Tail(or)-made Protein Secretion in Pathogenic Mycobacteria

Daleke, M.H.

2012

document version
Publisher's PDF, also known as Version of record

1

GENERAL INTRODUCTION

Tuberculosis

Tuberculosis (TB) is an airborne infectious disease that has plagued mankind for thousands of years. TB is caused by a group of closely related bacterial species known as the *Mycobacterium tuberculosis* complex (MTBC) (Fig. 1), of which *M. tuberculosis* is the primary causative agent in humans. The most common form of the disease is pulmonary TB, but virtually any part of the body can be affected (1). In a small percentage of infected individuals TB can cause characteristic changes to the skeleton, such as bone lesions or deformation of the spine, also known as Pott's disease (2). Such changes have been observed in skeletal remains from Neolithic Sweden (3200-2300 BC) (3), Pharaonic Egypt (3000-500 BC) (4) and Iron Age Britain (1200 BC) (5). Detection of ancient MTBC DNA in bone samples and preserved tissue from these sites confirms that TB was widespread and relatively frequent during antiquity (3-5). In agreement with this finding, the Greek physician Hippocrates (460-370 BC) described TB, or phthisis, as the most prevalent disease of his time (6). The earliest evidence of human TB is found in 9,000-year-old bone samples (7), but TB is thought to have been present in the human population much earlier than that. Based on genome comparisons and analyses of mutation rates the MTBC has been proposed to originate from a common ancestor present in East Africa 40,000-15,000 years ago, which coincides with the migration of modern humans out of Africa (8, 9). Consistent with this idea, MTBC DNA was detected in the bones of a 17,000 year old extinct bison in North America, indicating that the disease had spread to the Americas prior to the closing of the Bering Strait approximately 9,000 years ago (10).

In the 18th and 19th centuries, TB reached epidemic proportions in Europe and North America, claiming nearly 1,000 lives per 100,000 population annually. Fortunately, as living standards improved TB prevalence began to decline. The elucidation of the etiology of TB by the German physician Robert Koch in 1882 paved the way for the development of tools to detect and combat TB. In 1943 the Swedish-Danish physician Jörgen Lehmann discovered the first tuberculostatic, paraamino salisylic acid (PAS). This finding was shortly followed by the discoveries of more efficient antituberculosis drugs, such as streptomycin, isoniazid and rifampicin, which allowed treatment to cure TB (6). Nevertheless, in the 1980s the TB incidence began to increase rapidly again, and TB is now responsible for 1.4 million deaths worldwide annually. The reemergence of TB is strongly associated with the growing HIV pandemic; of the 8.8 million new cases of TB that were reported in 2010, 1.1 million were in people living with HIV (11). The TB prevalence is highest in the high HIV-burden regions Africa, South-East Asia and the Western Pacific, where TB is a major cause of death in people co-infected with HIV (1). Another key factor contributing to the resurgence of TB is the emergence of multidrug-resistant *M. tuberculosis* (MDR-TB), which is a particularly large problem in China, India and Russia. In *M. tuberculosis*, drug resistance arises as a result of spontaneous mutations and failure to follow treatment will allow drug-resistant subpopulations to become dominant within a patient. MDR-TB, *i.e.* strains that are resistant to at least the first-line drugs isoniazid and rifampicin, can be treated with second-line drugs. However, lack of compliance has led to the emergence of extensively drug-resistant TB (XDR-TB), and recently to strains that are resistant to all available antituberculosis drugs (TDR-TB) (12). The World Health Organization (WHO) has launched strategies

aimed at reducing TB prevalence and death rates and ultimately at eliminating TB as a public health problem (13). These efforts have contributed to a global decline in TB incidence rates since 2002; however, as a consequence of the world's rising population, the incidence rates only drop with 1.3% yearly (11).

The only available vaccine against TB is bacille Calmette-Guérin (BCG), a live attenuated strain of *Mycobacterium bovis* that became avirulent due to loss of genetic regions during serial passage *in vitro*. BCG has been used to protect infants and young children against severe forms of TB since 1921. However, BCG confers highly variable protection against pulmonary disease in adults. The underlying reasons for this are not clear, although previous exposure to environmental mycobacteria may limit the efficacy of BCG (14). In addition, BCG may have lost too many virulence factors to induce adequate protective immunity (15).

In 90% of cases, inhalation of *M. tuberculosis* does not result in active disease, but rather in an asymptomatic latent infection (14). Upon entry in the lung, *M. tuberculosis* is taken up by alveolar macrophages and delivered into phagosomes. Normally, bacteria-containing phagosomes become acidified and fuse with hydrolytic lysosomes, resulting in degradation of the bacteria. However, *M. tuberculosis* has evolved mechanisms to evade the various bactericidal mechanisms, and can survive and even replicate within host cells (16). The infected macrophages attract other immune cells, including macrophages, granulocytes and lymphocytes, leading to the formation of a tight and stratified cluster of cells known as the granuloma, which allows the host to contain and control the infection (17). However, it also protects the bacteria from eradication by the host immune system, and *M. tuberculosis* can survive in granulomas for years or even decades (18). It is estimated that about two billion people carry a latent TB infection (19). As 5-10% of persons with a latent infection will eventually develop active TB, which cannot be prevented by BCG, these individuals form a huge reservoir of potential disease and transmission (14). Efforts are being made to develop improved vaccines against TB, and currently several vaccine candidates are in pre-clinical and clinical trials. These include both novel priming vaccines to replace BCG, such as recombinant BCG and attenuated *M. tuberculosis* strains, and boosts for BCG, such as subunit vaccines and live vector vaccines expressing mycobacterial antigens (11). Future vaccine candidates will also include post-exposure vaccines to prevent reactivation of latent TB (20).

The mycobacteria

In addition to *M. tuberculosis*, more than 100 other mycobacterial species have been identified (23). These can be divided in two groups based on their growth rate: the rapid-growers, which produce visible colonies on solid medium within seven days, and the slow-growers, which typically require 10-28 days for visible growth (24). Analysis of 16S rRNA indicates that the slow-growers have evolved from the rapid-growers (25) (Fig. 1). This division also reflects the virulence of mycobacterial species, as most fast-growers are non-pathogenic, free-living saprophytes, whereas the slowly growing group contains a number of important human and animal pathogens (25). The latter include, apart from the species belonging to the *M. tuberculosis* complex, *Mycobacterium leprae*, the causative agent of leprosy; *Mycobacterium ulcerans*, the etiological agent of Buruli ulcer; *Mycobacterium*

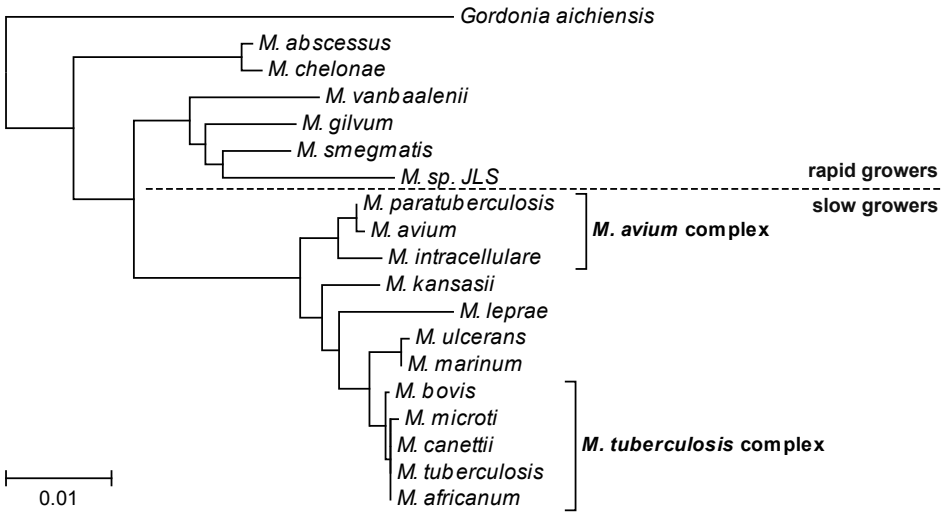

Fig. 1. Phylogenetic tree of selected members of the genus *Mycobacterium*. The evolutionary relationship was inferred from 1368 aligned nucleotides of 16S rRNA DNA sequences of selected mycobacterial species and the outgroup *Gordonia aichiensis*, using the Neighbor-Joining method (21). The subgroups of rapid- and slow-growing mycobacteria are separated by a dotted line, and the *M. tuberculosis* and the *M. avium* complex are indicated. The tree is drawn to scale, and the scale bar represents 0.01 changes per nucleotide. Phylogenetic analyses were conducted in MEGA4 (22).

avium, which causes bird tuberculosis and opportunistic infections in immunocompromised humans; and *Mycobacterium marinum*, an aquatic organism that causes chronic, granulomateous infections in ectotherms and skin infections in humans (26). Mycobacteria are rod-shaped, non-flagellated bacteria that belong to the high C+G Gram-positive bacteria. Although they are classified as Gram-positive they have a highly characteristic staining profile, known as acid fast or Ziehl Neelsen. This staining characteristic is due to the presence of a special and complex cell envelope structure characterized by complex, long-chain (C_{70}-C_{90}) fatty acids known as mycolic acids (further discussed below). Mycobacteria are phylogenetically most closely related to the genera *Corynebacterium*, *Nocardia* and *Rhodococcus*, which also produce mycolic acids, although of shorter length and lower complexity (24).

Models for *M. tuberculosis* pathogenesis

Successful development of novel drugs and vaccines against TB requires a better understanding of the pathogenesis of *M. tuberculosis*; however, research on this topic is limited by several factors. First, with its generation time of 22 hours, *M. tuberculosis* renders experiments time consuming (24). Second, *M. tuberculosis* is highly pathogenic and must be handled in biosafety level (BSL) 3 facilities (27). In addition, the most frequently used animal model, the mouse, fails to mimic structured granuloma formation and disease progression of human TB (28). Other

infection models that mimic human TB better, such as guinea pigs and macaques, are more costly and, especially the use of monkeys raises ethical concerns (28). Therefore, several laboratories study TB using mycobacterial model organisms, of which the rapid-growing saprophyte *Mycobacterium smegmatis*, *M. bovis* BCG and *M. marinum* are most widely used. *M. smegmatis* and *M. bovis* BCG are non-pathogenic to healthy humans and thus safe alternatives to *M. tuberculosis*. However, both species lack important virulence factors; *M. smegmatis* as a consequence of its distant relatedness to *M. tuberculosis* (Fig. 1), and *M. bovis* BCG due to genetic deletions (29), as will be discussed further below. An increasing number of laboratories use *M. marinum*, one of the closest relatives of *M. tuberculosis* outside the MTBC (Fig. 1) (30). Other advantages of working with this species include its fairly high growth rate (it has a doubling time of 4-6 h), its relatively high amenability to genetic manipulations, and its mild pathogenicity to humans (29). Human infections may result from contact of injured skin with infected fish or contaminated water, and are usually limited to the skin of extremities as a consequence of the low optimal growth temperature (30°C) of *M. marinum*, although disseminated disease has (rarely) been reported in immunocompromised patients (31). Another important advantage of *M. marinum* is the availability of natural animal models, most notably the genetically tractable zebrafish model (*Danio rerio*). Early stages of granuloma formation can be studied in real time in the transparent zebrafish embryos, which only have an innate immune system (32, 33). In addition, *M. marinum* infections of adult zebrafish, which have both innate and adaptive immunity, allow the study of caseating granulomas very similar to those observed in human TB (34). Genome comparisons show that many TB-related virulence factors are conserved in *M. marinum* (35), and infections of macrophages or zebrafish have shown that virulence determinants of *M. tuberculosis* can rescue the mutated version of *M. marinum* orthologous genes (36, 37).

The mycobacterial cell envelope

Mycobacteria are surrounded by a unique cell envelope, which is composed of a plasma membrane (PM), an unusually lipid-rich cell wall core and an outer capsule-like layer (Fig. 2). The cell wall core is distinctive for mycobacteria and their close relatives and consists of peptidoglycan (PG), arabinogalactan (AG) and mycolic acids. Mycobacterial PG forms a relatively thin layer outside the PM and is structurally very similar to that of many other bacteria, except that it contains N-glycolyl muramic acid instead of the more common N-acetylated form. What is even more unusual is that PG is covalently linked to another branched polysaccharide, *i.e.* arabinogalactan. The terminal arabinose residues of AG are in turn esterified with mycolic acids (38). Mycolic acids are 2-alkyl, 3-hydroxyl fatty acids, which contain a shorter, saturated α-branch and a long meromycolate chain. The latter can be modified at two positions with a variety of functional groups, giving rise to a large repertoire of different mycolic acids, some of which are species-specific (39). Mycolic acids are linked to AG to form a super-molecular complex of PG-AG-mycolic acid, but also exist in the cell wall as non-covalently attached glycolipids, usually in the form of trehalose dimycolate (TDM or cord factor), which associate with other free lipids and glycolipids. The composition of this mycolic acid-containing layer

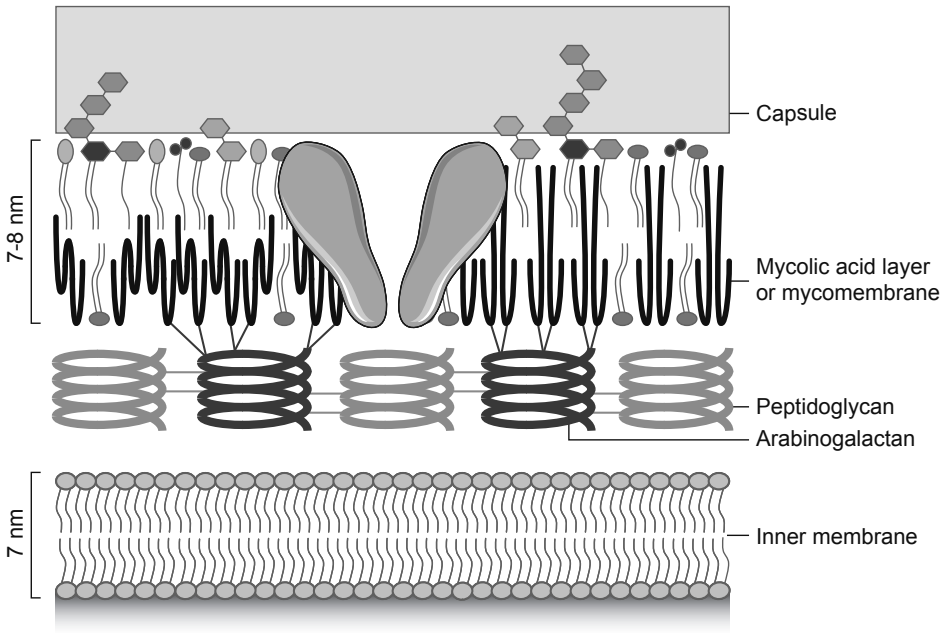

Fig. 2. Model of the mycobacterial cell envelope. On the outside of a typical bacterial inner membrane is the cell wall core, consisting of covalently linked peptidoglycan-arabinogalactan-mycolic acids. The mycolic acids form part of the inner leaflet of a second hydrophobic layer that is reminiscent of the outer membrane of Gram-negative bacteria. This layer also contains a wide array of free lipids, many of which are specific for mycobacteria. The precise arrangement of the mycolic acids is still a matter of debate, and the two current models are presented here: the mycolic acids may exist in an extended form, with the proximal end of the meromycolate chain extending into the outer leaflet (right), or they may adopt a compactly folded conformation (left). In analogy with the outer membrane of Gram-negative bacteria, proteins are found in the mycolic acid-containing layer, here depicted as a channel. The outermost layer of the mycobacterial cell envelope is the capsule, which consists of polysaccharides and proteins. This figure was adapted from (50) with permission from Nature Publishing Group, and made with the aid of Servier Medical Art.

varies between mycobacterial species, and frequently contains immunologically active lipids such as lipooligosaccharides (LOSs), lipomannan (LM), phenolic glycolipids, glycopeptidolipids and sulfolipids. The cell wall is also rich in lipid-anchored antigenic glycoconjugates, such as lipoarabinomannan (LAM) (38). The outermost layer of the mycobacterial cell envelope is formed by a loosely attached polysaccharide- and protein-rich capsule, which, in analogy to other pathogenic bacteria, may be involved in interaction with the host (40).

The structural organization of the mycobacterial cell wall lipids has long been a matter of speculation. Recently, cryo-electron microscopy (EM) analysis confirmed the presence of a lipid bilayer external to the PG in the cell envelopes of mycobacteria and *Corynebacterium glutamicum*, of which the integrity depended on the presence of mycolic acids (41, 42). Especially the latter species was helpful to elucidate this, because mycolic acids are, in contrast to in the mycobacteria, not essential in *C. glutamicum*.

This bilayer is analogous to the outer membrane (OM) of Gram-negative bacteria, albeit completely different in organization and lipid composition, and was hence named mycolic acid-containing outer membrane (MOM) or "mycomembrane". The AG-anchored mycolic acids are expected to form part of the inner leaflet; however, as the width of the MOM is too small to accommodate the long meromycolate chains in an extended conformation, it is not completely clear how they are arranged (41, 42). Different models have been proposed, in which the meromycolate chains are either linear and possibly extended into the outer leaflet of the MOM (41), or restricted to the inner leaflet by adopting a folded conformation (Fig. 2) (42, 43). Biophysical experiments using Langmuir monolayers supported the model of compactly folded mycolic acids forming the inner layer of the MOM (44, 45).

The complex organization and concomitant low permeability of the cell wall efficiently protect mycobacteria from noxious compounds and harsh conditions such as drought, osmotic shock and low pH (38, 46). However, this rigid and impermeable cell wall also prevents diffusion of nutrients. Therefore, mycobacteria are predicted to produce outer membrane channels, known as porins, for the uptake of hydrophilic compounds such as glucose, amino acids and metal ions (46, 47). However, thus far not many porins have been identified in mycobacteria (47). Certain hydrophobic compounds appear to be able to diffuse through the cell wall, whereas others, such as cholesterol, are taken up by dedicated transport systems (48). It is not known how fatty acids, the preferred carbon source of *M. tuberculosis in vivo* (49), are acquired, but they might be taken up via one of the transport systems homologous to the MCE4 cholesterol pathway (47). Importantly, the cell wall is not only an efficient permeability barrier for compounds from the environment, it is also hinders the secretion of proteins. Therefore, mycobacteria utilize specialized secretion systems to translocate proteins across their cell envelopes.

The type VII protein secretion pathway

The virulence of bacterial pathogens depends to a large extent on the ability to secrete virulence factors to the bacterial surface, external milieu or directly into host cells (51). Because the biological membranes in the bacterial cell wall hinder export of proteins, translocation across these barriers is mediated by dedicated protein secretion systems. Similar to other bacteria, mycobacteria secrete proteins across the inner membrane via the ubiquitous general secretory (Sec) pathway or twin-arginine translocation (Tat) system (52). Proteins designated for transport via these pathways have a specific N-terminal signal sequence that is usually removed upon translocation (53, 54). In addition to these housekeeping secretion pathways, mycobacteria also contain a copy of the accessory SecA2 system, which mediates secretion of a subset of proteins via a poorly understood mechanism (55). In Gram-positive bacteria, which have only one lipid bilayer, the Sec/Tat pathways are generally sufficient for protein export. Gram-negative bacteria, however, have evolved a number of specialized secretion systems for transport of proteins across the outer membrane. Six pathways, generally known as the type I to type VI secretion (T1S to T6S) systems, have been identified in Gram-negative bacteria. These pathways either secrete proteins that are delivered into the periplasm by the Sec or Tat systems (type II and V) in a two-step process, or via a signal peptide-independent one-step mechanism across the

entire cell envelope (type I, III, IV and VI) (56). Although mycobacteria also contain a diderm cell envelope, they lack the type I to type VI pathways, probably due to the convergent evolution of the diderm cell envelope. However, they have, perhaps unsurprisingly given the complexity of the mycobacterial cell wall, evolved a unique specialized secretion system. This system is known as the ESX or type VII secretion pathway, and is responsible for the secretion of a large number of proteins that lack N-terminal signal peptides (50).

The five ESX systems and their substrates

The discovery of the type VII pathway began with the characterization of the potent, low-molecular-weight T cell antigens ESAT-6 (early secreted antigenic target of 6 kDa) and CFP-10 (culture filtrate protein of 10 kDa) (57, 58). ESAT-6 and CFP-10, which are encoded by an operon, are both efficiently secreted in the culture medium of *M. tuberculosis*, despite the fact that they lack classical signal peptides (59, 60). Comparative genomics methods revealed that the *esat-6* and *cfp-10* genes are found in a region known as region of difference 1 (RD1), which is present in virulent strains of *M. tuberculosis* and *M. bovis*, but absent from the genomes of *M. bovis* BCG (61) and the avirulent vole bacillus *Mycobacterium microti* due to independent deletion events (Fig. 3) (62). Because this locus contains genes coding for several predicted integral membrane proteins and ATPases it was suggested already more than a decade ago that the genes surrounding *esat-6* and *cfp-10* encode a system responsible for secretion of the ESAT-6 and CFP-10 proteins (63-65). Later, this cluster, now named ESX-1, was indeed shown to encode a functional secretion system, as complementation of *M. bovis* BCG with the entire genomic locus from *M. tuberculosis* restored secretion of ESAT-6 (66), and inactivation of separate transcriptional units or single genes within this cluster in *M. tuberculosis* abolished secretion of ESAT-6 and/or CFP-10 (67-69). Furthermore, studies in *M. marinum* and *M. smegmatis* showed that ESX-1 genes are required for secretion of ESAT-6 and CFP-10 also in these species, indicating that the ESX-1 cluster encodes a functionally conserved secretion system in mycobacteria (36, 70). Recently, it has become clear that ESX-1 is responsible for secretion of multiple substrates. These include *e.g.* the ESX-1-encoded EspB (71, 72), EspE (73), EspF and EspK (74), all identified in studies of *M. marinum*, and *M. tuberculosis* EspA (75) and EspC (76), which are encoded elsewhere in the genome (Fig. 3) (Esp stands for ESX-1 secretion-associated proteins (77)).

ESAT-6 and CFP-10 belong to the WXG100 family of proteins, which are characterized by a helix-turn-helix structure, a conserved Trp-X-Gly motif (WXG) that is located in the turn between the helices, and an approximate length of 100 amino acids (78-80). The *esat-6* and *cfp-10* genes are typically organized in operons with a *cfp-10* homologue followed by an *esat-6* homologous gene (59, 63). In the *M. tuberculosis* genome, 23 *wxg100* genes are arranged in 11 such pairs, with one single *wxg100* gene located near one pair (64). In four cases these pairs are surrounded by gene clusters with homology to the ESX-1 locus (Fig. 3) (64, 65). Phylogenetic analyses and genomic comparisons suggest that these clusters, which are named ESX-2 to ESX-5, have arisen by duplication events. ESX-4, which also contains the smallest number of genes, is thought be the most ancestral cluster. This cluster appears to have been duplicated to give rise to ESX-1, -3, -2 and finally, ESX-5. Interestingly, whereas ESX-1 to ESX-4 are widely distributed in the

genomes of mycobacteria, the most recently evolved cluster, ESX-5, is restricted to the slow-growing species. Thus, while absent from the genome of fast-growing saprophyte *M. smegmatis*, ESX-5 is present in mycobacterial pathogens such as *M. tuberculosis, M. leprae* and *M. marinum* (64). Both ESX-3 and ESX-5 have been shown to encode functional secretion systems responsible for secretion of WXG100 proteins (81, 82). In addition, ESX-5 seems to be a major secretion pathway for the enigmatic PE and PPE proteins (81, 83). These proteins, which constitute two large virulence-associated mycobacterial protein families, will be further discussed below.

Structural components of the ESX machineries

Analysis of the five different ESX loci shows that, in addition to the *wxg100* genes, there are six other genes that are conserved in the ESX clusters, which are thought to encode structural components of the ESX secretion machineries (77). These include three predicted integral membrane proteins with no clear homology to known proteins, *i.e*. EccB, EccD and EccE (where Ecc stands for ESX-conserved component), a predicted serine protease of the subtilisin family, *i.e* the mycosin MycP, a soluble ATPase of the AAA+ family designated EccA and a putative membrane-anchored

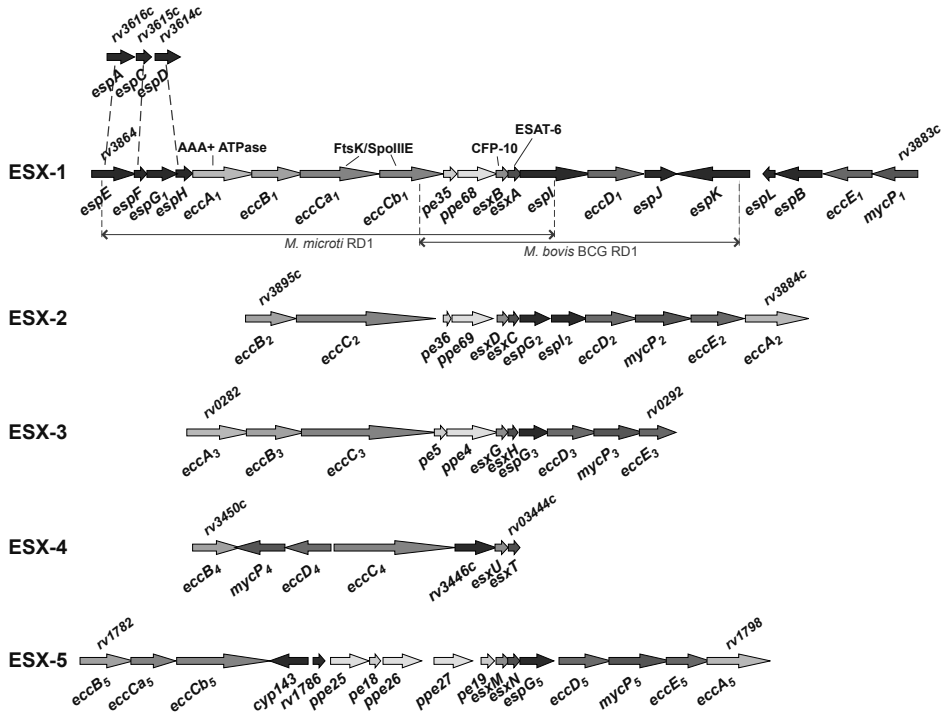

Fig. 3. Genetic organization of the five ESX clusters in *M. tuberculosis*. The RD1 deletions of *M. bovis* BCG and *M. microti* are marked in the ESX-1 cluster. The genes that encode the ESX-1 substrates EspA and EspC are encoded at a distinct genetic location and show homology to genes near the 5′ end of the ESX-1 locus (marked by dotted lines). *ecc* stands for ESX conserved component and *esp* for ESX-1 secretion-associated protein. Modified from (77).

protein of the FtsK/SpoIIIE family of ATPases, *i.e.* EccC (Fig. 3) (77). Notably, EccC belongs to the same family as the energy-providing coupling protein VirD4 of the type IV secretion system (84), and is therefore thought to play a critical role in ESX secretion. Most of the available data on the role of these components in secretion derive from studies on the ESX-1 cluster. Inactivation of *eccCb*$_1$, *eccD*$_1$ and *mycP*$_1$ in *M. tuberculosis* showed that the encoded components are indeed critical for ESX-1-mediated secretion (67-69, 85). Moreover, also *eccB*$_1$ and *eccE*$_1$ have been shown to be required for ESX-1 secretion in *M. smegmatis* (70). In *M. marinum*, transposon mutants with strongly reduced ESX-5 secretion have been obtained in genes coding for the cytosolic proteins EccA$_5$ (83) and EspG$_5$ (81) (the latter will be further discussed in chapter 5). Notably, despite the screening of large libraries of transposon mutants, these were the only two genes for which mutants with transposon insertions could be isolated in this species (81, 83). This is in contrast to *M. tuberculosis*, in which the crucial role for the EccC$_5$ and EccD$_5$ proteins in ESX-5 secretion was recently confirmed with targeted knockouts (86) and transposon mutants (Houben *et al.*, submitted), suggesting that ESX-5 is essential in *M. marinum*.

In a recent study, EccB$_5$, EccC$_5$, EccD$_5$, EccE$_5$ and MycP$_5$ were shown to localize in the cell envelope of *M. marinum* (Houben *et al.*, submitted). Furthermore, a 1500 kDa complex corresponding to the core ESX-5 secretion machinery was isolated from cell envelopes of *M. marinum* and *M. bovis* BCG, and was shown to be composed of EccB$_5$, EccC$_5$, EccD$_5$ and EccE$_5$, but notably not MycP$_5$ (Houben *et al.*, submitted). As all known structural components of the ESX systems are inner membrane-associated or cytosolic (Houben *et al.*, submitted), the mechanism of substrate transport across the mycomembrane remains an open question. Thus far, only a small number of outer membrane proteins have been identified in mycobacteria (47), and it is possible that a yet-to-be identified protein is responsible for translocation across the mycolic acid-containing layer (Fig. 4*A*). Such a component could be associated with each of the ESX systems, or be shared by all of them. Another possibility is that (part of) the large ESX core membrane complex extends across the periplasmic space and the mycomembrane, thereby forming one large channel across the entire cell envelope (Fig. 4*B*) (Houben *et al.*, submitted).

Targeting of substrates to ESX-1

As discussed above, the T7S substrates lack detectable secretion signals, and studies on ESX-1 indicate that substrate targeting is a complex process. The extreme C terminus of CFP-10, which forms a flexible tail (80), contains a seven amino acid signal that is recognized by the soluble part of the EccC$_1$ protein, *i.e.* EccCb$_1$ (87). This signal is required for secretion of CFP-10 and its partner ESAT-6, which lacks its own signal and thus piggybacks on CFP-10 for secretion (87). Recently, also another ESX-1 substrate, EspC, was shown to depend on its C terminus for secretion. Intriguingly, this C terminus was shown to interact with the cytosolic ATPase EccA$_1$, and to mediate secretion in a manner functionally distinct from the C-terminal signal of CFP-10 (76).

Yet another level of complexity is added to substrate targeting since several of the ESX-1 substrates are mutually dependent for secretion. For reasons that are not clear, secretion of ESAT-6, CFP-10, EspA or EspC is abolished in the absence of any of the other three substrates (75, 76, 88). It has been speculated that this codependency

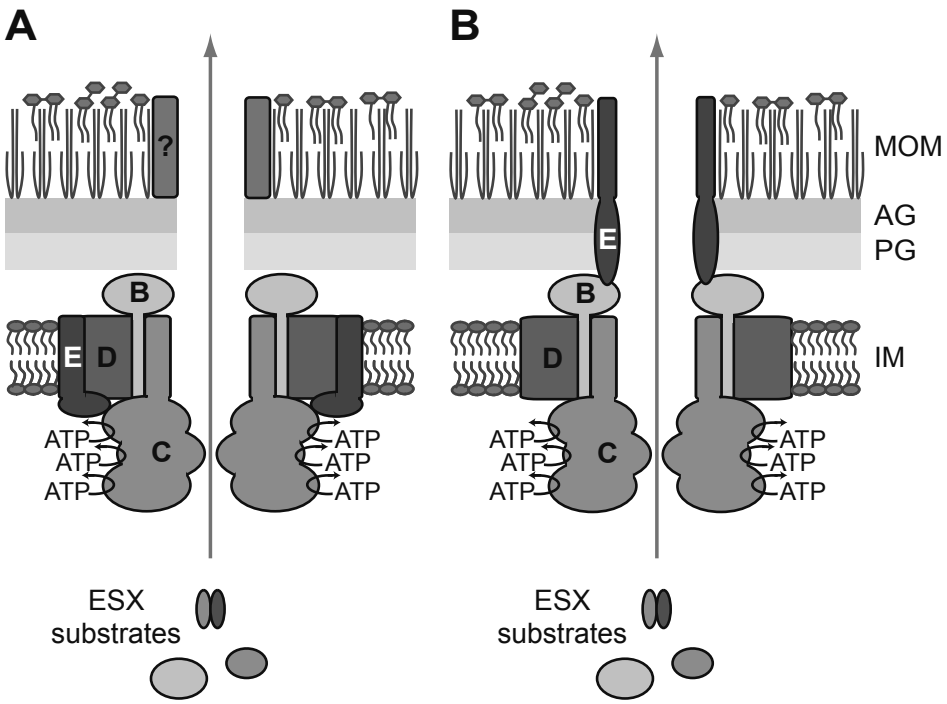

Fig. 4. Models of the architecture of the type VII secretion systems. Studies in *M. marinum* and *M. bovis* BCG recently identified an ESX-5 core membrane complex composed of the FtsK/SpoIIIE-like ATPase $EccC_5$ and the transmembrane proteins $EccB_5$, $EccD_5$ and $EccE_5$, and two models were proposed regarding the mechanism of ESX secretion (Houben *et al.*, submitted). (*A*) In the first model, substrates are translocated across the inner membrane (IM) by the EccBCDE complex and across the mycolic-acid containing outer membrane (MOM) by an unknown component, in a two-step procedure. (*B*) Alternatively, the large ESX core complex may form a channel that spans the entire cell envelope, thereby mediating secretion in a one-step procedure. In this model, $EccE_5$, the only component that is restricted to species with a mycolic acid-containing cell envelope, was proposed to extend into the MOM. AG, arabinogalactan; PG, peptidoglycan. By courtesy of E. Houben.

is due to the need for the substrates to interact in the cytosol prior to secretion, to bring together the structural components of the secretion system at the membrane and/or to subsequently be secreted as a multi-protein complex (75, 76). Alternatively, these proteins might be secreted components of the translocation machinery (75, 76), in similarity to the codependent T6S substrates Hcp and VgrG, which have been proposed to form the extracellular portion of the T6S translocation channel (89).

Function of the type VII secretion systems

Re-introduction of the RD1 region into *M. bovis* BCG and/or *M. microti* and the generation of genetic deletions in the RD1 region of *M. tuberculosis* and *M. bovis* have made it clear that ESX-1 plays a critical role in mycobacterial virulence (66, 68, 90-92). Detailed investigations have linked cytotoxicity, cytolysis, cell-to-cell spread, granuloma formation and *in vivo* growth and virulence to the presence of a

functional ESX-1 secretion system in *M. marinum* and *M. tuberculosis* (36, 37, 67-69, 93). The mechanism behind these effects is not clear, but it appears to involve the ability of ESX-1-secreted proteins to disrupt biological membranes (36, 68). Indeed, *M. tuberculosis, M. leprae* and *M. marinum* have all been shown to escape in an ESX-1-dependent fashion from the phagosome into the host cell cytosol during infection (94, 95). Due to its ability to lyse phospholipid membranes and host cells *in vitro*, these membranolytic properties have been attributed to ESAT-6 (68, 96, 97). Interestingly, an *M. tuberculosis* strain expressing a normally secreted version of ESAT-6 that lacked its last 12 amino acids remained in the phagosome of infected macrophages and was attenuated in a mouse model, indicating that the extreme C terminus of this protein is critical for its function (98, 99). It should, however, be pointed out that ESX-1 is widely distributed amongst non-pathogenic mycobacteria, and the presence of a functional ESX-1 system does not correlate with virulence (99). Moreover, in the avirulent *M. smegmatis*, ESX-1 is involved in conjugal DNA transfer (100, 101). Therefore, while it is possible that ESAT-6 has acquired different functions in pathogenic mycobacteria, ESX-1-dependent virulence might be related to other secreted effector proteins.

Of the five ESX loci, only ESX-3 is essential for *in vitro* growth of *M. tuberculosis* (102). This locus is under transcriptional control of the iron- and zinc-dependent regulators IdeR and Zur (103, 104), and it could be conditionally inactivated upon supplementation of the growth medium with iron and/or zinc (105). The encoded ESX-3 secretion system mediates secretion of at least one WXG100 protein, and the secretion levels have been found to be inversely correlated to the amounts of extracellular iron (82). Furthermore, work on the non-essential ESX-3 in *M. smegmatis* showed that the production and ability of the mycobacterial iron siderophores mycobactin and carboxymycobactin to bind iron are independent of ESX-3, but that the mycobacteria could not utilize the sequestered iron (82). The ESX-3 secretion system thus appears to play an essential role in mycobacterial iron uptake, possibly by acquiring iron from siderophores. Interestingly, the ESX-3 locus is intact in *M. leprae* (64), a bacterium that has undergone extensive genome reduction and is thought to maintain only those genes that are essential for survival, which further underlines the essential role of this locus in the mycobacterial life cycle.

Of the remaining three loci, virtually nothing is known about ESX-2 and ESX-4, whereas on the contrary, a number of studies have shed light on the function of the most recently evolved system, ESX-5 (86, 94, 106, 107). Host macrophages respond to mycobacterial infection with a number of different antimicrobial mechanisms, such as the production of reactive oxygen and nitrogen species, and by stimulation of a cellular immune response through the production of proinflammatory cytokines (108). However, pathogenic mycobacteria have developed ways to interfere with the various host defense mechanisms, including suppression of the signaling pathways required for secretion of proinflammatory cytokines (109). The ESX-5 system was shown to play a role in immunomodulation, as infection of macrophages with an isogenic *M. marinum* ESX-5 mutant strain led to increased production of proinflammatory cytokines such as TNF-α, IL-6 and IL-12 as compared to the wild-type strain (106). In addition, ESX-5 was shown to induce cell death, which might facilitate cell-to-cell spread (94). As ESX-1, but not ESX-5, is required for escape

from the phagosome, it was hypothesized that ESX-5 effectors will interact with and manipulate the host cell after ESX-1-mediated translocation into the cytosol (94). Recently, an *M. tuberculosis* ESX-5 mutant was shown to be attenuated in growth both during *in vitro* infection of macrophages and upon infection of severe combined immune-deficient (SCID) mice (86). A similar attenuation was observed during the early stages of infection of zebrafish embryos with an *M. marinum* ESX-5 mutant strain (107). Surprisingly, the same ESX-5 mutant strain exhibited a hypervirulent phenotype in adult zebrafish, with increased bacterial replication and rapid granuloma formation. This hypervirulence could not be explained by the presence of a fully developed adaptive immune system in the adult fish, and was attributed to an undefined *in vivo* growth advantage of the mutant (107).

Finally, it should be pointed out that the type VII secretion systems are not restricted to mycobacteria, and clusters with similarity to the ancestral ESX-4 locus are found in other mycolates, *i.e. Nocardia farcinica* and *Corynebacterium diphteriae* (64). Furthermore, WXG100 homologues and other ESX-like genes have also been identified in Actinobacteria (high G+C Gram-positive bacteria) that lack mycolic acids, such as *Streptomyces coelicolor* (64, 110), and even among the distantly related Firmicutes (low G+C Gram-positive bacteria), such as *Listeria monocytogenes*, *Staphylococcus aureus*, *Bacillus subtilis* and *Bacillus anthracis* (79). This suggests that the type VII pathway originally evolved as a Sec-independent system for protein export across the cytoplasmic membrane in Gram-positive bacteria (50), where it is involved in diverse functions as sporulation in *S. coelicolor* (110) and persistence in *S. aureus* infection (111). Subsequently, this pathway adapted to translocate proteins across the mycolic-acid containing layer, possibly by acquiring additional genes, and evolved further to perform critical roles in the pathogenesis of mycobacteria (50).

The PE and PPE protein families

One of the major surprises of the *M. tuberculosis* genome sequencing project was the discovery of the *pe* and *ppe* multigene families, which with 99 and 69 copies each were found to constitute nearly 10% of the coding capacity of the genome. The *pe/ppe* genes code for members of two large protein families, the PE and PPE proteins, which are named after the conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) motifs near their respective N-termini (63). These motifs are found within conserved, unrelated N-terminal domains of approximately 110 (PE) and 180 (PPE) amino acids (63). PE/PPE proteins can consist of only these so-called PE and PPE domains (Fig. 5*A*), or they may have extended C termini (Fig. 5*B*). In the latter case, the C-terminal domains are composed of unique sequences or of sequences with characteristic glycine-rich repeats, such as those encoded by the polymorphic CG-rich-repetitive sequence (PGRS) (112) subfamily of PE proteins and the major polymorphic tandem repeat (MPTR) (113) PPE subfamily (63, 114).

Inspection of the genomic sequences of various mycobacterial species and closely related bacteria has shown that the *pe/ppe* genes are largely unique to mycobacteria (114). Moreover, whereas fast-growing mycobacteria carry a small number of these genes, *pe/ppes*, especially of the *pe_pgrs* and *ppe_mptr*

subfamilies, are highly enriched in the genomes of slow-growing, pathogenic species (114). Very little is known about the function of these enigmatic proteins, and thus far LipY, a triacyl glycerol lipase, is the only PE protein for which a function has been characterized (115). Nevertheless, the accumulation of *pe/ppe* genes in the genomes of mycobacterial pathogens is intriguing and suggestive of a role of the PE/PPE proteins in mycobacterial pathogenesis. In agreement with this, *in vivo* analysis of the proteome of *M. tuberculosis* during infection of guinea pigs revealed that PE and PPE proteins were amongst the ten most abundant proteins during both the early and the chronic stages of infection (116). Numerous PE and PPE proteins localize to the mycobacterial cell envelope and are at least partly exposed on the bacterial surface (74, 117-123), and are thus ideally placed to interact with

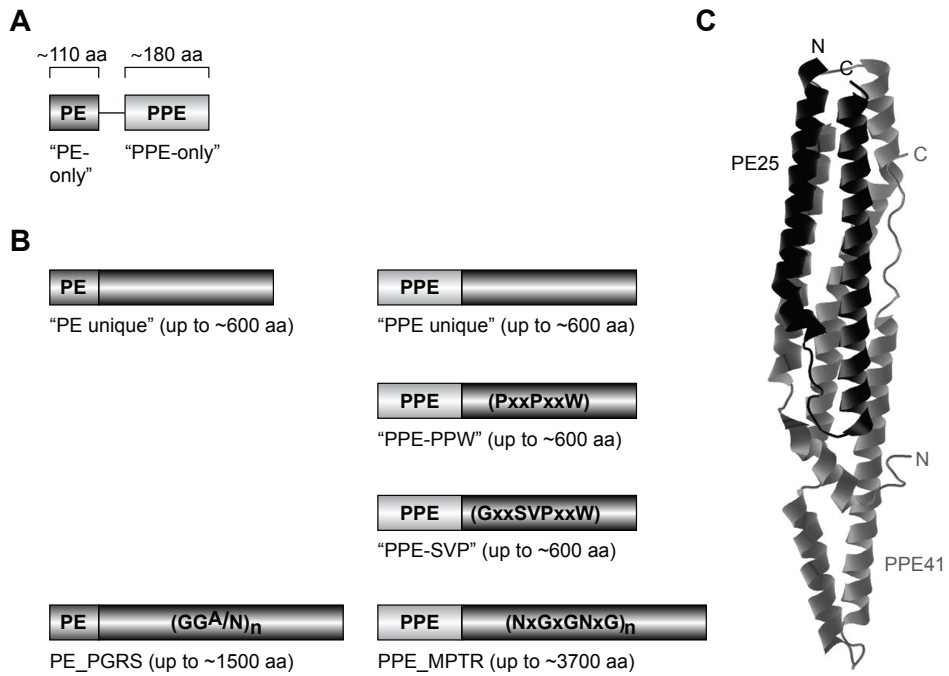

Fig. 5. PE/PPE domain organization, subfamilies and structure. Members of the PE/PPE subfamilies share conserved N-terminal domains of 110 and 180 amino acids (aa), which contain the characteristic Pro-Glu (PE) and Pro-Pro-Glu (PPE) motifs near the amino termini, respectively. (*A*) The most ancestral members of these families consist of only the PE/PPE domains, and the genes coding for them are often organized in operons. (*B*) During evolution, the PE/PPE proteins appear to have acquired extended C-terminal domains, which may be entirely unrelated in sequence, or share characteristic motifs. The most recently evolved PE/PPEs, the PE_PGRS and PPE_MPTR proteins, have C-terminal domains that contain multiple tandem repeats of glycine-rich motifs (114, 124). (*A* and *B*) were adapted from (124). (*C*) PE25 (black) and PPE41 (grey) interact to form a highly alpha-helical, elongated structure (PDB ID: 2g38). Note that the extreme N and C termini of PE25 and the C terminus of PPE41 were not resolved in the determined structure (127). The figure was generated with Jmol.

the host. Consistently, an increasing body of evidence suggests that PE and PPE proteins interfere with host immune responses (see references within (124)). For instance, inactivation of *ppe25* rendered *M. avium* incapable of preventing phagolysosomal fusion, and affected the survival of this bacterium in macrophages and in mice (125). Furthermore, the PE/PPE proteins have immunogenic properties (124), and some have been proposed to serve as source of antigenic variation, or to interfere with antigen processing (126). However, these hypotheses are based on the polymorphic nature and apparent redundancy of the repetitive sequences in the C-terminal portions of PE_PGRS and PPE_MPTR proteins, and still await experimental confirmation (124, 126).

PE and PPE proteins tend to be unstable and/or poorly soluble when expressed individually in heterologous hosts such as *Escherichia coli* (127), and consequently very little structural information is available for these proteins. Strong *et al* (127) hypothesized that the failure to produce soluble protein could be due to the absence of a partner protein, and reasoned that coexpression of a cognate partner might increase the solubility of the recombinant proteins. By coexpressing a *pe* and a *ppe* gene that are encoded in an operon (128), they could successfully purify PE25 and PPE41, and solve the structure of the heterodimer (Fig. 5*C*) (127). This structure revealed a mainly α–helical conformation for both proteins, with PE25 folding in two antiparallel helices that are connected by a short loop, and PPE41 being composed of five helices. The two helices of PE25 interact with two antiparallel helices of PPE41 to form an elongated four-helix bundle, which is stabilized by extensive hydrophobic interactions in the center of the complex. Indeed, the complex formation prevents exposure of apolar regions in both proteins, which might explain the poor solubility of PE25 and PPE41 when expressed individually (127). Interestingly, the four-helix bundle formed by PE25/PPE41 is reminiscent of the structure of the complex formed by ESAT-6 and CFP-10 (80). There are 13 additional *pe/ppe* operons in *M. tuberculosis*, whose gene products are likely to interact in a similar manner to PE25 and PPE41 (65, 114, 127-129). However, the majority of the PE and PPE proteins are encoded by isolated genes (65, 114), and it is unclear if the structure of the PE25/PPE41 complex is representative for these proteins. Based on the generated structure, a number of unpaired PE and PPE proteins were predicted to have cognate interaction partners that are encoded elsewhere in the genome (129). Therefore it is possible that many, if not all, PE and PPE proteins interact to form structurally similar complexes via their N-terminal domains (129).

Several *pe/ppe* genes are located within ESX clusters (65, 114), and bioinformatics analyses suggest that the *pe/ppe* families and the ESX loci are also evolutionarily related (114). Comparative genomics and phylogenetics indicated that the *pe/ppe* gene pair within the ESX-1 region comprises the most ancestral members of the two families, and that the duplications of the ESX loci gave rise to additional *pe/ppe* genes. These analyses also showed that the high numbers of *pe/ppe* genes found in the genomes of mycobacterial pathogens are the result of an extensive expansion that took place after the evolution of the most recent ESX cluster, *i.e.* ESX-5 (114). Interestingly, the PE/PPE families and the ESX systems also appear to be functionally linked, as several PE and PPE proteins have been shown to be substrates of the ESX systems, particularly the ESX-5 system. Indeed, while the export of a single PE and

PPE protein has been attributed to the ESX-1 system (74, 75), the ESX-5 system appears to be a major export pathway for PE/PPEs, especially for the most recently evolved members, the PE_PGRS and PPE_MPTR proteins (81, 83, 86). Analyses of truncated forms of one of the most well studied PE proteins, PE_PGRS33, suggested that the conserved N-terminal PE domain targets this protein to the cell wall (118, 130). However, similar to other T7S substrates the PE/PPE proteins lack detectable secretion signals and it is not known how they are targeted for secretion.

SCOPE OF THE THESIS

The aim of the research described in this thesis was to investigate how proteins secreted by the mycobacterial type VII secretion systems are recognized and targeted for secretion, with main focus on the enigmatic PE and PPE proteins that are secreted via the ESX-5 system.

In **chapter 2** the *M. tuberculosis* PE protein LipY and its *M. marinum* homologue, which contains an N-terminal PPE domain, are shown to be exported to the surface of *M. marinum* by the ESX-5 secretion system. These two proteins are subsequently used as model substrates to investigate the role of the N-terminal PE and PPE domains in ESX-5 secretion. The role of the PE domain in secretion is further investigated in **chapter 3**, using the newly identified ESX-5 substrate PE_PGR33 as model.

Chapter 4 describes a detailed search for specific sequences that are required for ESX-5 secretion. In this study, secretion of the previously identified ESX-5 substrate PPE41 (83), and its partner PE25 is investigated. These two proteins form a complex for which the structure has been elucidated (127), and inspection of this structure allowed us to select flexible or exposed residues that might be involved in protein-protein interaction and secretion. The role of these residues and regions were examined by deletion and mutation analysis.

The study described in **chapter 4** resulted in the identification of a motif that is required for secretion via ESX-5, but which is not sufficient to direct the substrates to this system. This means that an additional signal must be present that targets the substrates to the correct secretion machinery. Proteins destined for export across the inner membrane or out of the bacterial cell are often delivered to their cognate secretion system by cytosolic chaperones (131-133). In **chapter 5**, we characterize the function of the cytosolic ESX-5 system component $EspG_5$, and investigate if this protein interacts with substrates of ESX-5.

Finally, the results obtained in this thesis are summarized and discussed in **chapter 6**.

REFERENCES

1. Lawn SD & Zumla AI (2011) Tuberculosis. *Lancet* 378:57-72.
2. Ortner DJ & Putschar WGJ (1981) Identification of pathological conditions in human skeletal remains. Smithsonian Institution Press, Washington DC.
3. Nuorala E, *et al.* (2004) MTB complex DNA in a Scandinavian neolithic passage grave. *In* Molecular palaeopathology. Ancient DNA analyses of the bacterial diseases tuberculosis and leprosy. Archaeological Research Laboratory, Stockholm.
4. Zink A, *et al.* (2001) Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J Med Microbiol* 50:355-366.
5. Taylor GM, *et al.* (2005) Genotypic analysis of the earliest known prehistoric case of tuberculosis in Britain. *J Clin Microbiol* 43:2236-2240.
6. Daniel TM (2006) The history of tuberculosis. *Respir Med* 100:1862-1870.
7. Hershkovitz I, *et al.* (2008) Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS One* 3:e3426.
8. Brosch R, *et al.* (2002) A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* 99:3684-3689.
9. Wirth T, *et al.* (2008) Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog* 4:e1000160.
10. Rothschild BM, *et al.* (2001) *Mycobacterium tuberculosis* complex DNA from an extinct bison dated 17,000 years before the present. *Clin Infect Dis* 33:305-311.
11. World Health Organization (2011) Global tuberculosis control: WHO Report 2011. WHO/HTM/TB/2011.16, Geneva.
12. Gandhi NR, *et al.* (2010) Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *Lancet* 375:1830-1843.
13. World Health Organization (2006) The Stop TB Strategy: building on and enhancing DOTS to meet the TB-related Millennium Development Goals. Geneva.
14. World Health Organization (2004) BCG Vaccine. *Wkly Epidemiol Rec* 4:27-38.
15. Russell DG, *et al.* (2010) Tuberculosis: what we don't know can, and does, hurt us. *Science* 328:852-856.
16. Pieters J (2008) *Mycobacterium tuberculosis* and the macrophage: maintaining a balance. *Cell Host Microbe* 3:399-407.
17. Cosma CL, *et al.* (2003) The secret lives of the pathogenic mycobacteria. *Annu Rev Microbiol* 57:641-676.
18. Russell DG (2007) Who puts the tubercle in tuberculosis? *Nat Rev Microbiol* 5:39-47.
19. World Health Organization (2012) Tuberculosis fact sheet N°104. http://www.who.int/mediacentre/factsheets/fs104/en/index.html, March 2012.
20. Kaufmann SH, *et al.* (2010) New vaccines for tuberculosis. *Lancet* 375:2110-2119.
21. Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
22. Tamura K, *et al.* (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.
23. Tortoli E (2006) The new mycobacteria: an update. *FEMS Immunol Med Microbiol* 48:159-178.
24. Wayne LG & Kubica GP (1986) Section 16: the mycobacteria, p. 1435-1457. *In* Buchanan RE & Gibbons NE (eds.), Bergey's manual of systematic bacteriology, 9th Ed. The Williams & Wilkins Co, Baltimore.
25. Stahl DA & Urbance JW (1990) The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. *J Bacteriol* 172:116-124.
26. Rastogi N, *et al.* (2001) The mycobacteria: an introduction to nomenclature and pathogenesis. *Rev Sci Tech* 20:21-54.
27. Alderton H & Smith D (2001) Safety in the laboratory, p. 367-369. *In* Parish T & Stoker NG (eds.), Methods in molecular medicine: *Mycobacterium tuberculosis* protocols, vol. 54. Humana Press, New Jersey.
28. Gupta UD & Katoch VM (2005) Animal models of tuberculosis. *Tuberculosis (Edinb)* 85:277-293.
29. Shiloh MU & DiGiuseppe Champion PA (2009) To catch a killer. What

can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis? *Curr Opin Microbiol* 13:86-92.

30. Tonjum T, *et al.* (1998) Differentiation of *Mycobacterium ulcerans*, *M. marinum*, and *M. haemophilum*: mapping of their relationships to *M. tuberculosis* by fatty acid profile analysis, DNA-DNA hybridization, and 16S rRNA gene sequence analysis. *J Clin Microbiol* 36:918-925.
31. Petrini B (2006) *Mycobacterium marinum*: ubiquitous agent of waterborne granulomatous skin infections. *Eur J Clin Microbiol Inf Dis* 25:609-613.
32. Davis JM, *et al.* (2002) Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. *Immunity* 17:693-702.
33. Stoop EJ, *et al.* (2011) Zebrafish embryo screen for mycobacterial genes involved in the initiation of granuloma formation reveals a newly identified ESX-1 component. *Dis Model Mech* 4:526-536.
34. Swaim LE, *et al.* (2006) *Mycobacterium marinum* infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity. *Infect Immun* 74:6108-6117.
35. Stinear TP, *et al.* (2008) Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res* 18:729-741.
36. Gao LY, *et al.* (2004) A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol Microbiol* 53:1677-1693.
37. Volkman HE, *et al.* (2004) Tuberculous granuloma formation is enhanced by a *Mycobacterium* virulence determinant. *PLoS Biol* 2:e367.
38. Brennan PJ & Nikaido H (1995) The envelope of mycobacteria. *Annu Rev Biochem* 64:29-63.
39. Barry CE, 3rd, *et al.* (1998) Mycolic acids: structure, biosynthesis and physiological functions. *Prog Lipid Res* 37:143-179.
40. Daffe M & Etienne G (1999) The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity. *Tuber Lung Dis* 79:153-169.
41. Hoffmann C, *et al.* (2008) Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc Natl Acad Sci U S A* 105:3963-3967.
42. Zuber B, *et al.* (2008) Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J Bacteriol* 190:5672-5680.
43. Bhamidi S, *et al.* (2011) Detailed structural and quantitative analysis reveals the spatial organization of the cell walls of *in vivo* grown *Mycobacterium leprae* and *in vitro* grown *Mycobacterium tuberculosis*. *J Biol Chem* 286:23168-23177.
44. Villeneuve M, *et al.* (2005) Temperature dependence of the Langmuir monolayer packing of mycolic acids from *Mycobacterium tuberculosis*. *Biochim Biophys Acta* 1715:71-80.
45. Villeneuve M, *et al.* (2007) Conformational behavior of oxygenated mycobacterial mycolic acids from *Mycobacterium bovis* BCG. *Biochim Biophys Acta* 1768:1717-1726.
46. Draper P (1998) The outer parts of the mycobacterial envelope as permeability barriers. *Front Biosci* 3:D1253-1261.
47. Niederweis M, *et al.* (2010) Mycobacterial outer membranes: in search of proteins. *Trends Microbiol* 18:109-116.
48. Van der Geize R, *et al.* (2007) A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proc Natl Acad Sci U S A* 104:1947-1952.
49. Bloch H & Segal W (1956) Biochemical differentiation of *Mycobacterium tuberculosis* grown *in vivo* and *in vitro*. *J Bacteriol* 72:132-141.
50. Abdallah AM, *et al.* (2007) Type VII secretion--mycobacteria show the way. *Nat Rev Microbiol* 5:883-891.
51. Finlay BB & Falkow S (1997) Common themes in microbial pathogenicity revisited. *Microbiol Mol Biol Rev* 61:136-169.
52. DiGiuseppe Champion PA & Cox JS (2007) Protein secretion systems in mycobacteria. *Cell Microbiol* 9:1376-1384.
53. Berks BC (1996) A common export pathway for proteins binding complex

redox cofactors? *Mol Microbiol* 22:393-404.

54. von Heijne G (1985) Signal sequences. The limits of variation. *J Mol Biol* 184:99-105.
55. Rigel NW & Braunstein M (2008) A new twist on an old pathway--accessory Sec systems. *Mol Microbiol* 69:291-302.
56. Tseng TT, *et al.* (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* 9 Suppl 1:S2.
57. Andersen P, *et al.* (1995) Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice. *J Immunol* 154:3359-3372.
58. Skjot RL, *et al.* (2000) Comparative evaluation of low-molecular-mass proteins from *Mycobacterium tuberculosis* identifies members of the ESAT-6 family as immunodominant T-cell antigens. *Infect Immun* 68:214-220.
59. Berthet FX, *et al.* (1998) A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). *Microbiology* 144:3195-3203.
60. Sorensen AL, *et al.* (1995) Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect Immun* 63:1710-1717.
61. Mahairas GG, *et al.* (1996) Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* 178:1274-1282.
62. Brodin P, *et al.* (2002) Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect Immun* 70:5568-5578.
63. Cole ST, *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393:537-544.
64. Gey Van Pittius NC, *et al.* (2001) The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol* 2:RESEARCH0044.
65. Tekaia F, *et al.* (1999) Analysis of the proteome of *Mycobacterium tuberculosis in silico*. *Tuber Lung Dis* 79:329-342.
66. Pym AS, *et al.* (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol Microbiol* 46:709-717.
67. Guinn KM, *et al.* (2004) Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* 51:359-370.
68. Hsu T, *et al.* (2003) The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc Natl Acad Sci U S A* 100:12420-12425.
69. Stanley SA, *et al.* (2003) Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc Natl Acad Sci U S A* 100:13001-13006.
70. Converse SE & Cox JS (2005) A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. *J Bacteriol* 187:1238-1245.
71. McLaughlin B, *et al.* (2007) A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. *PLoS Pathog* 3:e105.
72. Xu J, *et al.* (2007) A unique *Mycobacterium* ESX-1 protein co-secretes with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation. *Mol Microbiol* 66:787-800.
73. Carlsson F, *et al.* (2009) Polar localization of virulence-related Esx-1 secretion in mycobacteria. *PLoS Pathog* 5:e1000285.
74. Sani M, *et al.* (2010) Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathog* 6:e1000794.
75. Fortune SM, *et al.* (2005) Mutually dependent secretion of proteins required for mycobacterial virulence. *Proc Natl Acad Sci U S A* 102:10676-10681.
76. DiGiuseppe Champion PA, *et al.* (2009) ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. *Mol Microbiol* 73:950-962.
77. Bitter W, *et al.* (2009) Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog* 5:e1000507.
78. Ilghari D, *et al.* (2011) Solution structure of the *Mycobacterium tuberculosis*

EsxG-EsxH complex: functional implications and comparisons with other *M. tuberculosis* Esx family complexes. *J Biol Chem* 286:29993-30002.

79. Pallen MJ (2002) The ESAT-6/WXG100 superfamily -- and a new Gram-positive secretion system? *Trends Microbiol* 10:209-212.
80. Renshaw PS, *et al.* (2005) Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J* 24:2491-2498.
81. Abdallah AM, *et al.* (2009) PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. *Mol Microbiol* 73:329-340.
82. Siegrist MS, *et al.* (2009) Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. *Proc Natl Acad Sci U S A* 106:18792-18797.
83. Abdallah AM, *et al.* (2006) A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Mol Microbiol* 62:667-679.
84. Christie PJ, *et al.* (2005) Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol* 59:451-485.
85. Ohol YM, *et al.* (2010) *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host Microbe* 7:210-220.
86. Bottai D, *et al.* (2012) Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. *Mol Microbiol 83:1195-209.*
87. Champion PA, *et al.* (2006) C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. *Science* 313:1632-1636.
88. MacGurn JA & Cox JS (2007) A genetic screen for *Mycobacterium tuberculosis* mutants defective for phagosome maturation arrest identifies components of the ESX-1 secretion system. *Infect Immun* 75:2668-2678.
89. Pukatzki S, *et al.* (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc Natl Acad Sci U S A* 104:15508-15513.
90. Lewis KN, *et al.* (2003) Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. *J Infect Dis* 187:117-123.
91. Pym AS, *et al.* (2003) Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat Med* 9:533-539.
92. Wards BJ, *et al.* (2000) An *esat6* knockout mutant of *Mycobacterium bovis* produced by homologous recombination will contribute to the development of a live tuberculosis vaccine. *Tuber Lung Dis* 80:185-189.
93. Brodin P, *et al.* (2006) Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. *Infect Immun* 74:88-98.
94. Abdallah AM, *et al.* (2011) Mycobacterial secretion systems ESX-1 and ESX-5 play distinct roles in host cell death and inflammasome activation. *J Immunol* 187:4744-4753.
95. van der Wel N, *et al.* (2007) *M. tuberculosis* and *M. leprae* translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* 129:1287-1298.
96. de Jonge MI, *et al.* (2007) ESAT-6 from *Mycobacterium tuberculosis* dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. *J Bacteriol* 189:6028-6034.
97. Smith J, *et al.* (2008) Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in *Mycobacterium marinum* escape from the vacuole. *Infect Immun* 76:5478-5487.
98. Brodin P, *et al.* (2005) Functional analysis of early secreted antigenic target-6, the dominant T-cell antigen of *Mycobacterium tuberculosis*, reveals key residues involved in secretion, complex formation, virulence, and immunogenicity. *J Biol Chem* 280:33953-33959.
99. Houben D, *et al.* (2012) ESX-1 Mediated translocation to the cytosol controls virulence of Mycobacteria. *Cell Microbiol 14:1287-1298.*
100. Coros A, *et al.* (2008) The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. *Mol Microbiol* 69:794-808.
101. Flint JL, *et al.* (2004) The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. *Proc Natl Acad Sci U S A* 101:12598-12603.

102. Sassetti CM, *et al.* (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48:77-84.

103. Rodriguez GM, *et al.* (2002) *IdeR*, an essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun* 70:3371-3381.

104. Maciag A, *et al.* (2007) Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. *J Bacteriol* 189:730-740.

105. Serafini A, *et al.* (2009) Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: essentiality and rescue by iron and zinc. *J Bacteriol* 191:6340-6344.

106. Abdallah AM, *et al.* (2008) The ESX-5 secretion system of *Mycobacterium marinum* modulates the macrophage response. *J Immunol* 181:7166-7175.

107. Weerdenburg EM, *et al.* (2012) ESX-5-deficient *Mycobacterium marinum* is hypervirulent in adult zebrafish. *Cell Microbiol* 14:728-739.

108. Flynn JL & Chan J (2001) Immunology of tuberculosis. *Annu Rev Immunol* 19:93-129.

109. Beltan E, *et al.* (2000) Secretion of cytokines by human macrophages upon infection by pathogenic and non-pathogenic mycobacteria. *Microb Pathog* 28:313-318.

110. Akpe San Roman S, *et al.* (2010) A heterodimer of EsxA and EsxB is involved in sporulation and is secreted by a type VII secretion system in *Streptomyces coelicolor*. *Microbiology* 156:1719-1729.

111. Burts ML, *et al.* (2005) EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *Proc Natl Acad Sci U S A* 102:1169-1174.

112. Poulet S & Cole ST (1995) Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch Microbiol* 163:87-95.

113. Hermans PW, *et al.* (1992) Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. *J Bacteriol* 174:4157-4165.

114. Gey van Pittius NC, *et al.* (2006) Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (*esx*) gene cluster regions. *BMC Evol Biol* 6:95.

115. Deb C, *et al.* (2006) A novel lipase belonging to the hormone-sensitive lipase family induced under starvation to utilize stored triacylglycerol in *Mycobacterium tuberculosis*. *J Biol Chem* 281:3866-3875.

116. Kruh NA, *et al.* (2010) Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome *in vivo*. *PLoS One* 5:e13938.

117. Banu S, *et al.* (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol Microbiol* 44:9-19.

118. Brennan MJ, *et al.* (2001) Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun* 69:7326-7333.

119. Delogu G, *et al.* (2004) *Rv1818c*-encoded PE_PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure. *Mol Microbiol* 52:725-733.

120. Demangel C, *et al.* (2004) Cell envelope protein PPE68 contributes to *Mycobacterium tuberculosis* RD1 immunogenicity independently of a 10-kilodalton culture filtrate protein and ESAT-6. *Infect Immun* 72:2170-2176.

121. He Z & De Buck J (2010) Localization of proteins in the cell wall of *Mycobacterium avium* subsp. *paratuberculosis* K10 by proteomic analysis. *Proteome Sci* 8:21.

122. Le Moigne V, *et al.* (2005) Expression, immunochemical characterization and localization of the *Mycobacterium tuberculosis* protein p27. *Tuberculosis (Edinb)* 85:213-219.

123. Malen H, *et al.* (2010) Definition of novel cell envelope associated proteins in Triton X-114 extracts of *Mycobacterium tuberculosis* H37Rv. *BMC Microbiol* 10:132.

124. Sampson SL (2011) Mycobacterial PE/PPE proteins at the host-pathogen interface. *Clin Dev Immunol* 2011:497203.

125. Li Y, *et al.* (2005) A *Mycobacterium avium* PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. *Cell Microbiol* 7:539-548.

126. Brennan MJ & Delogu G (2002) The PE multigene family: a 'molecular mantra' for mycobacteria. *Trends Microbiol* 10:246-249.

127. Strong M, *et al.* (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 103:8060-8065.

128. Tundup S, *et al.* (2006) Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. *FEBS Lett* 580:1285-1293.

129. Riley R, *et al.* (2008) Identifying cognate binding pairs among a large set of paralogs: the case of PE/PPE proteins of *Mycobacterium tuberculosis*. *PLoS Comput Biol* 4:e1000174.

130. Cascioferro A, *et al.* (2007) PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall. *Mol Microbiol* 66:1536-1547.

131. Pattis I, *et al.* (2007) The *Helicobacter pylori* CagF protein is a type IV secretion chaperone-like molecule that binds close to the C-terminal secretion signal of the CagA effector protein. *Microbiology* 153:2896-2909.

132. Randall LL & Hardy SJ (2002) SecB, one small chaperone in the complex milieu of the cell. *Cell Mol Life Sci* 59:1617-1623.

133. Wattiau P, *et al.* (1994) Individual chaperones required for Yop secretion by *Yersinia*. *Proc Natl Acad Sci U S A* 91:10493-10497.