were isolated from the growth medium of *M. capsulatus* (Bath) during early, mid and late exponential growth and stationary phase and identified by LC-MS analysis. 150 proteins, most of which were predicted to be intracellular, were found in the growth medium of a stationary phase culture. Only four predicted extracellular proteins were identified. (All 150 proteins are presented on CD accompanying the thesis. Se results for the four predicted extracellular proteins). Other extracellular proteins may have been present but masked by abundant intracellular proteins. During stationary phase the number of dying and newly dividing bacteria is in balance. The large number of intracellular proteins is expected to be a result of bacterial lysis. During exponential growth only 14 proteins were identified: Four of these were predicted by *in silico* analysis to be cytoplasmic, one protein was predicted to be located in the inner membrane, and five were predicted to be periplasmic. Three proteins were predicted to be extracellular and one to be lipid-anchored to the OM. Two of the predicted cytoplasmic proteins, MCA0707 and MCA1704 showed strong homology to heat shock protein 60 (Hsp60) of *Legionella pneumophila*. Hsp60 is a member of the GroEL family. These proteins are generally considered to be cytoplasmic and do not contain signal peptides or any other motif indicating that they are secreted. Several reports have challenged this view and immunolocalization techniques has demonstrated hsp60 of *L. pneumophila* to be surface- located (Garduno et al. 1998). These two proteins may therefor provide examples of extracellular proteins that cannot be predicted by *in silico* methods.

4.3 Identification of *M. capsulatus* proteins with potential roles in microbe-host interactions

Virulence factors are microbial molecules that enhance a microbe's potential to colonize, invade and survive within a host, and particular emphasis was placed on searching for proteins involved in such processes. Among proteins predicted to be, or found experimentally to be extracellular or associated with the OM, several proteins potentially involved in interaction with vertebrate host cells were found. Proteins from the growth medium of *M. capsulatus* (Bath) were identified as adhesins and potential virulence factors expressed under culturing conditions, and a number of additional virulence-related proteins was found to be encoded by the *M. capsulatus* genome by *in silico* analysis.

4.3.1 Identification of potential adhesins mediating adherence to DC

SEM images of *M. capsulatus* (Bath) co-cultivated with human $CD14^+$ monocyte-derived DCs for 3 h shows strong adhesion to the DCs. Previous studies have shown that *M. capsulatus* (Bath) do not adhere to $CD14^+$ monocytes, showing that the *M. capsulatus*-"receptor" is not expressed at this level of cell differentiation. Examination of predicted surface-associated and extracellular proteins revealed several proteins/protein complexes with similarities to well recognized adhesins in other bacterial species. Components of a type IV pili complex; a CU pili complex; afimbrial adhesin and proteins with possible adhesin function were identified among predicted β-barrels, lipoproteins and extracellular proteins.

Chaperone-usher pili

A chaperone, an usher and two proteins with homology to fimbria pilin and fimbria tip-adhesins were found among proteins predicted to be surface-associated and extracellular. Chaperone-usher pili proteins are found in a number of prokaryote species and are important virulence factors in many pathogenic strains. The binding partner of CU pili varies within and between species and depends on the type of pili.

The *M. capsulatus* (Bath) chaperone was placed within the σ-clade according to a classification scheme developed by Nuccio and Baumler (2007). This clade is widely distributed phylogenetically, but poorly characterized, and groups the *M. capsulatus* (Bath) CU pili with Orf1 of *Pseudomonas aeruginosa* and CsuD of *Vibrio parahaemolyticus*, two other gammaproteobacteria. Unfortunately, little is known about the morphology and function

of the surface structures of σ- fimbria, and the ligand/receptor for these pathogens is not known. Therefore, the role of the *M. capsulatus* (Bath) chaperone-usher pili in host interactions remains unknown.

Type IV Pili

All neccessary components to assemble a type IV pili (T4P) complex was found in the *M. capsulatus* (Bath) genome, and the structural subunit and tip-adhesin of the complex was identified in the growth medium of the bacterium. Type IV pili are critical virulence factors in pathogenic *Neisseria sp*., pathogenic strains of *Escherichia coli* and in *Pseudomonas aeruginosa.* Pilins *found in Methylococcus capsulatus*(Bath) is of the Iva type and therefor closer related to the pilins of *Neisseria sp.* and *P. aeruginosa (Craig & Li 2008)*. Possible roles of T4P in *Methylococcus*-host interactions and structural/ functional details are discussed later.

Filamentous hemagglutinin

A protein predicted to be extracellular and annotated as hemagglutinin-like protein was found to be homologous to filamentous hemagglutinin (FHA). FHA is secreted via a type Vb secretion system and requires an additional transporter protein to be translocated across the outer membrane. Type Vb transporters usually contain a POTRA domain, are able to form β-helical structures and are encoded close to the passenger protein in the genome (Leo et al. 2012). The transporter of *M. capsulatus* (Bath) filamentous hemagglutinin was identified among the predicted β-barrel proteins by domain and genomic context analysis, showing that *M. capsulatus* (Bath) encodes a filamentous hemagglutinin and is able to express it on the surface.

FHA is regarded as a protein with wide binding ability, and it expresses at least three different attachment activities in *Bordetella pertussis*. FHA is required for optimal adherence to ciliated respiratory cells in culture, non-ciliated epithelium like cells and macrophages (Locht et al. 1993). While both simple and complex carbohydrates inhibit adhesion to ciliated respiratory cells, sulphated saccharides (but not their unsulphated counterparts) blocked the binding to WiDr cells, heparin the binding to HeLa and integrin- specific antibodies the binding to macrophages. These data suggest that several different binding sites may be present on FHA and that different host receptors/ligands may be utilized at different stages of colonization. Noticeably, extracellular mycobacterial FHA can interact directly with dextran sulphate (Menozzi et al. 1991), emphasizing the importance of confirming the immunomodulatory

effects of *M. capsulatus* (Bath) on colitis in murine models where colitis is not induced by DSS.

Whether the *M. capsulatus* FHA is involved in adhesion to DC or if it is involved in adhesion to other cell types or substances is not clear, however an Arg-Gly-Asp (RGD) motif that is known to bind to integrins (Locht et al. 1993) was found in this protein, allowing interaction with cells expressing integrins.

4.3.2 Differences in adherence by Texas and Bath strains

Previous *in vitro* experiments have demonstrated that two strains of *M. capsulatus*, namely Bath and Texas, have different ability to adhere to DCs. Recently the genomic sequence of *M. capsulatus* (Texas) was published (Kleiveland et al. 2012a) allowing a comparison of amino acid sequences of proteins from the two strains.

A pairwise sequence alignment of T4P pilins and tip-adhesins from the Texas and Bath strains was performed. An 11 residue gap was found in the Bath sequence in an alignment between pilE (MCA0086) of the Bath strain and its Texas homolog. However, this gap was found in an N-terminal region assumed to be part of a signal peptide. The gapped extended N-terminal sequence found in the Texas strain is therefore not expected to affect the folding of the protein, but may affect the translocation of the protein and therefor the ability to quickly assemble a pili structure. Some proofs that the length of a signal peptide affect the rate of translocation exists: Autotransporter proteins often exhibit unusually long signal peptides. In an experiment the extended signal peptide was changed for a shorter “normal” one. Results showed that proteins were still translocated, but piltended to accumulate and misfold in the periplasm. The authors hypothesized that extended signal peptides slow down the Sec-dependent translocation by tethering the translocated protein to the IM allowing time for proper folding to take place (Szabady, Peterson et al. 2005). Transferred to the situation in Texas/Bath, a longer signal peptide in Texas may slow down translocation and reduce the ability to efficiently assemble a pilus upon encountering a DC. However, at this point the biological implication of an increased signal peptide length and amino acid substitutions found in the Texas pilin remain speculative.

In other species T4P is shown to exhibit PilC phase variation and PilE antigenic (sequence) variation contributing to altered binding properties (Hill & Davies 2009) Such differences will not be detected by *in silico* methods used in this study.

A pairwise alignment of the tip-adhesin of the T4P of the two strains showed three

semiconserved and two conserved substitutions in this protein. The tip-adhesin was then modeled to see if the substitutions found between the two strains had any effect on folding of the protein, or were exposed on the surface of the protein and could therefore affect binding. Because of the size of the tip-adhesin, the whole protein could not be modeled in a single process. The C-terminal part of the *M. capsulatus* Bath and Texas tip-adhesins was modeled after a template from *Pseudomonas aeruginosa* PilY1C-terminal domain. This was not an ideal template as the template only covers a short region of PilY1. The sequence alignment between the template and protein obtained low scores, but no closer related template exists. Only one substitution was found between the Bath and Texas tip-adhesin within the modeled part of the protein, a substitution from N → D (asparagine to aspartic acid). This substitution was found on the surface of the protein, but is considered to be a conserved substitution, and the biological relevance is not clear.

The binding site expected to be involved in adherence to integrins of host cells in *Pseudomonas aerugionosa* has been identified to be a RGD-motif (Johnson et al. 2011). Interestingly, the *M. capsulatus* (Bath and Texas) tip-adhesin does not contain the RGD motif, but two other motifs associated with integrin-binding is found in both the Bath and Texas tip adhesins: An LDV-motif associated with adhesion to α4β1 and α4β7 integrin is found N-terminally, while the motif leucine-aspartic acid-threonine (LDT) reported to be associated with adhesion to α4β7 integrin (Viney, Jones et al. 1996) is found C-terminally. The N → D substitution between Bath and Texas was not found within any of the putative integrin-binding sites and the contribution of this substitution to the different binding properties of *M. capsulatus* Texas and Bath could not be determined. Modeling of the C-terminal part of the tip-adhesin showed that both the putative integrin-binding site and the residue differing between the Texas and Bath strain is found on the surface of the protein.

The N-terminal part of pilY contains four substitutions, all within the first 113 residues. In order to determine if the N-terminal substitutions found in the T4P tip-adhesin of *M. capsulatus* affect structure and adhesive properties, the 3D-structure of the protein should be examined. Unfortunately, no suitable template could be found for the N-terminal region of the protein. Integrins are heterodimeric cell surface receptors that confer binding to other cells. The binding specificity is determined by the specific subunits that associate to form the receptor. $\alpha_4\beta_7$ integrin preferentially binds mucosal addressin cell adhesion molecule-1 (MAdCAM-1)and is expected to promote homing of $CD4^+$ T-cells to intestinal sites especially Peyer´s patches and mesenteric lymph nodes. Expression of α4β7 integrin is enhanced upon T-cell activation (DeNucci et al. 2010). Interaction of *M. capsulatus* (Bath) may be imagined

to affect T-cell effector function either through initiating signaling in the cells as a result of binding or through blocking the integrin, preventing homing.

Internalization of M. capsulatus (Bath) *in DC and host cell cytoskeleton rearrangements*

Many bacterial species are capable of entering, surviving and even multiplying in host cells. Entry into the cell may happen by phagocytosis or through bacteria-induced mechanisms. Both leukocytes and epithelial cells may be used as protective niches. Previous adhesion Leukocyte and endothelial cell adhesion molecules as targets for therapeutic interventions in inflammatory disease experiments of *M. capsulatus* (Bath) co-incubated with DCs have shown that single bacteria are internalized after adhering to the cells. Co-incubation followed by visualization by transmission electron microscopy demonstrates that bacteria are found in membrane-enclosed vesicles within the DC after internalization. Whether bacteria are internalized by phagocytosis or more active invasive strategies, and if the bacteria remain intact or are digested upon internalization, have not yet been determined.

After 3 hours of co-incubation, the adherence appears to be localized. Bacteria bind to specific regions of the DCs. Interestingly SEM images of earlier trials, with bacteria co-incubated with DC for only two hours do not show the same pattern, the adhering bacteria are found more dispersed on the surface of the DC. Although it should be noted that SEM images from two and three hours co-incubation are derived from two independent experiments, and not performed on cells from the same donor, the results point to the possibility that *M. capsulatus* (Bath) either moves on the surface of DC to regions rich in a particular ligand/receptor, or that adherence triggers a polarization of the DC.

Neisseria gonorrhea has been shown to specifically adhere to the uropods of polymorphonuclear granulocytes (PMN) in a T4P dependent manner (Soderholm, Vielfort et al. 2011). Bacteria adhered directly to the uropod, or could bind to pseudopods and then be transported to the uropod in the plasma membrane. The authors suggested uropod adhesion as a strategy for avoiding phagocytosis by the PMN, as phagocytosis occurs on the pseudopod pole of these cells. Single *Neisseria* are internalized at the uropod, but internalization is expected to be bacteria-initiated, and bacteria taken up via the uropod may avoid the degradation that follows phagocytosis. It has been proposed that *Neisseria* are able to "hitch-hike" on or within the neutrophil. Hitch-hiking may aid the bacterium in disseminating in the host and may even enable it to migrate on the neutrophil and cross the epithelial barrier. (Soderholm et al. 2011) Given that *M. capsulatus* (Bath) harbors a pilus complex

homologous to the *Neisseria* T4P complex, a similar strategy of avoiding phagocytosis and using leukocytes as vessels for dissemination is possible.

Some images of *M. capsulatus* (Bath) adherence to DCs after 2 hours co-incubation show protrusions from the DC membrane beneath the adhering bacteria. In order to elucidate the mechanisms of bacterial uptake, a search for proteins with possible involvement in cell invasion was conducted. Adhesion is typically the first step in invasion followed by host-microbe signaling events and changes in host membrane and actin cytoskeleton. Several pathogens trigger host cytoskeleton rearrangements as part of pathogenesis: *Neisseria sp.* type IV pili-mediated adhesion has been shown to lead to formation of cortical plaques, localized changes in the host membrane, enriched in components of the cortical cytoskeleton (F-actin and ezrin) and a subset of integral membrane proteins (EGFR, CD44 and ICAM-1). The T4P of *Methylococcus capsulatus* (Bath) may trigger similar changes in host-cell membranes and cytoskeleton allowing further adhesion interactions and signaling events.

E. coli (EPEC) adheres via T4P to the surface of intestinal epithelial cells. Similar to what is observed for *M. capsulatus* (Bath) adhering to DCs and *Neisseria gonorrhea* to PMNs, EPEC bind in a pattern described as localized adherence, forming large micro-colonies rather than adhering dispersed on the entire cell surface. Localized adherence is followed by injection of effectors by a TIIISS, rearrangement of host actin cytoskeleton and formation of large actin-rich pedestals (Donnenberg 2000). EPEC resting on the pedestal is then able to "surf" on the surface of the epithelial cell, taking advantage of the host actin pathways. As no evidence of a type III secretion system was found in *M. capsulatus* (Bath), the formation of pedestal-like structures in DC upon binding must be initiated by a different set of mechanisms than those used by EPEC.

Legionella pneumophila is capable of entering phagocytes through an unusual mechanism called coiling phagocytosis. A phagocyte pseudopod coils around the bacterium before internalizing it. Internalization is assumed to be mediated by a surface component since high titers of *L. pneumophila* antibodies abolishes coiling phagocytosis and leaves the bacteria to be engulfed by conventional phagocytosis (Horwitz 1984). A type IV pili system, a major outer membrane protein and a heat shock protein (Hsp60) are all surface-related proteins/protein complexes expected to be involved in adhesion and invasion in *L. pneumophila*. Hsp60 shows strong homology to the possibly extracellular MCA0707 and MCA1704 proteins found in *M. capsulatus* (Bath). Another *M. capsulatus* (Bath) protein, the hypothetical protein MCA0405, contained the RGD integrin-binding motif and showed significant homology to Enhanced entry protein C (EnhC) of *L. pneumophila*. EnhC has been

found to affect entry into host cells and has been suggested to be involved in a novel uptake mechanism such as coiling phagocytosis in this species (Cirillo, Lum et al. 2000). The finding of several homologs in *M. capsulatus* (Bath) of *Legionella* proteins involved in invasion suggests that *M. capsulatus* (Bath) may utilize similar mechanisms as *Legionella* in invasion of human cells.

4.3.3 Immune modulatory effects of *M. capsulatus* (Bath)

Kleiveland et al. (2012b) reported that mice with DSS-induced colitis that were fed a bacterial meal containing *M. capsulatus* (Bath) showed markedly reduced signs of intestinal inflammation as indicated by reduced cytokine mRNA expression and neutrophil infiltration; increased expression of the mucin 2 gene (*Muc2*), increased epithelial cell proliferation and colon crypt depths and decreased clinical symptoms (weight loss, diarrhea, blood in feces). In order to elucidate potential mechanisms involved, the *M. capsulatus* (Bath) proteome was examined for proteins with potential effect on host immune regulation or barrier function.

A MIF homolog identified in M. capsulatus (Bath)

The mammalian macrophage migration inhibitory factor (MIF) is a component of the host antimicrobial alarm system, and an integral component of host inflammatory responses (For a review see Calandra and Roger (2003)). A *M. capsulatus* (Bath) protein (MCA2795) annotated phenylpyruvate tautomerase were shown to be homologous to MIF. This protein is secreted by non-classical secretion mechanisms and was predicted to be extracellular in *M. capsulatus* (Bath). In mammals this cytokine is expressed by a range of leukocytes, epithelial cells and several tissues of the endocrine system, and is involved in both innate and adaptive immune responses. MIF has several points of action in the innate immune response: It activates extracellular signal regulated kinase (ERK1/ERK2) signaling and upregulates TLR4, promoting the detection of LPS of Gram-negative bacteria and enabling macrophages to act efficiently upon invasion. MIF suppresses apoptosis and growth arrest mediated by p53, a negative regulator of cell proliferation and it inhibits 'c-Jun activation domain-binding protein-1' JAB1, a co-activator of a transcription factor implicated in cell growth, and cell death (Calandra and Roger 2003). Human MIF homologues are found in helminths and protozoan parasitic species infecting mammalian hosts. In *Plasmodium berghei* MIF is expressed during parasite development in humans and is secreted from infected red blood cells. Gene deletion experiments show that MIF is not required for completion of the

parasite's life cycle, and it was suggested that the parasites express MIF to modulate the host immune response (Augustijn, Kleemann et al. 2007). The idea that a pro-inflammatory cytokine may be an advantage to a pathogen may sound far-fetched, but MIF's effects may turn out to be dose-dependent. Alternatively, the bacterial protein may compete with the vertebrate MIF for binding partners, thus having a desensitizing effect.

A SEFIR domain containing protein identified in M. capsulatus

The IL17-IL17R (IL-17 receptor) 'connection to IKK and SAPK/JNK' (CISK) signaling pathway play important roles in host pathogen defense. The IL-17 receptor and its associated adaptor protein CISK contain a domain, SEF/IL-17 receptor (SEFIR) SEF/IL-17 receptor (SEFIR), that mediates protein-protein interactions between components in this pathway (Wu, Gong et al. 2012). The SEFIR domain has been shown to occur in a limited number of prokaryotes, and interaction partners of the domain appear not to be found in these species, suggesting that SEFIR proteins interact with proteins from other organisms. Wu, Gong et al. (2012) performed a structural analysis of prokaryote SEFIR domains and concluded that they shared structural similarities with human CIKS and proposed that bacterial SEFIR domains may sabotage host pathways, hijacking IL-17R signaling by mimicry. MCA3012 a hypothetical protein predicted to be located in the cytoplasm contained the SEFIR domain

The IL-17 family of cytokines is central in host defense responses and inflammation. The production of IL-17 has been ascribed in particular to Th17 cells although several immune cells including macrophages, DCs, natural killer cells (NKs) and NKT cells also produce this cytokine. IL-17 family cytokines exert their effects through interaction with IL-17 receptors (IL17Rs). Stimulation of these receptors triggers signaling events resulting in production of cytokines (TGF-β, IL-6, IL-1β) chemokines (CXCL1, CXCL8, and CXCL10), granulocyte-macrophage and granulocyte colony-stimulating factor, anti-microbial peptides and acute-phase responses (Pappu, Ramirez-Carrozzi et al. 2011). The role of IL-17 in inflammatory disorders of the intestines are still being unraveled, but increased elevated levels of both IL-17A and its receptor is associated with IBD and the role of inflammatory cytokines induced by IL-17 is well established in Crohn´s disease. Also, neutrophile infiltration, signaling through MAPK kinases and activation of the transcription factor NF-κB is typical for inflammatory conditions, and IL-17 has been demonstrated to affect all these components of the immune system (Siakavellas & Bamias 2012). The fact that *M.capsulatus* (Bath) contains a domain with potential to interact with such a central pro-inflammatory pathway, is clearly

interesting in the light of the effects this non-commensal has been demonstrated to assert on soya induced enteritis in *Salmo salar* and DSS-induced colitis in mice.

A SIMPL domain containing protein identified in in M. capsulatus (Bath)

SIMPL (Signaling molecule that associates with mouse pelle-like kinase) is a coactivator in TNF-α mediated activation of nuclear factor-kB. Overexpression of SIMPL is found to lead to activation of NF-κB dependent promoters and transcription of genes involved in acute and chronic inflammatory response. A protein containing a SIMPL domain was found in *M.capsulatus*

4.3.4 Life within a host? Proteins conferring advantages in colonizing/invading bacteria

A number of proteins with a potential to increase the possibility for survival within a host was encountered during this study, but as no proof of *M. capsulatus* (Bath) residing within vertebrates has been found to this date, the discussion of such proteins will be limited to a brief comment.

Avoiding attack by the host complement system is a key to survival for many pathogens. OmpA, Neisseria T4P and filamentous hemagglutinin have all been described to bind C4bp, inactivating C4b and avoiding downstream activation of the complement cascade (Sukumaran et al. 2003). Surrounding itself with a capsule may be another mode of escaping eradication by the complement system. *M. capsulatus* (Bath) has been described to have a capsule. Ward et al. (2004) identified several possible pathways for synthesis of a capsule, and suggested colanic acid, alginate and O antigen as possible capsular materials. Capsules may help *M. capsulatus* (Bath) in avoiding phagocytosis or it may provide a protective envelope improving intracellular survival of the bacterium.

Invasive bacteria may enter phagocytic host cells by allowing itself to be phagocytosed or by bacteria-induced internalization events. Three *M. capsulatus* (Bath) proteins were found to contain mammalian cell entry (Mce) protein domains. In one of these proteins (MCA0363) three copies of the domain was found. The Mce protein family was first recognized as virulence factors when a DNA sequence from *Mycobacterium tuberculosis* was shown to enable recombinant *E.coli* to invade mammalian cells. *M. tuberculosis* has since been shown to contain four similar *mce* operons (1-4), each containing six *mce* genes (A-F) (Zhang and Xie 2011). Gene products of several of these genes have been shown to facilitate interaction

and internalization of bacteria and mammalian cells. Blast searches with the *M. capsulatus* (Bath) sequences against *M. tuberculosis* revealed MCA1967 and MCA0172 to be closest related to the gene product of *mce1B* while MCA0363 was closest related to the product of *mce3A.* All gene products of the *mce3* operon are putative invasion-like exported proteins suggested to be involved in entry and survival within macrophages, and Mce3A-coated latex beads have been demonstrated to be internalized by HeLa cells (El-Shazly, Ahmad et al. 2007). The function of Mce1B is predicted to be similar to the Mce1A protein (Zhang and Xie 2011). All three Mce domain-containing *M. capsulatus* (Bath) proteins therefore may potentially be involved in invasion and intracellular survival.

Ability to manipulate the host cytoskeleton is a common feature of invading pathogens. An actin-binding protein was found within *M. capsulatus* (Bath).

Pathogens that multiply within a host must have a way of escaping from the vacuole and the phagocyte. Lysins and phospholipases are commonly used for escape. *M. capsulatus* (Bath) expresses both an extracellular lysin and a putative phospholipase.

Some proteins or domains are typically associated with pathogenic species or strains, or are shown to increase virulence functions of pathogens, but does not have a defined function. One protein, MCA0553, was found to contain a F5/8 or discoidin domain (DS). This domain defines similar regions between the N-terminal region of discoidins, an adhesion protein from the slime mold *Dictyostelium discoideum* and C-terminal repeats from human blood coagulation factors V and VIII. The DS domain occurs in a number of eukaryotic proteins, most of which are involved in cell adhesion or developmental processes but has also been identified in prokaryotic species (Baumgartner, Hofmann et al. 1998). The domain binds a broad range of ligands, including phospholipids, carbohydrates and proteins.

Baumgartner, Hofmann et al. (1998) found homologues in bacterial sialidases and in the Mu-toxin of *Clostridium perfringens* and hypothesized that bacterial DS domains should be relevant from a clinical perspective, as they all occur in proteins expected to be virulence factors of human pathogens. The function of the DS-containing protein in *M. capsulatus* (Bath) remains unknown, but given its surface location and the fact that it contains a domain shared with coagulation-, adhesion- and virulence factors, the function of this protein may prove important in host interactions.

5 Concluding remarks

Microbes adapted to life within a host have developed a multitude of strategies and structures enabling them to interact with the host and shape their own environment. Typical of pathogenic microbes is that they attack a few key signaling- or cytoskeleton remodeling pathways at several different points. Although the pathways targeted are common for many different pathogens, the strategies and effectors utilized may differ greatly between species and in fact between strains of bacteria within the same species. Rather surprisingly, this work has identified a number of proteins in a non-pathogen, non-commensal bacterium with potential effects in a vertebrate host. Among the proteins described are adhesins with putative binding sites for vertebrate cell surface adhesion molecules, proteins possibly allowing invasion of host cells and proteins with potential immunomodulatory effects. In all, three proteins with potential regulatory effect on components of NF-κB activation pathways, a MIF protein, the SEFIR domain protein and a SIMPL domain protein, was found. *M. capsulatus* (Bath) may therefore share with several pathogens the ability to regulate NF-κB transcription activation by attacking at several different points. The SEFIR domain involved in IL-17 signaling potentially confer the bacterium regulatory control of another key component in inflammatory responses. These and other *Methylococcus* proteins should be examined further for structural properties and biological effects.

The discovery of immunomodulatory proteins in the *M. capsulatus* proteome may shed some light on the mechanism by which this bacterium affect inflammatory conditions in the intestines, but it also raises many questions: Why has a bacterium never described to be part of intestinal microflora come to harbor such a large number of proteins seemingly providing adaptions to life within a host? What is the advantage for the bacterium in having complex and energetically costly pili systems with adhesion sites for vertebrate cells? And how has *M. capsulatus* (Bath) gained homologs of vertebrate immunomodulatory proteins?

In conclusion, this study has by no means uncovered the full potential of the *M. capsulatus* (Bath) proteome. Non-commensal bacteria may hide many secrets for interacting with mammalian cells and the mammalian immune system just waiting to be discovered and exploited for different purposes.

6 References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25 (17): 3389-402.

Arrieta, M. C. & Finlay, B. B. (2012). The commensal microbiota drives immune homeostasis. *Front Immunol*, 3: 33.

Bendtsen, J. D., Kiemer, L., Fausboll, A. & Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiol*, 5: 58.

Berven, F. S., Flikka, K., Jensen, H. B. & Eidhammer, I. (2004). BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Res*, 32 (Web Server issue): W394-9.

Berven, F. S., Karlsen, O. A., Straume, A. H., Flikka, K., Murrell, J. C., Fjellbirkeland, A., Lillehaug, J. R., Eidhammer, I. & Jensen, H. B. (2006). Analysing the outer membrane subproteome of Methylococcus capsulatus (Bath) using proteomics and novel biocomputing tools. *Arch Microbiol*, 184 (6): 362-77.

Bhavsar, A. P., Guttman, J. A. & Finlay, B. B. (2007). Manipulation of host-cell pathways by bacterial pathogens. *Nature*, 449 (7164): 827-34.

Craig, L. & Li, J. (2008). Type IV pili: paradoxes in form and function. *Curr Opin Struct Biol*, 18 (2): 267-77.

Davidsen, T., Beck, E., Ganapathy, A., Montgomery, R., Zafar, N., Yang, Q., Madupu, R., Goetz, P., Galinsky, K., White, O., et al. (2010). The comprehensive microbial resource. *Nucleic Acids Res*, 38 (Database issue): D340-5.

DeNucci, C. C., Pagan, A. J., Mitchell, J. S. & Shimizu, Y. (2010). Control of alpha4beta7 integrin expression and CD4 T cell homing by the beta1 integrin subunit. *J Immunol*, 184 (5): 2458-67.

Donnenberg, M. S. (2000). Pathogenic strategies of enteric bacteria. *Nature*, 406 (6797): 768-74.

Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*, 2 (4): 953-71.

Finlay, B. B. & Falkow, S. (1997). Common themes in microbial pathogenicity revisited. *Microbiol Mol Biol Rev*, 61 (2): 136-69.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., et al. (2010). The Pfam protein families database. *Nucleic Acids Res*, 38 (Database issue): D211-22.

Fronzes, R., Remaut, H. & Waksman, G. (2008). Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria. *EMBO J*, 27 (17): 2271-80.

Garduno, R. A., Faulkner, G., Trevors, M. A., Vats, N. & Hoffman, P. S. (1998). Immunolocalization of Hsp60 in Legionella pneumophila. *J Bacteriol*, 180 (3): 505-13.

Hill, S.A. & Davies, J.K.(2009) Pilin gene variation in Neisseria gonorrhoeae:reassessing the old paradigms. *FEMS Microbiol Rev, 33:521-530*

Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C. & Ala'Aldeen, D. (2004). Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev*, 68 (4): 692-744.

Hybiske, K. & Stephens, R. S. (2008). Exit strategies of intracellular pathogens. *Nat Rev Microbiol*, 6 (2): 99-110.

Ishibashi, Y., Claus, S. & Relman, D. A. (1994). Bordetella pertussis filamentous hemagglutinin interacts with a leukocyte signal transduction complex and stimulates bacterial adherence to monocyte CR3 (CD11b/CD18). *J Exp Med*, 180 (4): 1225-33.

Ivanov, II, Frutos Rde, L., Manel, N., Yoshinaga, K., Rifkin, D. B., Sartor, R. B., Finlay, B. B. & Littman, D. R. (2008). Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe*, 4 (4): 337-49.

Jarchum, I. & Pamer, E. G. (2011). Regulation of innate and adaptive immunity by the commensal microbiota. *Curr Opin Immunol*, 23 (3): 353-60.

Johnson, M. D., Garrett, C. K., Bond, J. E., Coggan, K. A., Wolfgang, M. C. & Redinbo, M. R. (2011). Pseudomonas aeruginosa PilY1 binds integrin in an RGD- and calcium-dependent manner. *PLoS One*, 6 (12): e29629.

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. & Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci*, 12 (8): 1652-62.

Kall, L., Krogh, A. & Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res*, 35 (Web Server issue): W429-32.

Kleiveland, C. R., Hult, L. T., Kuczkowska, K., Jacobsen, M., Lea, T. & Pope, P. B. (2012a). Draft Genome Sequence of the Methane-Oxidizing Bacterium Methylococcus capsulatus (Texas). *J Bacteriol*, 194 (23): 6626.

Kleiveland, C. R., Hult, L. T., Spetalen, S., Kaldhusdal, M., Christofferesen, T. E., Bengtsson, O., Romarheim, O. H., Jacobsen, M. & Lea, T. (2012b). The non-commensal bacteria Methylococcus capsulatus (Bath) ameliorates dextran sulfate sodium (DSS)-induced ulcerative colitis by influencing mechanisms essential for maintenance of barrier function. *Appl Environ Microbiol*.

Kovacs-Simon, A., Titball, R. W. & Michell, S. L. (2011). Lipoproteins of bacterial pathogens. *Infect Immun*, 79 (2): 548-61.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*, 305 (3): 567-80.

Kunisawa, J. & Kiyono, H. (2012). Immune regulation and monitoring at the epithelial surface of the intestine. *Drug Discov Today*.

Leo, J. C., Grin, I. & Linke, D. (2012). Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos Trans R Soc Lond B Biol Sci*, 367 (1592): 1088-101.

Locht, C., Bertin, P., Menozzi, F. D. & Renauld, G. (1993). The filamentous haemagglutinin, a multifaceted adhesion produced by virulent Bordetella spp. *Mol Microbiol*, 9 (4): 653-60.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*, 39 (Database issue): D225-9.

McGuckin, M. A., Linden, S. K., Sutton, P. & Florin, T. H. (2011). Mucin dynamics and enteric pathogens. *Nat Rev Microbiol*, 9 (4): 265-78.

Menozzi, F. D., Gantiez, C. & Locht, C. (1991). Interaction of the Bordetella pertussis filamentous hemagglutinin with heparin. *FEMS Microbiol Lett*, 62 (1): 59-64.

Morand, P. C., Tattevin, P., Eugene, E., Beretti, J. L. & Nassif, X. (2001). The adhesive property of the type IV pilus-associated component PilC1 of pathogenic Neisseria is

supported by the conformational structure of the N-terminal part of the molecule. *Mol Microbiol*, 40 (4): 846-56.
Newton, H. J., Sansom, F. M., Dao, J., McAlister, A. D., Sloan, J., Cianciotto, N. P. & Hartland, E. L. (2007). Sel1 repeat protein LpnE is a Legionella pneumophila virulence determinant that influences vacuolar trafficking. *Infect Immun*, 75 (12): 5575-85.
Nuccio, S. P. & Baumler, A. J. (2007). Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev*, 71 (4): 551-75.
Prasadarao, N. V., Blom, A. M., Villoutreix, B. O. & Linsangan, L. C. (2002). A novel interaction of outer membrane protein A with C4b binding protein mediates serum resistance of Escherichia coli K1. *J Immunol*, 169 (11): 6352-60.
Romarheim, O. H., Overland, M., Mydland, L. T., Skrede, A. & Landsverk, T. (2011). Bacteria grown on natural gas prevent soybean meal-induced enteritis in Atlantic salmon. *J Nutr*, 141 (1): 124-30.
Romine, M. F. (2011). Genome-wide protein localization prediction strategies for gram negative bacteria. *BMC Genomics*, 12 Suppl 1: S1.
Rose, R. W., Bruser, T., Kissinger, J. C. & Pohlschroder, M. (2002). Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol Microbiol*, 45 (4): 943-50.
Schulz, G. E. (2002). The structure of bacterial outer membrane proteins. *Biochim Biophys Acta*, 1565 (2): 308-17.
Serino, L., Nesta, B., Leuzzi, R., Fontana, M. R., Monaci, E., Mocca, B. T., Cartocci, E., Masignani, V., Jerse, A. E., Rappuoli, R., et al. (2007). Identification of a new OmpA-like protein in Neisseria gonorrhoeae involved in the binding to human epithelial cells and in vivo colonization. *Mol Microbiol*, 64 (5): 1391-403.
Seydel, A., Gounon, P. & Pugsley, A. P. (1999). Testing the '+2 rule' for lipoprotein sorting in the Escherichia coli cell envelope with a new genetic selection. *Mol Microbiol*, 34 (4): 810-21.
Siakavellas, S. I. & Bamias, G. (2012). Role of the IL-23/IL-17 axis in Crohn's disease. *Discov Med*, 14 (77): 253-62.
Smith, S. G., Mahon, V., Lambert, M. A. & Fagan, R. P. (2007). A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiol Lett*, 273 (1): 1-11.
Soderholm, N., Vielfort, K., Hultenby, K. & Aro, H. (2011). Pathogenic Neisseria hitchhike on the uropod of human neutrophils. *PLoS One*, 6 (9): e24353.
Solis, N. & Cordwell, S. J. (2011). Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. *Proteomics*, 11 (15): 3169-89.
Sukumaran, S. K., Shimada, H. & Prasadarao, N. V. (2003). Entry and intracellular replication of Escherichia coli K1 in macrophages require expression of outer membrane protein A. *Infect Immun*, 71 (10): 5951-61.
Thanassi, D. G. & Hultgren, S. J. (2000). Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr Opin Cell Biol*, 12 (4): 420-30.
Tokuda, H. (2009). Biogenesis of outer membranes in Gram-negative bacteria. *Biosci Biotechnol Biochem*, 73 (3): 465-73.
Tommassen, J. (2010). Assembly of outer-membrane proteins in bacteria and mitochondria. *Microbiology*, 156 (Pt 9): 2587-96.
Tseng, T. T., Tyler, B. M. & Setubal, J. C. (2009). Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol*, 9 Suppl 1: S2.
van den Berg, B. M., Beekhuizen, H., Mooi, F. R. & van Furth, R. (1999). Role of antibodies against Bordetella pertussis virulence factors in adherence of Bordetella pertussis and

Bordetella parapertussis to human bronchial epithelial cells. *Infect Immun*, 67 (3): 1050-5.
Waksman, G. & Hultgren, S. J. (2009). Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat Rev Microbiol*, 7 (11): 765-74.
Walters, M. S. & Mobley, H. L. (2009). Identification of uropathogenic Escherichia coli surface proteins by shotgun proteomics. *J Microbiol Methods*, 78 (2): 131-5.
Ward, N., Larsen, O., Sakwa, J., Bruseth, L., Khouri, H., Durkin, A. S., Dimitrov, G., Jiang, L., Scanlan, D., Kang, K. H., et al. (2004). Genomic insights into methanotrophy: the complete genome sequence of Methylococcus capsulatus (Bath). *PLoS Biol*, 2 (10): e303.
Whittenbury, R., Phillips, K. C. & Wilkinson, J. F. (1970). Enrichment, isolation and some properties of methane-utilizing bacteria. *J Gen Microbiol*, 61 (2): 205-18.