PPE’S AND THE ESX 5 SECRETION SYSTEM
SHINING LIGHT ON THE DARK MATTER OF MYCOBACTERIA

Louis Simon Ates
PPE’S AND THE ESX-5 SECRETION SYSTEM
SHINING LIGHT ON THE DARK MATTER OF MYCOBACTERIA

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op woensdag 30 maart 2016 om 9.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Louis Simon Ates
geboren te Amsterdam
Promotoren: prof.dr. W. Bitter
           prof.dr. C.M.J.E. Vandenbroucke-Grauls

Copromotor: dr. E.N.G. Houben
Nobody said it was easy.
No one ever said it would be this hard.
I’m going back to the start.
*Coldplay, The Scientist*

You can’t lose ‘em all – and how to find what your not looking for
*Coparck*
<table>
<thead>
<tr>
<th>Chapter 1</th>
<th>General Introduction and Scope of the Thesis</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoofdstuk 1</td>
<td>Achtergrond en Samenvatting van het Proefschrift</td>
<td>17</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>Type VII Secretion: A Highly Versatile Secretion System</td>
<td>25</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Essential Role Of The ESX-5 Secretion System In Outer Membrane Permeability Of Pathogenic Mycobacteria</td>
<td>53</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>The ESX-5 System of Pathogenic Mycobacteria Is Involved In Capsule Integrity and Phagosomal Rupture Through Its Substrate PPE10</td>
<td>95</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Hypervirulent Mycobacterium tuberculosis Isolates Do Not Secrete PE_PGRS Proteins Due To Mutations In The ppe38 Locus</td>
<td>123</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>Exploiting ESX-5: optimizing LipY as a carrier for heterologous secretion in Mycobacteria</td>
<td>151</td>
</tr>
<tr>
<td>Summarizing Discussion</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>Dankwoord/Acknowledgements</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>Literature</td>
<td>191</td>
<td></td>
</tr>
</tbody>
</table>
GENERAL INTRODUCTION

Tuberculosis

Mycobacteria form a group of environmental and pathogenic bacteria that distinguish themselves by unique capacities of their cell envelope. Several species of mycobacteria are able to cause diseases in humans and other animals. The most well-known of these pathogenic mycobacteria is *Mycobacterium tuberculosis*, the causative agent of tuberculosis. In western countries tuberculosis is often seen as a solved problem of the past. However, *M. tuberculosis* is still the most deadly bacterial pathogen worldwide. Each year nine million new cases of tuberculosis occur, and one-and-a-half million people die of this infectious disease (World Health Organization 2013). Tuberculosis is spread through inhalation of contaminated aerosols and only a low dose of *M. tuberculosis* bacilli is required to establish infection. Although *M. tuberculosis* spreads efficiently, the bacteria themselves grow very slowly. In comparison: *Escherichia coli* is able to divide every 20 minutes, while *M. tuberculosis* divides every 24 hours. This sounds as a disadvantage for the bacterium, but is actually an important part of the successful infection strategy of *M. tuberculosis*. The slow growth of mycobacteria is combined with the ability to slow down the bacterial metabolism into a state of dormancy, where by active infection turns to latent infection. This allows *M. tuberculosis* to stay undetected in the body of the host for years, until the infection is reactivated, usually because of an imbalance in the host's immune system. Malnourishment and HIV infection are important risk factors for reactivation to severe forms of tuberculosis that often result in death (World Health Organization 2013). Tuberculosis also used to be referred to as “the consumption”, which gives an indication of the state of a patient with progressive tuberculosis. *M. tuberculosis* primarily causes infection of the lungs, which in most cases remains occurs without symptoms. However, reactivation of the disease can occur in a situation of reduced immunity of the host, leading to dissemination of the bacteria within and outside of the lungs. This may lead to chronic coughing associated with bloody sputum and weight loss as important symptoms. In advanced tuberculosis, other organs can also be infected with very serious symptoms and death as a consequence. Treatment of tuberculosis consists of a cocktail of at least 4 different antibiotics that need to be taken for several months (World Health Organization 2013). This long and intense treatment leads to compliance problems in the treatment of tuberculosis, which in turn have led to antibiotic resistance of *M. tuberculosis* against first- and second-line antibiotics. Rising drug resistance has resulted in multi-drug-resistant (MDR) or extremely-drug-resistant (XDR) strains of *M. tuberculosis*. The prevalence of these resistant strains is increasing (World Health Organization 2013). Treatment of XDR-tuberculosis is extremely challenging, requiring cocktails of over ten different second-line antibiotics, associated with severe side-effects, that need to be taken for a period of at least 18 months years (World Health Organization 2013). In some cases, treatment of XDR-tuberculosis needs to be performed in quarantine of the patient in the first
months of treatment. This leads to stigmatization of patients, fear of healthcare facilities and rips apart families and social structures. Even if proper healthcare is available, treatment success rates of XDR-tuberculosis remain below 50%, especially in patients co-infected with HIV. Treatment of a single XDR-tuberculosis patient in a western country is estimated to cost approximately €450.000,- (World Health Organization 2013). Even though XDR-tuberculosis is still rare, the rise of these antibiotic resistant strains make it evident that improvements of current tuberculosis treatment and prevention strategies urgently need to be developed. The current vaccine against tuberculosis is a live attenuated bacterium called BCG. BCG was developed in the 1920’s by Calmette and Guérin in France. Their vaccine was based on Mycobacterium bovis, which causes a tuberculosis-like disease in cattle and humans. This strain was cultured in the laboratory for 10 years; over this period it accumulated genetic mutations. These mutations resulted in attenuation of the strain, i.e. it lost its ability to cause disease in humans, although vaccination with BCG still causes immune responses that help protect the vaccinated individuals. These immune-responses elicited by BCG confer partial protection against the most severe forms of tuberculosis in children, but unfortunately are not able to prevent reactivation of disease and development of open lung tuberculosis (hence transmission of M. tuberculosis) in adults and adolescents.

The worldwide increase in incidence of tuberculosis cases that was partially caused by the HIV epidemic has been halted in the last decade. However, only a marginal reduction in the number of tuberculosis cases and deaths has been achieved, indicating that the elimination of tuberculosis still requires intensified efforts (World Health Organization 2013). In order to achieve the millennium goal of the World Health Organization to eradicate tuberculosis by 2050, both new vaccines and new antibiotics are needed.

Mycobacteria are responsible for other diseases than just tuberculosis. Leprosy for instance, which is caused by Mycobacterium leprae, is still not eradicated and remains a health problem in developing countries. But also the less commonly known disease Buruli ulcer, which is caused by Mycobacterium ulcerans, is a debilitating disease that requires complex treatments. Several other species of mycobacteria can cause disease in animals or can cause opportunistic infections in humans. Mycobacterium marinum is the causative agent of a tuberculosis-like disease in poikilothermic animals such as fish and frogs. It is very closely related to M. tuberculosis and M. ulcerans, however, and can also cause an infection, known as fish tank granuloma, in humans (van der Sar et al. 2004). This has led to the use of M. marinum as a model organism for tuberculosis research.

A defining characteristic, which sets mycobacteria apart from Gram-negative and Gram-positive bacteria, is their unique cell envelope. Mycobacteria possess an outer membrane (Zuber et al. 2008; Hoffmann et al. 2008), which consists of mycolic acids. These long-chain (C_{n0}-C_{n9}) fatty acids are covalently linked to a peptidoglycan/arabinogalactan periplasmic layer (P J Brennan and Nikaido 1995). This hydrophobic
permeability barrier makes mycobacteria extremely resistant to the harmful effects of antibiotics and host defence mechanisms such as oxygen radicals or antibiotic peptides (Purdy, Niederweis, and Russell 2009; Gao et al. 2003). This feature also allows mycobacteria to withstand the extremely harsh environment in the macrophage compartment, known as the phagosome, for limited amounts of time. A more loosely associated capsular layer is present outside the outer membrane of the mycobacteria (Sani et al. 2010; Daffé and Etienne 1999; Ortalo-Magné et al. 1996). This capsule consists of proteins, polysaccharides and glycolipids, components that have been implicated in immune modulation and pathogenesis (Geurtsen et al. 2009; Gagliardi et al. 2007; Geijtenbeek et al. 2003), but the function of the capsule in pathogenesis remains largely unknown.

Pathogenesis and Protein secretion
When *M. tuberculosis* enters the body, usually by inhalation of infected aerosols, bacteria are recognized by lung macrophages. Macrophages take up the mycobacteria and contain them in their phagosomes. The phagosomal compartment is meant to fuse with the lysosome. For most other bacteria this process would mean quick death and eradication from the body as the lysosome is filled with anti-bacterial molecules such as antibiotic peptides, nitrogen and oxygen radicals, and proteases. However, *M. tuberculosis* is protected from these attacks by its impermeable cell envelope and additionally secretes proteinaceous virulence factors that allow the bacteria survive. These proteins allow *M. tuberculosis* to delay the fusion of phagosomes with lysosomes (Stewart et al. 2005), but also allow the mycobacteria to even rupture the phagosome and enter the macrophage cytoplasm (van der Wel et al. 2007; Simeone et al. 2012; Simeone et al. 2015). This environment is considered less hostile for mycobacteria than the phagolysosome, and it contains more nutrients. To secrete proteins through their unique cell envelope, mycobacteria have developed specialized secretion systems known as type VII secretion (T7S) systems (Bitter et al. 2009). *M. tuberculosis* contains five different of these T7S systems called ESX-1 to ESX-5 (Bitter et al. 2009). These secretion systems have only been discovered over the last two decades; they have been shown to be absolutely essential for the pathogenesis of tuberculosis (Reviewed in Chapter 2) (Pym et al. 2002). The ESX-1 system is the best-studied of the T7S systems and its substrates are responsible for the escape of pathogenic mycobacteria from the phagosome (van der Wel et al. 2007; Simeone et al. 2012). The ESX-3 system is involved in the uptake of iron and zinc, although it is not precisely understood by what mechanism this occurs (Serafini et al. 2013; Siegrist et al. 2009; Siegrist et al. 2014; Serafini et al. 2009). The ESX-5 system and its substrates are the subject of this thesis; they have many enigmatic characteristics. ESX-5, the most recently evolved T7S system, is only found in the slow-growing mycobacteria, most of which are pathogenic (Gey van Pittius et al. 2006). ESX-5 is responsible for the secretion of the majority of the so-called PE and PPE proteins (Abdallah et al. 2009; Ates et al. 2015). An important subgroup of the PE proteins, the PE_PGRS
proteins, so named because of their polymorphic GC-rich sequences (Cole et al. 1998). The biological role of ESX-5 and the PE and PPE proteins, is still largely unknown, although PE_PGRS proteins have been implicated to be in immune modulation (reviewed in chapter 2).

**SCOPE OF THIS THESIS**

The focus of this thesis is to acquire more insight into the functioning of the ESX-5 system, and the role of its substrates. A common first approach to investigate the role of a protein is to try to make a mutation in the gene that encodes this protein. However, work by my colleagues already indicated that it was not possible to make mutations in the conserved components of the ESX-5 system; i.e. the genes of interest were essential for growth. In chapter 3 we set out to prove that the ESX-5 system was indeed essential for growth. We discovered that when genes involved in the biogenesis of specific outer membrane lipids were mutated, the mutations in esx-5 were no longer lethal. Since the lack of these membrane lipids led to a more leaky outer membrane, we hypothesized that we could achieve the same effect by introducing an outer membrane porin from *M. smegmatis*, a non-pathogenic species of mycobacterium, which does not have ESX-5. After introduction of this porin in *M. marinum* we could indeed create several different ESX-5 mutants; this allowed us to study the ESX-5 system in much more detail than before. One of the results was that substrates of the ESX-5 system are involved in the uptake of fatty acids and probably also other nutrients.

In chapters 4 and 5 we focused on the role of two different substrates secreted by the ESX-5 system. Both studies were instigated by earlier screens in which we tried to identify ESX-5 negative secretion mutants (Abdallah et al. 2006; Ates et al. 2015; van der Woude et al. 2012). In chapter 4 we describe the identification of a mutant in the gene ppe10 which seemed to secrete increased amounts of PE_PGRS proteins. These proteins were, however, not produced or secreted in higher amounts, but were more loosely attached to the cell surface, because of reduced capsular integrity. We visualized the morphology of the capsule of this esx-5 mutant strain with electron microscopy and showed that the capsule of this mutant was malformed. Strains with reduced capsular integrity had fewer ESX-1 substrates on their cell surface and therefore were impaired in phagosomal rupture in the early stages of infection in both macrophage- and zebrafish infection experiments. That ESX-5 plays a role in capsule integrity is an important new finding; we were able to pinpoint this phenotype to a single ESX-5 substrate: PPE10.

In chapter 5 we identified a mutant in the gene encoding another substrate of ESX-5: PPE38. A *M. marinum* strain with a mutation in ppe38 appeared unable to secrete PE_PGRS proteins, although the ESX-5 system itself was not affected. When we made similar mutations in the orthologue of ppe38 in *M. tuberculosis* we saw that these mutants were also unable to secrete PE_PGRS proteins, while other T7S substrates were secreted normally. In earlier literature we found that certain clinical
**GENERAL INTRODUCTION AND SCOPE OF THE THESIS**

*M. tuberculosis* strains, belonging to the ‘Beijing’ lineage, have naturally occurring mutations in *ppe38*, and surrounding genes, due to genetic recombination events (McEvoy, Warren, et al. 2009; McEvoy, van Helden, et al. 2009). We obtained these clinical strains from the group of Rob Warren at Stellenbosch University in South Africa. We were able to show that these clinical strains are also deficient in the secretion of specific subsets of the PE and PPE proteins (PE_PGRS and PPE_MPTR). Furthermore, when we reintroduced *ppe38* and the surrounding genes in these isolates, these strains secreted PE_PGRS proteins like wild type *M. tuberculosis*. In collaboration with Professor Rogelio Hernandez-Pando in Mexico we performed *in vivo* experiments in mice to test the virulence of these strains. The naturally occurring *ppe38* mutants were hypervirulent in this murine model. Mice infected with clinical isolates with the *ppe38* mutation were more ill and died earlier than mice infected with wildtype strains. The hypervirulent phenotype could be partially reversed by reintroducing the *ppe38* locus in the *ppe38* mutants. This study showed that PPE38 is required for the secretion of specific groups of ESX-5 substrates and that loss of PPE38 and concomitant secretion defects lead to hypervirulence in mice of clinical *M. tuberculosis* isolates.

In chapter 6 we attempted to translate our findings on ESX-5 secretion to achieve secretion of non-mycobacterial proteins (heterologous secretion) through ESX-5; Heterologous secretion can be useful for vaccine design, or for fundamental research purposes. For this purpose, we investigated the possibility to use LipY, a substrate of the ESX-5 system which has been intensively studied by our group (Daleke et al. 2011; Daleke, Ummels, et al. 2012). We first defined the minimal domains that are required for the secretion of LipY. We fused these domains to an artificial protein derived from chicken egg ovalbumin (OVA), and investigated if this non-mycobacterial protein was secreted through the ESX-5 system. The secretion of the LipY-OVA fusion protein was indeed successful, but very inefficient. Therefore we set up a method to find mutations that would increase the secretion of this construct. By introducing random mutations in the construct with an error-prone DNA-polymerase and checking for colonies with more efficient secretion we were able to significantly optimize the secretion of LipY-OVA, thereby showing that this is a good method to optimize heterologous secretion.

The results obtained in this thesis are summarized and discussed in chapter 7, the summarizing discussion.
ACHTERGROND
EN SAMENVATTING
VAN HET PROEFSCHRIFT
ACHTERGROND EN SAMENVATTING VAN HET PROEFSCHRIFT

Tuberculose

Mycobacteriën zijn een groep van bacteriën die verschillende ziektes kunnen veroorzaken in mensen en dieren, maar ook in grote getalen als onschuldige bodembacteriën voorkomen. De bekendste ziekteverwekkende mycobacterie is Mycobacterium tuberculosis, de bacterie die de ziekte tuberculose veroorzaakt. Tuberculose wordt in westerse landen nog wel eens gezien als een probleem van vroeger, maar is wereldwijd nog steeds de meest dodelijke bacteriële ziekteverwekker. Anderhalf miljoen mensen sterven elk jaar aan tuberculose en negen miljoen nieuwe mensen worden ziek. Tuberculose wordt overgedragen via aërosolen in de lucht en kan zich op deze manier snel door de populatie verspreiden. *M. tuberculosis* groeit zelf echter extreem langzaam en kan zelfs in een staat van metabole inactiviteit geraken, welke latente tuberculose genoemd wordt. Dit klinkt wellicht als een nadeel, maar het stelt de bacterie in staat om jarenlang een onopgemerkte infectie te veroorzaken. Op het moment dat het immuunsysteem van de patiënt achteruit gaat, bijvoorbeeld door ondervoeding of door een secundaire infectie zoals HIV, kan de bacterie zich reactiveren met alle verschrikkelijke gevolgen voor de patiënt en zijn omgeving. De ziekte tuberculose werd in de volksmond voorheen wel tering genoemd, afkomstig van het woord vertering. Patiënten met actieve tuberculose vermageren en verliezen hun energie, totdat ze vel over been zijn. Op het moment dat de bacterie zich door de rest van het lichaam verspreid kunnen vrijwel alle lichaamsdelen geïnfecteerd en beschadigd raken, met de dood als gevolg, wanneer behandeling uitblijft. Een antibioticumkuur tegen *M. tuberculosis* bestaat uit een medicijncocktail van minstens 4 verschillende medicijnen die meerdere maanden geslikt moet worden. Door de bijwerkingen en de lange duur van de behandeling is het moeilijk voor patiënten om deze behandeling af te maken. Hierdoor zijn in de afgelopen decennia steeds meer antibioticumresistente *M. tuberculosis* stammen gaan circuleren. De meest ernstige varianten hiervan, genaamd XDR (extremely drug resistant)-tuberculose, zijn zelfs wanneer alle medische faciliteiten aanwezig zijn bijna niet te behandelen. Een XDR-tuberculose patiënt moet gedurende minstens twee jaar lang een cocktail van tien tot twintig verschillende medicijnen slikken en injecteren. Deze medicijnen hebben vaak zeer ernstige bijwerkingen en een deel van de behandeling moet in quarantaine worden uitgevoerd. Zelfs wanneer de patiënt in deze gevallen wordt opgenomen in een ziekenhuis en alle mogelijke zorg ontvangt is de kans dat de patiënt overlijdt aan XDR-tuberculose nog steeds vaak boven 50%. De behandeling van een XDR-tuberculose patiënt in een westers land kost naar raming bijna €500.000,- per patiënt. XDR-tuberculose is nu nog relatief zeldzaam, maar zonder nieuwe antibiotica en vaccin strategieën zullen deze gevallen ongetwijfeld meer en meer toenemen.

Er is een vaccin tegen tuberculose genaamd Bacille de Calmette et Guérin (BCG). Dit vaccin is al in de jaren ’20 van de vorige eeuw ontwikkeld in Frankrijk. Dit vaccin is gebaseerd op Mycobacterium bovis, de tegenhanger van *M. tuberculosis* die een
tuberculose-achtige ziekte veroorzaakt in koeien en mensen. Deze bacterie is jaren in het lab doorgekweekt en gedurende dat proces zijn allerlei mutaties opgetreden. Door deze mutatie is de bacterie nog wel in staat om ons immuunsysteem te activeren, maar kan deze de mens niet meer ziek maken. BCG is sindsdien al aan 3 miljard mensen toegediend en heeft waarschijnlijk vele levens gered. Het vaccin beschermt namelijk kinderen tegen de ernstigste vormen van TB, zoals hersenvliesontsteking en infectie van de bloedbaan. Helaas is het BCG vaccin niet in staat om een zodanig immuunrespons op te wekken dat het adolescenten en volwassenen kan beschermen tegen het oplopen van tuberculose en kan het daarom de verspreiding van tuberculose niet tegenhouden.

In het afgelopen decennium is de opmars van tuberculose wereldwijd een halt toegeroepen, maar een grote daling van aantal gevallen en doden is helaas nog niet in zicht. Om de wereldwijde tuberculose epidemic te kunnen bedwingen zullen er daarom ook nieuwe antibiotica op de markt moeten verschijnen en zal een beter vaccin ontwikkeld moeten worden, dat ook de verspreiding van tuberculose kan tegengaan.

Pathogenese en eiwitsecretie
Tuberculose is niet de enige ziekte die wordt veroorzaakt door mycobacteriën. Ook bijvoorbeeld lepra, veroorzaakt door Mycobacterium leprae, is nog steeds een probleem in meerdere landen, net als Buruli ulcer, veroorzaakt door Mycobacterium ulcerans. Bovendien zijn er vele soorten mycobacteriën die dieren kunnen infecteren of in de bodem of andere plekken voorkomen en in zeldzame gevallen ook ziekte in de mens kunnen veroorzaken. Mycobacterium marinum, veroorzaakt een tuberculose-achtige ziekte in vissen en koudbloedige amfibieën zoals kikkers, maar kan ook in mensen een zogenaamd aquarium granuloom veroorzaken. Tevens is M. marinum een belangrijk modelorganisme voor het onderzoek naar M. tuberculosis wat ook in mijn studies veelvuldig gebruikt is.

Mycobacteriën onderscheiden zich van andere bacteriën, onder andere, door hun unieke celervelop. Deze bestaat uit een binnenmembraan dat vergelijkbaar is met dat van andere bacteriën, maar ook uit een uniek buitenmembraan. Dit buitenmembraan bestaat uit extreem lange vetzuren die een zeer ondoorlopende beschermlaag om de bacteriën vormen. Deze beschermlaag maakt de bacterie extreem ongevoelig voor antibiotica en stelt de bacterie in staat om te overleven in menselijk macrofagen; immuuncellen die eigenlijk dienen om bacteriën detecteren en op te ruimen. Om het buitenmembraan heen zit het mycobacteriële kapsel, wat bestaat uit eiwitten, suikerstructuren en vetten. Deze laag is slechts losjes bevestigd aan de bacterie en de rol van dit kapsel wordt nog niet volledig begrepen. Wel is bekend dat componenten van dit kapsel belangrijk zijn voor de virulentie van de bacterie alsmede voor het moduleren van het immuunsysteem van de gastheer.

Wanneer ons lichaam via de longen wordt geïnfecteerd door M. tuberculosis, zal de bacterie herkend worden door macrofagen in de longen. Zij nemen de bacteriën op
en zullen deze isoleren in het fagosoom, een subcompartiment van macrofaag. Voor de meeste bacteriën zou dit het einde betekenen van hun avontuur in ons lichaam, aangezien het fagosoom gemaakt is om bacteriën op te ruimen. Het fagosoom kan fuseren met een ander onderdeel van de macrofaag, het lysosoom, welke is gevuld met zuren, radicale stikstof- en zuurstofmoleculen en eiwitten die de bacterie aanvallen en afbreken. De bacterie is door zijn ondoorgrondbare celenuitvouw in staat om deze aanval in een initieel stadium te overleven, maar heeft vele strategieën ontwikkeld om ook op langere termijn in dit milieu te kunnen overleven. De meeste bekende strategieën die de bacterie hiervoor gebruikt, worden uitgevoerd door eiwitten die de bacterie uitscheid. Het uitscheiden, of secreteren, van deze eiwitten is natuurlijk een uitdaging, omdat de bacterie een zeer ondoorgrondbare celenuitvouw heeft. Mycobacteriën hebben dan ook eigen secretiesystemen ontwikkeld om eiwitten te kunnen secreteren over hun speciale dubbele membraan, een fenomeen wat Type VII secretie (T7S) genoemd. M. tuberculosis heeft vijf verschillende T7S systemen die ESX-1 tot en met ESX-5 genoemd worden. Deze secretiesystemen zijn slechts relatief kort bekend, maar blijken extremee belangrijk te zijn in het functioneren van mycobacteriën. Het ESX-1 systeem is een van de belangrijkste virulentiefactoren van mycobacteriën. Wanneer ESX-1 gemuteerd is, zoals in de vaccinstam BCG ook het geval is, is de bacterie niet meer in staat om ziekte te veroorzaken. Ten minste een van de belangrijke functies van het ESX-1 systeem is het ontsnappen uit het fagosoom. De eiwitten gecodeerd door ESX-1 zijn in staat om de membraan van de fagosomen te doorbreken, waardoor de bacterie dit gevaarlijke milieu kan ontsnappen en in het cytosol van de macrofaag veel meer voedingsstoffen tot zijn beschikking heeft. Het ESX-3 systeem is betrokken in de opname van ijzer en zink, maar het moleculaire proces wat hier voor verantwoordelijk is, is nog onduidelijk. Het ESX-5 systeem is de focus van dit proefschrift en is een zeer intrigerend secretiesysteem. ESX-5 is het meest recent geëvolueerde van de T7S systemen van M. tuberculosis en komt alleen voor in de langzaamgroeiende mycobacteriën, waarne vrijwel alle pathogenen behoren. Het ESX-5 systeem is verantwoordelijk voor de secretie van tientallen eiwitten behorende tot de klasse van de zogenaamde PE en PPE eiwitten. De rol van deze eiwitten is tot op heden volledig onbekend. Een subgroep van de PE eiwitten, de zogenaamde PE_PGRS eiwitten worden vaak geassocieerd met een rol in het moduleren van de immuunresponsen van de mens, maar sluitend bewijs hiervoor, of een mechanisme is helaas nog niet gevonden.

HET PROEFSCHRIFT
Het doel van dit proefschrift is om meer inzicht te krijgen in de mechanismen van T7S, maar vooral ook om een rol te vinden voor de eiwitten die gecodeerd worden door het ESX-5 systeem. De eerste manier waarop een microbioloog over het algemeen de functie van een gen en eiwit probeert te bepalen is door een mutatie in het desbetreffende gen te maken. Het bleek echter niet mogelijk te zijn om mutaties in de belangrijkste componenten van het ESX-5 systeem te maken. In genetische
termen bleken deze genen ‘essentieel’ te zijn. In hoofdstuk 3 beschrijven we dit proces. Om meer inzicht te krijgen in de rol van het ESX-5 systeem gingen we op zoek naar bacteriële mutanten die wel konden overleven zonder het ESX-5 systeem. Wij vonden dat we in stammen met een “lek” buitenmembraan wel mutaties konden maken in het esx-5-locus. Vervolgens hebben we dit gebruikt om te laten zien, dat esx-5-mutanten in M. marinum niet in staat waren om vetzuren op te nemen en daarmee lijkt het aannemelijk dat ESX-5 betrokken is bij het opnemen van verschillende voedingsstoffen over het buitenmembraan.

In hoofdstuk 4 en 5 hebben we gekeken naar de functies van eiwitten die gesecreteerd worden door ESX-5. Bij de zoektocht naar mutanten in het ESX-5 systeem vonden we ook mutanten die niet minder, maar juist meer ESX-5 substraten leken te secreteren (Hoofdstuk 4). Eén van deze mutanten was gemuteerd in het gen wat codeert voor een eiwit dat zelf een ESX-5 substraat is: PPE10. Na nader onderzoek bleek de ppe10-mutant echter niet meer PE_PGRS eiwitten te secreteren, maar deze zaten losser aan het oppervlakte verbonden. Ook andere eiwitten die normaal gesproken op het oppervlakte van de bacterie gevonden kunnen worden, leken nu los rond te zweven in het groeimagazijn van de bacterie. Het bleek dat mutanten in ppe10 en ook esx-5 mutanten niet in staat waren om een goed gevormd kapsel te vormen. Hierdoor vielen de onderdelen van het kapsel af van het oppervlakte van de bacterie. Dit stelde ons in staat om de rol van het kapsel in mycobacteriën in infectie onderzoek. Hierbij vonden we dat de mutanten in ppe10 of esx-5 minder goed in staat waren om in de beginfase van infectie te overleven. Dit kwam doordat de kapselmutanten de ESX-1 eiwitten die op hun oppervlakte horen te zitten niet meer hadden en daardoor niet uit het fagosoom konden ontsnappen.

In hoofdstuk 5 beschrijven we een andere mutant van een substraat van het ESX-5 systeem. Dit gen, ppe38, bleek noodzakelijk te zijn om de PE_PGRS eiwitten te kunnen secreteren. Deze mutant werd, net als de mutanten in de andere hoofdstukken, eerst gevonden in M. marinum. Wij konden echter aantonen dat ook in M. tuberculosis een versie van dit gen zit en dat dit gen eenzelfde functie heeft. Uit literatuuronderzoek bleek dat stammen die zijn beschreven door de groep van Robin Warren in Zuid-Afrika natuurlijke mutaties in deze regio hadden. Wij hebben samengewerkt met zijn groep en konden aantonen dat ook in deze natuurlijke ppe38 mutanten niet in staat waren PE_PGRS eiwitten te secreteren. Bovendien gingen deze klinische stammen weer gewoon PE_PGRS eiwitten secreteren wanneer we dit gen weer inbrachten in deze stammen. In samenwerking met de groep van Rogelio Hernandez-Pando uit Mexico, zijn vervolgens dierproeven gedaan om de virulentie van deze stammen te bepalen. Het blijkt dat de natuurlijke ppe38 mutanten hypervirulent zijn; muizen geïnfecteerd met deze mutant hadden meer bacteriën in de longen en werden ernstiger ziek. Wanneer we het gen weer inbrachten in de stammen herstelde de virulentie zich. Deze data corresponderen met de observatie dat deze Zuid-Afrikaanse stam verantwoordelijk was voor een grote uitbraak onder patiënten. Deze data suggereren
dat de PE_PGRS eiwitten een onderduikmechanisme voor M. tuberculosis vormen en dat de bacterie agressiever wordt als het deze eiwitten niet kan secreteren.

In hoofdstuk 6 proberen we onze opgedane fundamentele kennis over ESX-5 secretie om te zetten naar een toepassing die gebruikt zou kunnen worden voor biotechnologie of vaccinontwikkeling. Wij hebben LipY, een substraat van ESX-5, genomen en in meer detail bepaald wat het gedeelte van LipY was dat verantwoordelijk was voor secretie. Vervolgens hebben we dit gedeelte van het gen gefuseerd aan OvaL, een stuk kunstmatig DNA wat gebruikt kan worden om vaccins te onderzoeken. Het ontstane fusie-eiwit werd maar mondjesmaat gecanteerd, maar door het te muteren konden we varianten vinden die betere secretie vertoonden. Hoewel een klinische toepassing van deze data nog ver weg is laten we met dit werk zien dat je T7S kan gebruiken voor het secreteren van eiwitten die normaal gesproken niet door mycobacteriën gemaakt worden en hoe dit secretieproces geoptimaliseerd kan worden.

In hoofdstuk 7 bespreek ik de tijdens mijn promotie verkregen data in de algemene discussie en probeer ik aan te geven wat deze resultaten betekenen voor het onderzoeksveld van de mycobacteriële eiwitsecretie.
TYPE VII SECRETION: A HIGHLY VERSATILE SECRETION SYSTEM

Louis S. Ates¹, Edith N.G. Houben² & Wilbert Bitter¹,²

CHAPTER SUMMARY

Type VII secretion (T7S) systems of mycobacteria secrete substrates over the unusual dierm cell-envelope. Furthermore, T7S gene clusters are present throughout the phylum Actinobacteria and functional T7S-like systems have been identified in Firmicutes. Most of the T7S substrates can be divided in two families: the Esx proteins, which are found in both Firmicutes and Actinobacteria, and the PE and PPE proteins, which are more mycobacterium-specific. Members of both families have been shown to be secreted as folded heterodimers, implying this is a conserved feature of T7S substrates. Most knowledge on the mechanism of T7S and the roles of T7S systems in virulence comes from studies on pathogenic mycobacteria. These bacteria can contain up to five different T7S systems, called ESX-1 to ESX-5, each having their own role in bacterial physiology and virulence.

Here, we discuss the general composition of T7S systems and the role of the individual components in secretion. These conserved components include two membrane proteins with (predicted) enzymatic activities: a predicted ATPase (EccC), likely to be required for energy provision of T7S, and a subtilisin-like protease (MycP) involved in processing of specific substrates. Additionally, we describe the role of a conserved intracellular chaperone in T7S substrate recognition, based on recently published crystal structures and molecular analysis.

Finally, we discuss system-specific features of the different T7S systems in mycobacteria and their role in pathogenesis and provide an overview of the role of T7S in virulence of other pathogenic bacteria.

1Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, the Netherlands
2Section Molecular Microbiology, Amsterdam Institute of Molecules, Medicine & Systems, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
TYPE VII SECRETION: A HIGHLY VERSATILE SECRETION SYSTEM

Bacterial secretion systems were initially studied in the Gram-negative bacterium *Escherichia coli* K-12. When researchers started to explore protein secretion in different Gram-negative bacteria and especially in bacterial pathogens, it was clear that *E. coli* K-12 was not able to present us with a complete picture of protein secretion systems. Type II, type III and type IV secretion systems were quickly discovered and revolutionized host-pathogen interaction studies. Gram-negative bacteria need these specialized secretion systems to transport proteins across two membranes (also called a diderm cell envelope). The presence of this complex cell envelope does not only mean that two membranes have to be crossed, but an additional problem is that there is no energy source at the outer membrane. This means that alternative mechanisms for protein transport needs to be present, such as coupling the energy of the inner membrane to protein transport across the outer membrane or to cross the entire cell envelope in a single step. Although the discovery of different secretion systems in Gram-negative bacteria was a major breakthrough, the downside has been that secretion systems in other bacteria have been neglected. It was generally thought that secretion in other bacteria, which are generally monoderm, would completely depend on the universal Sec or Tat system. Only in recent years this idea is shifting and again it started by studying pathogens, *i.e.* the pathogenic mycobacteria.

The genus *Mycobacterium* contains a number of important pathogens, including *Mycobacterium tuberculosis* and *Mycobacterium leprae*. These pathogens are generally slow-growing and have a distinctive growth pattern, known as cording. Although these bacteria genetically belong to the high-GC Gram-positive bacteria, they produce a second membrane composed of unique and complex lipids (Zuber et al. 2008; Sani et al. 2010). As such, the cell envelope of these bacteria is therefore diderm and protein secretion is as problematic as for Gram-negative bacteria. Among the first identified secreted proteins of *M. tuberculosis* were two small proteins, known as ESAT-6 and CFP-10 (Andersen et al. 1995). Later these proteins were renamed EsxA and EsxB (Bitter et al. 2009). These proteins lack any obvious canonical secretion signal (Sec or Tat) and therefore seemed to depend on a new secretion system. Detailed analysis of the tuberculosis vaccine strain *Mycobacterium bovis* BCG indicated that this strain not only had lost both the *esxAB* genes, but also several surrounding genes responsible for EsxAB secretion (Mahairas et al. 1996; Pym et al. 2003). Subsequent research showed that many of the genes surrounding *esxAB* are indeed part of this secretion system (Gey Van Pittius et al. 2001). This region is now known as the *esx-1* locus and contains in total 20 genes (Figure 1). With these genes identified, it quickly became clear that mycobacteria have several of these secretion systems. In fact, up to five different *esx*-loci can be present on the chromosome of mycobacterial species. Recently, additional ESX systems have been identified on conjugative mycobacterial plasmids (Ummels et al. 2014). On these plasmids, the esx clusters are located adjacent to a type IV-like secretion cluster and together they are required for the conjugation process (Ummels et al. 2014).
Figure 1. Genetic loci of different T7S (and T7S-like) systems. Depicted are the T7S loci, esx-1, esx-4 and esx-5 of M. tuberculosis H37Rv (Cole et al. 1998), as well the T7S-like systems of S. aureus (Str. USA300, annotation based on Anderson et al. (Anderson et al. 2013)) and B. subtilis subsp. subtilis (str. 168, annotation based on Huppert et al. (Huppert et al. 2014)). Color coding represents conserved T7S membrane components (dark blue), (putative) substrates of the systems (green), cytosolic chaperones (yellow) and Firmicute-specific T7S-like membrane components (light blue).
To emphasize that these ESX systems are required for secretion over the diderm mycobacterial cell envelope, they were named type VII secretion (Abdallah et al. 2007; Bitter et al. 2009). In literature the terms Wss (for WxG100 secretion system) or ESX have been infrequently used, but in this chapter we will use type VII secretion (T7S) as a general term for these secretion systems and ESX for the different mycobacterial secretion systems. T7S systems are not specific for pathogenic mycobacteria; they are also present in non-pathogenic mycobacteria and many other bacteria have homologous systems. Most of these bacteria, such as Rhodococcus, Corynebacterium and Nocardia species, are in fact closely related to mycobacteria and have a similar diderm cell envelope. Probably, the T7S systems of these species are involved in protein secretion as well, although currently there is no data supporting this. An interesting group of T7S homologs is present in several Firmicutes, including many different Bacillus and Staphylococcus species (Pallen 2002). However, these systems are only distantly related since only a homolog of the EccC membrane component (see below) and homolog(s) of the EsxA substrates are present in these bacteria. These latter systems have recently been shown to function as active secretion systems (Burts et al. 2005; Huppert et al. 2014). Clearly, there is an evolutionary link between the diderm T7S systems and the newly identified Firmicute secretion systems, but nomenclature could be an issue. Previously, we suggested that they might be called type VIIb system, in analogy to the type IV secretion systems where similar heterogeneity occurs (Abdallah et al. 2007). We will comprehensively discuss the mycobacterial T7S systems, which have been studied most intensively, and give an overview of the similarities with the Firmicute T7S-like secretion systems, with a focus on the nature of the substrates and the conserved secretion signal.

Composition And Functioning Of T7S In Mycobacteria

T7S gene clusters of M. tuberculosis usually share a number of conserved genes that belong to ten different gene families. A general nomenclature for these proteins was agreed on in an early phase (Bitter et al. 2009); the name ESX-conserved component (Ecc) is used for components that are present in most systems and ESX-1-specific components (Esp) for components that are (mostly) unique for the best-studied secretion system, ESX-1. Finally, for some conserved proteins the old and well-established names were kept, i.e. Mycosins (MycP), Esx, PE and PPE proteins. Within the esx loci we can find genes that encode both substrates and structural components. In addition to the Esx proteins, PE and PPE proteins and most Esp proteins are probably substrates. Thus far, there seem to be seven conserved genes that are required for the secretion process and can therefore be considered as core components of the secretion machinery. Two of these core components, EspG and EccA, are localized to the mycobacterial cytosol, while the other five, EccB, EccC, EccD, EccE and MycP, contain transmembrane domains (TMDs) and reside in the mycobacterial cell envelope (Figure 2). All esx gene clusters contain members of these ten conserved gene families, except the most archaic cluster, i.e. esx-4,
Figure 2. Model for T7S in Mycobacteria. The conserved membrane components (blue) form a complex, in which the EccC homolog is the ATPase possibly providing energy for the secretion process. The mycosin (MycP) is not part of the core complex, but is essential for successful secretion. The T7S substrates (green) are secreted dependently on the conserved signals YxxxD/E and WxG (red). Secretion of PE/PPE dimers is dependent on the cytosolic chaperones EspG and EccA (yellow). While EspG binds to the substrate pair in the cytosol, EccA might be involved in releasing this chaperone from the PE/PPE dimer upon contact with the membrane complex. In contrast, Esx proteins are not recognized by EspG and their dependence on the cytosolic chaperones might be indirect due to interdependence of Esx and PE/PPE for secretion. The EspB monomer has a similar fold to PE-PPE dimers and contains the putative secretion signal. Upon translocation EspB is processed and forms a heptamer with a barrel-like structure. Whether PE_PPE dimers adopt a similar quaternary structure is yet unknown. Secreted substrates can localize to the culture supernatant or remain attached in the capsular layer. Whether the secretion process is a one- or two-step process is not known, therefore a putative outer membrane component (grey) is indicated by the question mark.
that lacks pe, ppe, espG, eccA and eccE genes (Figure 1). First, we will discuss the two cytosolic conserved components of mycobacterial T7S systems and then the membrane components.

EspG
EspG was initially not considered a T7S core component (Bitter et al. 2009), as these proteins share relative low sequence identity (20-25%). However, recent experiments have clearly shown that these proteins are structural and functional equivalents. To acknowledge this, it would be more appropriate to rename this component Esx conserved component, but we realize that the current name is already established and should be continued. The first indication for a role of EspG in secretion was a study describing that EspG$_5$ is required for the secretion of PE/PPE proteins via the ESX-5 system (Daleke et al. 2011; Abdallah et al. 2009). Subsequently, immuno-precipitation experiments showed that EspG$_5$ specifically interacts with the heterodimeric model substrates PE25/PPE41 (Daleke, Woude, et al. 2012) (for dimer formation of T7S substrates see below). Additional biophysical analysis of the EspG$_5$-PE25/PPE41 complex showed a tight interaction. Because EspG$_5$ is located in the cytosol and could not be identified in the culture supernatant it probably dissociates from the substrate pair upon contacting the secretion machinery. Interestingly, the interaction of EspG with PE/PPE proteins shows considerable system specificity, as EspG$_5$ does not interact with the PE35/PPE68$_1$ protein pair that is secreted via ESX-1 (Daleke, Woude, et al. 2012). Conversely, EspG$_1$ of the ESX-1 system does not bind the ESX-5-dependent PE25/PPE41, but does interact with the ESX-1 protein pair PE35/PPE68$_1$. This suggests that EspG proteins are cytosolic chaperones that specifically recognize their cognate PE/PPE substrates. This component might therefore determine through which T7S system these substrates are transported. The discovery that EspG proteins are PE/PPE-specific chaperones also explains why the esx-4 locus is lacking espG, in addition to pe and ppe genes. 

Recently, crystal structures of EspG$_5$ in complex with PE25/PPE41 were obtained (Figure 3C) (Ekiert and Cox 2014; Korotkova et al. 2014). The structures reveal that the EspG$_5$ protein has a novel mixed α/β-fold. The crystal structure of EspG$_3$ of the ESX-3 system was solved in one of these studies (Ekiert and Cox 2014). Despite the low similarity on sequence level, EspG$_3$ has a highly similar fold, confirming that these proteins have a similar function. The structure of the EspG$_5$-PE25/PPE41 complex shows that EspG$_5$ binds to the tip of the PPE protein, which is not involved in the interaction with the PE protein. This tip region of PPE contains some hydrophobic residues that are now buried in a hydrophobic groove formed by a central β-sheet and C-terminal α-helical bundle of EspG$_5$. Subsequent mutagenesis in the EspG$_5$-recognition motif of various ESX-5 dependent PPE proteins showed that mutations that abolish EspG$_5$ binding in vitro usually affect substrate secretion (Korotkova et al. 2014). Additionally, co-expression of EspG proteins increased the in vitro solubility of PE/PPE pairs, which hints towards a role of EspG in keeping these protein pairs...
Figure 3. Crystal structures of T7S substrates. A) Structure of the heterodimer EsxB (dark blue) and EsxA (light blue) of M. tuberculosis (3FAV). The two proteins form a four-helix bundle. The Tyr of the YxxxD motif and the Gly and Trp residues of the WxG motif that are postulated to constitute together the T7S secretion signal are shown in red (Poulsen et al. 2014). B) The EsxA protein of S. aureus forms a homodimer that results in two putative secretion signals (VxxxD) on each end of the four-helix bundle (in red) (2VRZ) (Sundaramoorthy, Fyfe, and Hunter 2008). C) Crystal structure of PE25 (light blue) and PPE41 (dark blue) of M. tuberculosis in complex with the chaperone EspG₅ (yellow). EspG interacts with the PPE protein through hydrophobic interactions, but not directly with the PE protein. The WxG motif on the PPE and the YxxxE motif on the PE protein together form a putative T7S signal (red residues) (4KXR) (Korotkova et al. 2014). D) Crystal structure of monomeric EspB visualizing an extended secretion signal that includes the YxxxD/E and WxG motif (red residues) (3J83) (Solomonson et al. 2015).

soluble. Recently published x-ray diffraction and electron microscopy data of the ESX-1 substrate EspB not only shows that this protein has a highly similar fold as PE/PPE heterodimers (see below, Figure 3D), but also that it can form pore-like quaternary structures (Solomonson et al. 2015). Modelling suggests that EspB could be organized as a heptamer and that the multimerization is mediated by its putative
EspG-binding domain. Therefore, the EspG proteins could be responsible for preventing this multimerization to occur intracellularly. Together, these data suggest that EspG proteins are chaperones involved in substrate recognition and perhaps preventing their aggregation in the cytosol.

Although the EspG proteins seem to play a central role in recognition of the PE and PPE substrates in the cytosol, they do not interact with the Esx proteins (Daleke, Woude, et al. 2012). Perhaps the intrinsic solubility of Esx protein pairs makes a role for the EspG chaperone superfluous. EspG might therefore be a dedicated chaperone of PE/PPE proteins (and perhaps several Esp proteins such as EspB) that are more prone to aggregate.

EccA

EccA is the second cytosolic protein encoded by most mycobacterial T7S systems. EccA belongs to the AAA+ (ATPases associated with various cellular activities) protein family, members of which are involved in diverse processes, including protein degradation, signal transduction and (dis)assembly of protein complexes. AAA+ proteins typically form ring-shaped hexamers with a central pore. EccA, also forms oligomers, possibly hexamers, upon overexpression in E. coli (Wagner, Evans, and Korotkov 2014). ATP hydrolysis by these hexamers usually induces a conformational change in the complex that is transferred to the bound substrate(s). All EccA homologs are composed of two conserved domains joined by a linker region. The C-terminal domain of EccA is the ATPase domain and contains all the motifs that are characteristic for AAA+ proteins. In line with this, in vitro ATPase activity of EccA can be pinpointed to this C-terminal domain (Luthra et al. 2008). The N-terminus of EccA contains six tetratricopeptide repeat (TPR) motifs, known for mediating protein-protein interactions in other proteins (Cerveny et al. 2013). The structure of this N-terminal domain of M. tuberculosis EccA has recently been solved, showing an arrangement of these motifs in a right-handed superhelix (Wagner, Evans, and Korotkov 2014). Interestingly, the structure of this N-terminal domain of EccA resembles the structure of PilF, which is involved in the assembly of the type IV pili system in Pseudomonas aeruginosa (Wagner, Evans, and Korotkov 2014).

EccA is important for T7S, as disruption of eccA genes in esx-1 and esx-5 clusters results in the loss of secretion of Esx and PE/PPE proteins (Bottai et al. 2012; Abdallah et al. 2006; Gao et al. 2004). In addition, a mutation in the ATPase domain of EccA disrupts secretion (Joshi et al. 2012) and EccA is, similarly to other core components of the ESX-3 system, essential for viability (Sassetti, Boyd, and Rubin 2003). However, EccA-independent secretion through ESX-1 and ESX-5 (E. N. G. Houben et al. 2012; Converse and Cox 2005) (Houben and Bitter unpublished results) has been observed. These results suggest that either EccA plays different roles in different mycobacterial species/strains, or that the function of EccA is redundant in specific cases. Interestingly, EccA is, just like EspG, not present in ESX-4. This protein therefore could also play a role in the secretion of specific substrates, such as the PE and PPE proteins.
Perhaps in analogy with the AAA+ protein in the type VI secretion system, EccA could be involved in the disassembly of EspG chaperones from the PE/PPE dimers (Ekiert and Cox 2014).

The type VII secretion membrane complex

After synthesis and folding in the cytosol, the T7S substrates should be guided to the inner membrane, where their transport across the mycobacterial cell envelope is initiated. For the other specialized secretion systems, a complex machinery is involved in the actual secretion process. In line with this, all five conserved T7S membrane components, EccB, EccC, EccD, EccE and MycP have been shown to be essential for protein secretion by ESX-1 (Brodin et al. 2006; E. N. G. Houben et al. 2012; Stanley et al. 2003), ESX-5 and ESX-3 (E. N. G. Houben et al. 2012; Bottai et al. 2012; Siegrist et al. 2014). Most of these membrane components have large N- or C-terminal hydrophilic domains and only one or two predicted TMDs. EccD is the exception. This protein is highly hydrophobic, usually having 11 predicted TMDs and only a relatively small N-terminal hydrophilic part. Solely based on this characteristic, EccD has been postulated to form the membrane pore through which substrates are transported (Stanley et al. 2003), although there is no functional proof for this. First evidence about the composition of the transport channel was provided by the observation that four of the conserved components of the ESX-5 system, i.e. EccB, EccC, EccD and EccE, form a stable membrane complex of ~1.5 MDa (E. N. G. Houben et al. 2012). It is highly likely that this large complex forms the channel through which substrates are transported, although evidence for channel activity has not yet been provided. Because the substrates are probably secreted as (hetero)dimers, the translocation pore should be relatively large to allow passage of these folded structures. Of the four components of the membrane complex, only EccE is not present in all T7S systems; again ESX-4 lacks this component. This could suggest that this protein is located at the periphery of the complex, which is supported by the observation that EccE is highly sensitive to proteolytic degradation when the membrane complex is treated with trypsin (E. N. G. Houben et al. 2012).

The presence of classical TMDs with high hydrophobicity predicts that all the Ecc components are inserted in the inner membrane of the mycobacterial cell envelope. As substrates also need to cross the outer membrane to end up in the extracellular environment, the question remains whether the observed membrane complex also inserts into the outer membrane. The identification of an outer membrane channel is complicated by the fact that our knowledge of mycobacterial outer membrane proteins is limited (Michael Niederweis et al. 2010). The only mycobacterial outer membrane protein that is studied in detail, MspA, does show structural similarity to the outer membrane proteins of Gram-negative bacteria, as MspA spans the outer membrane using short β-sheet transmembrane domains that are organized in a so-called β-barrel. Although the size of the T7S membrane complex would allow it to span both the inner and the outer membrane, none of the subunits has a substantial domain with
predicted beta-sheets that could form a β-barrel. Perhaps, outer membrane transport is mediated by other (unidentified) proteins that more loosely associate with the T7S membrane complex. Of course other structures, like amphipathic α-helices, such as those involved in formation of the Type IV secretion complex (Chandran et al. 2009), could play a role. However, these latter unusual structures are more difficult to predict in silico.

Another option is that the T7S system is only involved in transport across the inner membrane and that outer membrane transport depends on a separate transport system, which would mean a two-step process. Although hard evidence for a one step secretion process across the complex mycobacterial cell envelope is still missing, there is currently also no data that the T7S substrates are exposed to the mycobacterial periplasm at any point during the translocation process (Rosenberger, Brülle, and Sander 2013).

**EccC**

Of the four subunits of the T7S membrane complex, only EccC has predicted functional domains; it contains three conserved nucleotide binding domains (NBDs). EccC is one of the most conserved T7S components; it is not only present in all mycobacterial T7S systems, but is also the only protein that is present in T7S-like systems in Firmicutes. Apparently, EccC homologs are central players in these secretion systems. While EccC is usually encoded by a single gene, the homolog in the ESX-1 system of *M. tuberculosis* is composed of two distinct proteins. However, it is most likely that these two EccC subunits together form a functional unit (Stanley et al. 2003) and originated from a single gene. All three predicted NBDs of EccC show homology to members of the FtsK/SpoIIIE family of ATPases. This large protein family consists of ATPases that are involved in a wide range of cellular processes, of which FtsK and SpoIIIE, involved in DNA transport (Burton and Dubnau 2010), are best-studied. Similar as for the AAA+ ATPases, FtsK/SpoIIIE-like ATPases usually form ring-like hexamers, which seems to suggest that EccC is functional as a hexameric complex as well. However, as EccC has three successive FtsK/SpoIIIE-like ATP-binding domains instead of one, this protein could have different characteristics. Other type IV secretion systems also contain a member of the FtsK/SpoIIIE protein family, the type IV coupling protein (T4CP). This protein plays a key role in substrate recognition and protein transport (Atmakuri, Cascales, and Christie 2004). Possibly, EccC performs similar functions in T7S by recognizing chaperone-substrate complexes at the membrane. Accordingly, EccC of the ESX-1 system was shown to interact with the substrate EsxB in both immuno-precipitation and yeast-two-hybrid experiments. This interaction was dependent on the presence of the C-terminal secretion signal of EsxB (Champion et al. 2006). Although EccC proteins have three NBDs (Ramsdell et al. 2014), mutational analysis of individual NBDs of EccC molecules of both mycobacteria and firmicutes suggests distinct roles of each domain in secretion; while ATP binding to the first domain is essential for secretion, the other two domains are not strictly required for
protein transport (Ates et al. 2015). In addition, ATP binding to any of the three NBDs of EccC does not seem to be required for formation of the ESX-5 membrane complex (Ates et al. 2015), suggesting that ATP hydrolysis by EccC is not involved in complex assembly, but that it is dedicated for substrate recognition and/or transport. Recent structural analysis of EccC from the thermophilic actinobacterium Thermomonospora curvata (Tc) with and without bound substrate TcEsxB provides important insight in the role of the individual NBDs in substrate recognition and transport (Rosenberg et al. 2015). In this study, Rosenberg et al. described the structure of TcEccC in complex with a peptide representing the TcEsxB secretion signal, revealing that the signal binds into a hydrophobic pocket of NBD3, while both NBD2 and NBD3 were bound to ATP (Figure 4). In contrast, NBD1 is visualized in a nucleotide-free

Figure 4. Crystal structures of EccC of Thermomonospora curvata. C-terminal domains of TcEccC containing all three NBDs (A) or containing only NBD2 and 3 and with a bound secretion signal of TcEsxB (B) are shown as described by Rosenberg et al. 2015 (Rosenberg et al. 2015). The secretion signal (in green) is bound to a hydrophobic pocket of NBD3. While NBD2 and 3 have a bound ATP molecule (red), NBD1 has a sulfate ion at the ATP binding site instead (in orange). NBD1 activity is inhibited by a linker domain of NBD2. This inhibition can be alleviated by changing arginine 543 (the orange residue) to an alanine.
state, suggesting that NBD2 and 3 are relevantly inefficient ATPases. The hydrophobic binding pockets of NBD1 and NBD2, do not bind the secretion signal, but are filled by a linker region of the adjacent NBD3. Mutating the linker of NBD2 that binds NBD1 (R543A) subsequently activated ATPase activity by NBD1 and this activity is even increased upon binding of TcEsxB to NBD3. These results suggest that substrate binding activates a chain of events leading to ATPase activation. In line with the previously obtained data in mycobacteria and B. subtilis (Ramsdell et al. 2014; Ates et al. 2015), the authors showed that mutations in catalytic residues of NBD1 completely abolished ATP hydrolysis, while similar mutations in NBD2 and NBD3 only reduced this activity. Rosenberg et al. provide evidence that TcEccC(R543A) multimerizes in vitro upon binding of TcEsxB, indicating that this substrate triggers both ATPase activity and multimerization. However, addition of TcEsxA, most-likely the binding partner of TcEsxB, abolishes both multimerization and ATPase activity of this TcEccC construct. How the in vitro multimerization of EccC correlates with the settings in the bacterial cell, where EccC is already multimeric as part of a large membrane complex (E. N. G. Houben et al. 2012), remains to be investigated.

Mycosin
The only conserved membrane protein of T7S systems that is not part of the ESX-5 membrane complex is mycosin (MycP). Mycosins are subtilisin-like proteases with a classical signal sequence that presumably directs the protease domain to the periplasm. In addition mycosins have a putative C-terminal TMD to anchor the protein in the membrane (Brown et al. 2000; Dave et al. 2002). Although mycosins are essential for T7S, their role in secretion is still an enigma. The only known substrate for any mycosin is the ESX-1 substrate EspB, which is an ESX-1 secreted protein cleaved by MycP₁ (Ohol et al. 2010). In addition, structural analysis of EspB with and without the C-terminus that is cleaved by MycP₁ shows that MycP₁ might be involved in inducing a conformational change of the EspB quaternary structure (Solomonson et al. 2015). Although this finding suggests that mycosins are involved in substrate processing, the proteolytic activity of MycP₁ seems to be dispensable for the secretion process (Ohol et al. 2010). Similarly, MycP₅ of the ESX-5 system is essential for ESX-5-dependent secretion, but its protease activity is not (van Winden, Houben and Bitter, unpublished results). This surprising observation suggests that mycosins have a second role in secretion, besides their function as a protease. Further research is needed to understand this crucial additional role of mycosins in T7S.

Recently, the structures of both M. smegmatis and M. thermoresistibile MycP were solved, which show a typical subtilisin core domain decorated with several extended loops (44, 45). The most distinctive structural feature of mycosins is the presence of an N-terminal extension. Many subtilisins are produced with an N-terminal extension, called a propeptide, that prevents premature substrate access to the active site (Shinde and Inouye 1995). However, classical subtilisin propeptides form a tightly folded structure near the active site and this propetide is usually degraded by
the subtilisin-like proteases after folding and/or transport. The N-terminal extension of mycosins does not display homology with subtilisin propeptides and is structurally different as well, as it is wrapped around the protease domain and does not block the active site (44, 45). In addition, MycP₁ with an intact N-terminal extension is able to cleave EspB in vitro (44, 45). Therefore, the term propeptide for this N-terminal extension of mycosins is probably incorrect. Interestingly, the structure of M. smegmatis MycP₃ was solved, revealing a very similar fold of the protease domain to the MycP₁ structure, including the N-terminal extension (45). The properties of the active site clefts are quite distinct between these two mycosins, implying different substrate specificities.

**T7S Substrates**

Although the different T7S systems that have been studied in detail secrete different classes of substrates, most of these substrates do belong to the EsxAB clan (Pfam CL0352) (47). This clan contains six protein families: Esx (WxG100), PE, PPE, LXG, DUF2563 and DUF2580. The best-known of these is the Esx family. The first Esx protein that was discovered is the EsxA (ESAT-6) protein of Mycobacterium tuberculosis, which is secreted as a heterodimer together with EsxB (CFP-10). Since then, members of this protein family have been described over a wide range of species, mostly in the phyla of the Actinobacteriae and the Firmicutes (11, 48), but have also been found in Verucomicrobia, Lentinospherae, Planctomycetes, Chloroflexi and even some in Proteobacteria (47). Esx proteins are also called WxG100 proteins (Pallen 2002), based on a short conserved motif, tryptophan-X-glycine (WxG), in the middle of the protein and their typical size of approximately 100 amino acids (Figure 3A). Esx proteins form two helices separated by a turn, which contains the conserved WxG motif. Although most Esx proteins are small, some contain an extended C-terminal domain. The C-terminal domains of these extended WxG100 proteins are highly variable and, although largely uncharacterized, are predicted to have highly divergent functions. Esx proteins of mycobacteria, such as EsxA and EsxB, are secreted as antiparallel heterodimers (49, 50). These co-secreted dimers are usually encoded by adjacent genes and are part of the same operon (51). In contrast, some Esx proteins from Streptococcus (52) and Staphylococcus (53) exclusively form homodimers and are encoded by genes that are generally mono-cistronic (Figure 3B). It has been suggested that the bi-cistronic heterodimers evolved after a duplication event, which suggests that the original substrates were homodimers (Poulsen et al. 2014).

The structure of the EsxAB complex (55) shows that the N- and C-termini of both EsxA and EsxB are predominantly unstructured. The longest of these unordered stretches, the C-terminus of EsxB, merits special attention since this region has been shown to contain a secretion motif that is required for secretion (see below). The overall structure of Esx dimers is also observed for other Esx proteins, such as EsxGH (56) and EsxRS (57).
PE and PPE proteins

Two other major classes of the T7S substrates belonging to the EsxAB clan are the PE and PPE proteins. Although they belong to different protein families, these proteins form secreted heterodimers and, similar to the mycobacterial Esx substrates, their genes are often located adjacent on the genome. Therefore, they will be discussed together in this section. PE proteins are named for their conserved proline-glutamic acid (PE) motifs at position eight and nine of the N-terminus of these proteins, whereas PPE proteins are defined by a proline-proline-glutamic acid motif at position 7 to 9. PE and PPE proteins have only been described in members of the Actinobacteria and are most widespread in the slow-growing species of Mycobacteria (Gey van Pittius et al. 2006). The structure of PE25 of \textit{M. tuberculosis} in complex with PPE41 has been solved by Strong and colleagues (Strong et al. 2006) and shows an antiparallel dimer forming a four-helix bundle, similar to that of the Esx proteins. The recently solved crystal structure of EspB has revealed that this protein is similar to a PE-PPE dimer (Figure 3D)(18). Interestingly, this protein multimerizes as a heptamer, indicating that a similar confirmation could be formed by PE-PPE dimers.

PE proteins are characterized by a conserved N-terminal PE domain of approximately 110 amino acids, which forms a helix-turn-helix structure similar to the Esx proteins. Only the turn is less defined as compared to the Esx proteins and does not contain the WxG motif. The PE domain interacts with a PPE domain through conserved apolar residues that establish strong hydrophobic interactions (Strong et al. 2006). A secretion motif in the mostly unstructured C-terminal domain of PE proteins is, similar as for EsxB, essential for secretion (Daleke, Ummels, et al. 2012; Poulsen et al. 2014). While the genes coding for the most ancient PE proteins are located within the \textit{esx} loci and mostly consist of only the PE-domain, more recently evolved members of this protein family have (large) extended C-terminal domains. For instance, \textit{M. tuberculosis} contains 99 genes encoding PE proteins, of which a 69 belong to the PE_PGRS group, named after polymorphic GC-rich-repetitive sequences that code for the C-terminal domains. These C-terminal domains are composed of glycine-rich repeats and can be up to 1550 residues long. These long PE proteins are secreted to the cell surface by a T7S system (i.e. the ESX-5 system) as well.

The PPE domain, with approximately 180 amino acids, is slightly larger than the PE domain and contains five $\alpha$-helices. The three N-terminal $\alpha$-helices interact with the PE partner protein. A typical WxG motif is present in the turn between the second and the third $\alpha$-helix, similar to that of the ESX proteins. The fourth and fifth $\alpha$-helices also pair together to form an extension of the PE-PPE dimer. This extended region forms the interaction site for the EspG chaperone (Daleke, Woude, et al. 2012; Korotkova et al. 2014). Similar to PE proteins, PPE proteins can have (largely) extended C-terminal domains.

As mentioned previously, the interaction of the PPE41 protein with EspG$_5$ is mediated by the conserved hydrophobic tip of the PPE proteins, mostly formed by $\alpha$-4 and $\alpha$-5. Bio-informatic analysis suggests that PPE proteins secreted by
the different ESX systems in mycobacteria can be grouped on the basis of these conserved residues in the hydrophobic patch (17), and this phenomenon could explain how the EspG chaperones establish system specificity (15). Thus far, the PE and PPE proteins are the only T7S substrates are shown to require such a specific chaperone.

Other T7S substrates

The EsxAB clan also contains the DUF2563, DUF2580 and the LxG family. The first one is a small Mycobacterium-specific protein family, of which not much is known. Members of the second family, DUF2580, are restricted to the mycobacteria as well, but some members of this family are indeed T7S substrates: both the ESX-1 substrates EspC and EspF belong to this family. Interestingly, the genes encoding these substrates are located in an operon with genes encoding other T7S substrates, i.e. EspA and EspE, respectively. Although EspA and EspE do not officially belong to the EsxAB clan, structure prediction program (Phyre2 (Kelley and Sternberg 2009)) predicts with high confidence a helix-turn-helix domain at the N-terminus of these proteins with Esx proteins as best template. Therefore, possibly these proteins form heterodimers with EspC and EspF. Another known ESX-1 substrate is EspB. Recently, the structure of this protein was elucidated, which showed that this protein has an N-terminal helix-turn-helix motif followed by a T7 secretion signal motif (see below, Figure 3D) (Daleke, Ummels, et al. 2012). The major surprise was that the adjacent region of this domain in fact showed structural homology to PPE proteins, which means that this protein forms a four-helix bundle by itself and therefore could be secreted as a monomer. The study by Solomonson et al. (Solomonson et al. 2015) furthermore shows that these monomeric structures multimerize as a heptamer with a barrel-shaped structure. This suggests that PE-PPE dimers could form similar quaternary structures, which might have important implications to predict their functions.

Firmicutes have different T7S substrates as well, both EsxAB-like proteins and proteins that do not seem to belong to the EsxAB clan. Recently, in S. aureus two new substrates were identified with the somewhat misleading names EsxC and EsxD. Although they officially do not belong to the EsxAB clan (and therefore also not to the Esx family), again structure prediction indicates a helix-turn-helix motif.

The final and perhaps most intriguing protein family within the EsxAB clan is LxG. This recently described extended protein family contains diverse proteins, including (putative) endonucleases and toxins (Zhang, Iyer, and Aravind 2011). These members have an N-terminal LxG domain and a nuclease domain (SUKH family) at the C-terminus. Members of this family are mainly found in the Firmicutes. Unfortunately, secretion of these proteins has not been studied. However, because they belong to the EsxAB clan, secretion via T7S seems likely. Supporting this prediction is the observation that homologs of SUKH endonucleases in other bacteria have different N-terminal domains that are linked to secretion; in Proteobacteria they are predicted to be secreted via the two-step secretion pathway and in Actinobacteria via a classical (Sec/Tat) secretion pathway.
T7 secretion signal

How are the T7S substrates recognized and what determines system specificity? The first indication of a secretion signal was identified for the EsxAB dimer. Deletion of the unstructured C-terminal tail of EsxB completely blocked secretion (Champion et al. 2006). Furthermore, in the same study it was shown that in yeast two-hybrid experiments the C-terminal tail of EsxB interacts with the EccC protein of ESX-1. This first study could not identify crucial residues within this C-terminal tail, but later it was shown that there was indeed a conserved consensus, albeit slightly more upstream (Daleke, Ummels, et al. 2012). The presence of a C-terminal secretion signal was also shown for other T7S substrates, including EspC and PE25. Detailed analysis showed that two residues within the C-terminal tail of PE25 are crucial and that the spacing between these residues is important as well. The identified signal, YxxxD/E, is crucial for the secretion of Esx, PE and Esp proteins in mycobacteria (Daleke, Ummels, et al. 2012). If we look at all T7 substrates from different organisms a broader consensus must be used, as described by Poulsen et al. (Poulsen et al. 2014). Interestingly, although the secretion signal was first identified in the unstructured C-terminal tail of T7S substrates, more recent structural studies on Esx, PE and EspB proteins showed that this signal is in fact part of the elongated second helix. In these structures the two crucial residues, i.e. tyrosine and the acidic amino acid, are located on the same side of the helix (Poulsen et al. 2014). Furthermore, the tyrosine is closely positioned to the conserved tryptophan residue of the WxG motif in the turn region of the dimer partner. It has been suggested that this tryptophan residue is therefore also part of the secretion signal (Poulsen et al. 2014). Because the deletion of several C-terminal residues beyond the YxxxD/E motif is already enough to block secretion, the signal probably extends at least 10 amino acids further than the consensus sequence, however without any clear conserved features. Recent data (Rosenberg et al. 2015) indicate that this unstructured tail is recognized by the C-terminal domain of EccC and responsible for the multimerization of EccC. Probably a similar secretion signal is present for T7S substrates in Firmicutes, but the role in secretion has not been studied in the same detail. Although indeed the C-terminus was shown to be important for these Firmicute substrates, the wrong amino acids were examined (i.e. an erroneous YxxxD/E motif), without any effect (Anderson et al. 2013; Poulsen et al. 2014).

One puzzling observation is that exchange of the secretion signal between an ESX-1 and an ESX-5 substrate does restore secretion, but does not redirect the substrate to another secretion system. Therefore, this secretion signal does not seem to determine system specificity. Apparently, a second signal is required for this. For the PE-PPE heterodimeric complex this second signal is probably provided for by the specific chaperone EspG, which was shown to specifically recognize substrates of the cognate secretion system. However, it is still unknown what characteristic of the Esx dimers is determining their systems specificity.
Other major questions concerning substrates and substrate specificity also still remain. For instance, the large majority of PE and PPE proteins in mycobacteria seem to be without an obvious partner protein. Are they secreted as single proteins, or are they also secreted as a heterodimer with an unknown partner? Furthermore, many of these PE and PPE proteins contain large C-terminal extensions. A good example is one of the few PE/PPE protein with a known function, LipY. This protein contains a PE domain in *M. tuberculosis*, a PPE domain in *M. marinum* and a classical signal peptide in the fast-growing *Mycobacterium gilvum* (Daleke et al. 2011). Interestingly, whereas both the PE and PPE domain were shown to be interchangeable for secretion to the cell surface via T7S, exchange for a classical signal sequence did not result in surface localization. What is interesting to note is that many of the WxG (and LxG) proteins in Actinobacteria and Firmicutes have C-terminal protein domains with described functions as well, indicating that there might be overlapping structural and mechanistic similarities between these proteins and the PE/PPE proteins. Unfortunately, no protein structures are known for these putative substrates and therefore many questions about the folding, chaperones, secretion partners and functions of these proteins remain.

The Role Of T7S Systems In Virulence
ESX-1 system of pathogenic Mycobacteria

The specific roles of T7S systems and their substrates have been best-described in mycobacteria. As mentioned above, the ESX-1 system of *M. tuberculosis* was the first T7S system identified; this region was lost when *M. bovis* was cultured for 11 years by Calmette and Guerin to create the live attenuated-vaccine strain *M. bovis* BCG (Berthet et al. 1998; Philipp et al. 1996). It was named region of difference 1 (RD1) and Pym et al. showed that this region is (mainly) responsible for the attenuation of *M. bovis* BCG and the vole bacillus *Mycobacterium microti* (Pym et al. 2002).

Since then, the importance of the ESX-1 system for the virulence of mycobacteria has been strongly established (Hsu et al. 2003; Lewis et al. 2003; Brodin et al. 2006; Fortune et al. 2005), but the role of the individual secreted substrates has been more difficult to investigate. As mentioned above, the first ESX-1 substrates that were identified were EsxA and EsxB (Andersen et al. 1995), which are major secreted proteins and dominant T cell antigens. As such they were regarded as the principle substrates and responsible for the attenuation of *M. bovis* BCG. The EsxA protein indeed seemed to be the prime candidate for mediating phagosomal escape (see below). Later it was shown that more proteins are secreted via this system; these proteins are known as ESX-1-associated proteins (Esp). Especially EspA and EspB seem to be involved in virulence as well (Pang et al. 2013; Hunt et al. 2012; Garces et al. 2010; Solomonson et al. 2015). A complicating factor in determining the role of individual components in virulence is their co-dependence for secretion. Secretion of EspA and possibly EspB is shown to be linked to functional secretion of EsxAB and
vice versa (Fortune et al. 2005). Redirection of ESX-1 substrates through other T7S systems could help to solve this conundrum.

Virulence

Although the interdependence of ESX-1 substrates makes the study of individual substrates difficult, major steps have been taken to elucidate the function of ESX-1 and its substrates in virulence. The ESX-1 system plays an important role in the macrophage infection cycle of pathogenic mycobacteria. Several species of Mycobacteria, including *M. tuberculosis*, *M. marinum* and *M. leprae* have been shown to translocate from the phagosome to the cytosol in late stages of the macrophage infection cycle (van der Wel et al. 2007). Upon translocation to the cytosol, bacteria start to replicate and ultimately induce a necrosis-like cell death. After necrosis of the host cell the bacteria can spread to neighboring macrophages within the host (Simeone et al. 2012). Therefore, escape into the cytosol is essential for full virulence. This crucial step is dependent on a functional ESX-1 system (van der Wel et al. 2007), since phagosomal escape is not observed for *M. bovis* BCG or esx-1 mutants of *M. tuberculosis* and *M. marinum* (D. Houben et al. 2012; van der Wel et al. 2007). Furthermore, it has also been shown that mycobacteria can lyse erythrocytes in an ESX-1- and contact-dependent manner (Gao et al. 2004; Koo et al. 2008). This phenomenon is seen as a direct effect of the membrane disrupting potential of ESX-1 substrates (Smith et al. 2008; Koo et al. 2008). Several studies have implicated EsxA as the main membrane-disrupting protein of the ESX-1 system (de Jonge et al. 2007; Smith et al. 2008), although some questions remain. Most importantly, EsxA and EsxB are present in non-pathogenic mycobacteria as well as pathogenic mycobacteria and seem to have similar functions (Flint et al. 2004). Furthermore, *Mycobacterium kansasii* has several subtypes of which only one shows phagosomal escape and a concomitant increased virulence. Avirulent subtypes of *M. kansasii* do efficiently secrete EsxA (D. Houben et al. 2012). Perhaps different ESX-1 dependent substrates are together required for phagosome escape. EspA would be an important candidate for this, since this protein is highly upregulated within the phagosomal environment. Another important substrate is EspB, which was shown to be crucial for the virulence of *M. marinum*. Interestingly, this protein forms ring-shaped heptamers with a hydrophobic domain (Solomonson et al. 2015). As such, this protein could be involved in perturbation of host cell membranes. In *M. marinum*, ESX-1 proteins have been shown to localize in the capsular layer at the cell surface (Sani et al. 2010). This surface localization of ESX-1 substrates would be in line with the membrane-disrupting capacities of mycobacteria (Kennedy et al. 2014). Whether this phenomenon is the same for *M. tuberculosis* remains to be established. Furthermore, the components of the ESX-1 machinery have been shown to be localized to the cell poles of multiple species of mycobacteria (Wirth et al. 2012; Carlsson et al. 2009). The ESX-1 system and its substrates seems to be enriched in new cell poles with active peptidoglycan
synthesis, indicating that there might be a role for the ESX-1 system in cell-wall growth (Carlsson et al. 2009).

**ESX-1 regulation**

The ESX-1 system is already active in culture medium, but is upregulated when mycobacteria encounter host cells. This process is regulated by multiple transcriptional regulators. The best-described transcriptional regulator is the two-component sensor kinase PhoPR system. The PhoPR system is activated by a decrease in pH (Abramovitch et al. 2011), as well as increased Cl⁻ concentration (Tan et al. 2013), which are conditions occurring in maturing phagosomes. The avirulent *M. tuberculosis* strain H37Rv contains a single point mutation in the DNA binding-domain of PhoP, which abrogates DNA binding and results in a reduced virulence (J. S. Lee et al. 2008). PhoP was later shown to regulate secretion of ESX-1 indirectly through regulation of the espACD locus (Frigui et al. 2008).

Another gene that is part of the PhoP regulon is *whiB6* (*Rv3862c*), which is a transcriptional regulator situated in close genetic proximity to the *esx-1* gene cluster (Solans et al. 2014). In clinical *M. tuberculosis* strains, PhoP binds to the promoter region of *whiB6*, which will induce transcription. Increased production of WhiB6 positively regulates the expression of several *esx-1* genes and in that way is able to increase ESX-1 secretion. Point mutations in the promoter region of *whiB6* in laboratory strains H37Rv and H37Ra result in an inverted regulation of *whiB6* by PhoP, making PhoP a negative regulator of ESX-1 secretion (Solans et al. 2014).

The espACD operon is also regulated by other factors, one of which is EspR. This protein was discovered as a regulator of ESX-1 secretion and was first hypothesized to be an ESX-1 substrate itself (Raghavan et al. 2008), but was later confirmed to be a more general nucleoid-associated protein that is not secreted (Blasco et al. 2012). EspR is a transcriptional regulator with a unique helix-turn-helix structure and a C-terminal domain that is essential for dimerization. In its dimerized form it binds two specific operator sites situated 177 bp apart, thereby probably creating a loop in the promoter region of espACD (Rosenberg et al. 2011).

Finally, the espACD operon is also regulated by the two-component regulator MprAB (Pang et al. 2013) and the regulator Lsr2 (Gordon et al. 2010). Altogether a picture is emerging of a secretion system that is tightly regulated by multiple layers of regulation, adapted to different conditions. These networks can be highly relevant for the virulence capacities of clinical strains, since small mutations in regulators or promoter regions can have significant effects on virulence (Gonzalo-Asensio et al. 2014; J. S. Lee et al. 2008).

**Role of ESX-1 and T7S in horizontal gene transfer**

Although the role of the ESX-1 system in pathogenic mycobacteria is tightly linked with intracellular routing and host cell death, a completely different role has been described in *M. smegmatis*. This bacterium is capable of horizontal gene transfer.
(HGT), in which large genomic fragments are exchanged between different strains (reviewed in (Derbyshire and Gray 2014)). This process is an atypical form of conjugation and is called distributive conjugal transfer (DCT). Interestingly, this process is dependent on ESX-1 in different and opposing ways: esx-1 mutants of donor strains are shown to be hyperconjugative (Flint et al. 2004), whereas esx-1 mutants in the recipient strains are hypoconjugative (Coros et al. 2008). This paradoxical dual role of ESX-1 in mycobacteria has not been mechanistically explained yet, but it will be highly interesting to discover if DNA is transported through this T7S system or if secreted proteins located at the cell surface play a role in cell-cell adhesion. An interesting observation is that the hyperconjugative phenotype of the M. smegmatis esx-1 mutants can be complemented with its M. tuberculosis counterparts. However, DCT has not been described in M. tuberculosis or other slow-growing mycobacteria (Gray et al. 2013; Gutierrez et al. 2005).

The ESX-1 system of M. smegmatis is not the only ESX system that has been linked to DNA transport. Recently a conjugative plasmid (pRAW) was discovered in M. marinum (Ummels et al. 2014). This plasmid efficiently conjugated between different slow-growing species of Mycobacteria, including M. tuberculosis. No conjugation to fast-growing mycobacteria was observed, indicating specific genetic requirements for recipient strains. The pRAW plasmid was shown to contain a newly identified T7S system called ESX-P1. This ESX system was shown to be essential for conjugal transfer of pRAW. ESX-P1 has the highest homology with the ESX-5 system (Ummels et al. 2014).

ESX-3
The ESX-3 system is essential for the growth of M. tuberculosis, as was shown by both directed mutagenesis and saturated transposon mutagenesis (Griffin et al. 2011; Sassetti, Boyd, and Rubin 2003). ESX-3 is regulated by iron- and zinc-dependent transcription regulators ideR (Rodriguez et al. 2002) and Zur (FurB) (Maciag et al. 2007). Sloan-Sigrist et al. have shown that expression of the esx-3 system is highly upregulated when iron chelators are added to the growth medium (Siegrist et al. 2009). Additionally, several groups have reported that ESX-3 is only essential for growth under iron-limiting conditions (Siegrist et al. 2009; Serafini et al. 2009). It is postulated that a functional ESX-3 system is crucial for the uptake of iron via mycobactin, a siderophore produced by mycobacteria, but a mechanism for this phenomenon remains to be elucidated. In conditions of high zinc availability, the essentiality of ESX-3 can be rescued as well. The most efficient complementation of ESX-3 essentiality is achieved by supplementing both iron and zinc, which suggests that mycobactin uptake is not the only function of the ESX-3 system (Serafini et al. 2009). There seems to be only a partial correlation of functional ESX-3 secretion and mycobactin uptake, indicating that the role of ESX-3 is probably broader than metal ion acquisition. It is somewhat counterintuitive that a protein secretion system should be involved in the uptake of metal ions, but perhaps ESX-3 substrates
play a role in the uncoupling of the Fe-mycobactin complex or in the actual binding of this siderophore at the cell surface.

The only known substrates of the ESX-3 system are EsxG and EsxH, which are both encoded by the esx3 locus itself. These two proteins are secreted as a heterodimer, just like their ESX-1 homologs EsxAB. Two other putative ESX-3 substrates are PE5 and PPE4, which are encoded by the esx-3 cluster and which are also essential. However, experimental evidence for the secretion of these proteins via ESX-3 is lacking.

ESX-5

The ESX-5 system is probably the most-recently evolved T7S system in mycobacteria and is only present in slow-growing species (Gey van Pittius et al. 2006). The slow-growing mycobacteria include most pathogenic species, such as *M. tuberculosis*, *M. leprae*, *M. marinum*, *M. ulcerans* and *M. avium*. The evolution of the ESX-5 system, probably through a duplication of the esx-2 gene cluster, was followed by a large expansion of the *pe* and *ppe* genes in species like *M. tuberculosis* (Cole et al. 1998) and *M. marinum*. Genomic analysis suggests that ESX-5 could be responsible for the secretion of most PE and PPE proteins and that the ESX-5 system plays an important role in virulence and/or the slow-growing phenotype of mycobacteria. Different biochemical methods have experimentally validated the first hypothesis; indeed many PE and PPE proteins of *M. marinum* and *M. tuberculosis* are secreted via ESX-5, especially the recently evolved subclasses of PE and PPE proteins, such as the PE-PGRS and PPE-MPTR proteins (Abdallah et al. 2009). After secretion these proteins can be identified mainly at the cell-surface but also in the culture filtrate (Abdallah et al. 2009; Bottai et al. 2012; Mishra et al. 2008). There are indications that a number of these surface-localized PE-PPE proteins have a direct role in virulence by interacting with host immune receptors, such as Toll-like receptor (TLR) 2 and the inflammasome (reviewed in (Sampson 2011)), but recent data indicate that there could be an important second role for ESX-5 substrates. Just like for ESX-3, the membrane components of the ESX-5 system are essential for the *in vitro* growth of *M. bovis* BCG, *M. marinum* (Ates et al. 2015) and *M. tuberculosis* H37Rv on culture plate or broth (Di Luca et al. 2012; Griffin et al. 2011). However this essentiality is not observed in all *M. tuberculosis* strains, for instance in CDC1551 esx-5 mutations can be readily identified (Bottai et al. 2012; E. N. G. Houben et al. 2012). Recent data could explain this apparent discrepancy; the essentiality of esx-5 is linked to the permeability of the outer membrane. The ESX-5 system is no longer essential when outer membrane permeability is increased, either by mutations in lipid biosynthesis or by the introduction of the pore-forming protein MspA from *M. smegmatis*. Slow-growing mycobacteria do not produce MspA-like pores and as such probably need alternative mechanisms for nutrient transport over the outer membrane. The ESX-5 system and its substrates are responsible for the uptake of fatty acids and probably other nutrients (Ates et al. 2015). The expression of MspA-like pores makes mycobacteria more vulnerable to defense systems of the host
(Lamrabet et al. 2014), which suggests that the slow-growing mycobacteria utilize ESX-5 substrates to acquire nutrients, while maintaining resistance against bactericidal host factors.

In *M. marinum*, esx-5 mutants with a strongly reduced secretion phenotype (recall that knock-out mutants are not viable) are attenuated in cell-infection experiments and show delayed phagosomal escape. The ESX-5 system also seems to be involved in inducing cell death through inflammasome activation via an unknown mechanism (Abdallah et al. 2011). An espG5 transposon mutant of *M. marinum* is attenuated in zebrafish embryo experiments, but surprisingly this mutant is hypervirulent in adult zebrafish (Weerdenburg et al. 2012). This hypervirulence is characterized by increased bacterial numbers in the organs, an altered pro-inflammatory cytokine response and more severe immunopathology (Weerdenburg et al. 2012). This suggests that the ESX-5 system might be responsible for manipulation of the adaptive immune response, which is not yet developed in zebrafish embryos. However, the same effect was observed in *rag/-* zebrafish that lack B-cells and T-cells, indicating that antigen-specific adaptive immune responses are probably not responsible for the hypervirulence of the esx-5 mutant. Additionally, expression of dormancy-related genes was not clearly affected in the esx-5 mutant strain, indicating that hypervirulence is not due to an inability to enter the dormant phase of infection (Weerdenburg et al. 2012). In contrast to these results, Bottai et al. have shown that an eccD5 mutant of *M. tuberculosis* has reduced cell wall integrity and is significantly attenuated in a SCID mouse model (Bottai et al. 2012).

One of the few ESX-5 substrates with a well-studied and defined function is the LipY protein. This protein is a PE protein in *M. tuberculosis*, but a PPE protein in *M. marinum*, illustrating that the PE/PPE domains are probably not involved in the function of these proteins, but responsible for the secretion (Daleke et al. 2011; Mishra et al. 2008). LipY is a lipase, which can efficiently degrade long chain triacylglycerols (TAGs) that are present in the host cell or stored inside the mycobacteria (Daniel et al. 2011; Deb et al. 2006). Utilization of TAGs is important during latent infection and expression of LipY is upregulated in *in vitro* models of dormancy (Deb et al. 2006). *M. bovis* BCG overexpressing lipY was shown to lose its capacity to induce protection against *M. tuberculosis* (Singh et al. 2011) and an *M. tuberculosis* strain which overexpressed LipY was hypervirulent in a mouse model (Singh et al. 2014). These data indicate that LipY is an important virulence factor in *M. tuberculosis*.

**PE_PGRS**

As mentioned previously, the major group of PE proteins are the so-called PE_PGRS proteins, characterized by a long C-terminal region with glycine-rich repeats. Notably, PE_PGRS proteins are not present in all ESX-5-containing mycobacteria. *M. avium* and *M. xenopi* for instance do not contain any PE_PGRS protein (Gey van Pittius et al. 2006). PE_PGRS proteins have been hypothesized to be involved in several pathogenesis-
related or immunological mechanisms. The PE_PGRS wag22 was proposed to contain a C-terminal fragment with fibronectin binding properties (Espitia et al. 1999). In a subsequent study the wag22 gene was shown to be a vaccine candidate that prevents reactivation of \textit{M. tuberculosis} infection in a mouse model (Campuzano et al. 2007). PE-PGRS proteins have also been proposed to protect against proteosomal degradation and therefore reduced activation of CD8+ T-cells (Koh, Lehming, and Seah 2009; M. J. Brennan and Delogu 2002). The glycine-alanine rich repeat regions of PE_PGRS proteins are similar in composition to the EBNA1 protein of the human Eppstein-Barr virus (EBV). The EBNA1 protein was shown to be a virulence factor of EBV that is able to block its own proteasomal processing (Levitskaya et al. 1995) and the same phenomenon was shown for PE_PGRS proteins, suggesting a role in immune evasion for these proteins (M. J. Brennan and Delogu 2002; Koh, Lehming, and Seah 2009). However, in recent publications it has been shown that repetitive protein sequences of EBNA1 are not involved in preventing antigenic presentation. Instead, expression of these proteins seems to be self-limited by the mRNA sequence coding for these proteins, which is very rich in purines. These unstructured, purine-rich mRNA sequences make translation efficiencies high enough to create sufficient protein levels for infection, but low enough to avoid immune recognition (J. Tellam et al. 2008; J. T. Tellam et al. 2012). These results, question the function of PE_PGRS proteins in blocking epitope processing.

The PE_PGRS proteins have also been hypothesized to be a source of antigenic variation in mycobacteria (Banu et al. 2002). However, a conserved hierarchy of immune recognition of different PE/PPE proteins suggests that differential immune responses are not driven by antigenic variation, but by infection stage-specific expression of these proteins (Vordermeier et al. 2012). Additionally, sequencing of multiple clinical strains and their pe_pgrs sequences has led to the observation that pe_pgrs diversity as a whole is not driven by antigenic pressure (Copin et al. 2014; McEvoy et al. 2012). In fact, bio-informatic analysis indicates that the PE_PGRS proteins contain very few epitopes. Of the 1,649 known epitopes of \textit{M. tuberculosis}, only 3 are situated in the PGRS domain of PE_PGRS proteins (Copin et al. 2014). Nonetheless, there seems to be antigenic pressure on some of the pe_pgrs genes, while others seem to evolve neutrally. This suggests that the group of PE_PGRS proteins may not be as homogenous as sometimes thought and that individual genes or subgroups of these genes and proteins of these groups need to be examined for their biological roles.

One of the PE_PGRS proteins that has been studied in more detail is PE_PGRS33 (Rv1818c). Polymorphisms in pe_pgrs33, including large insertions, deletions or truncations, correlate with an absence of lung cavities in patients (Talarico et al. 2007). As shown for other PE proteins, the PE domain of PE_PGRS33 is necessary for the surface localization of this protein (Cascoferro et al. 2011; Delogu et al. 2004). The PGRS domain of PE_PGRS33 is shown to localize to the mitochondria of the host cell and able to induce apoptosis (Cadieux et al. 2011). Surprisingly, the full length PE_PGRS33 was also able to induce necrosis, showing that the PE and linker
domains can perform additional functions in this mechanism. Another study showed that PE_PGRS33 was able to induce host-cell apoptosis by specifically binding to the TLR2 (Basu et al. 2007).

Together, these data show that the PE_PGRS proteins are important manipulators of the host immune response. However, the role of the PE_PGRS proteins remains to be fully elucidated. Understanding the role of these proteins will be a major step in increasing our knowledge of M. tuberculosis, and will most likely contribute to our understanding of disease pathogenesis as well as vaccine design.

Type VII(-like) Secretion In Monoderm Bacterial Species

T7S systems are widely present in mycobacteria and closely-related bacteria. Mycobacteria are characterized by a diderm cell envelope of which the outer membrane is characterized by the presence of unusual lipids known as mycolic acids. These mycolic acids are partially linked to the arabinogalactan layer and partially coupled to trehalose molecules. Mycolic acids are not unique for mycobacteria, they can also be found in several close relatives, including Corynebacteria and Rhodococcus species. These bacteria therefore have a diderm envelope similar to mycobacteria and usually have a locus potentially coding for a T7S system (E. N. G. Houben, Korotkov, and Bitter 2013). Unfortunately, T7S systems have not been analyzed in these bacteria.

Interestingly, a number of monoderm bacteria also have T7S(-like) systems and some of these have been studied in more detail. The first one that we would like to discuss is the T7S system present in streptomycetes. Although these bacteria are monoderm and do not produce mycolic acids, they do belong to the same taxonomic order (Actinomycetales) as mycobacteria and have a well-conserved T7S system; in Streptomyces species all genes coding for the 5 conserved membrane proteins, i.e. homologs of EccBCDE and MycP, are usually present (E. N. G. Houben, Korotkov, and Bitter 2013; Akpe San Roman et al. 2010). A number of substrates seem to be present in these loci, of which the two small Esx proteins have indeed been shown to be secreted by Streptomyces coelicolor (Akpe San Roman et al. 2010). In two different species of Streptomyces, i.e. S. coelicolor and S. scabies (Akpe San Roman et al. 2010; Fyans et al. 2013), deletion of genes coding for the Esx-like T7S substrates resulted in abnormal spore formation and/or altered timing of spore formation. Surprisingly, absence of the T7S components did not show the same effect as the substrates, which seems to suggest that these Esx proteins have an intracellular effect on spore formation.

A number of Firmicutes species have a T7S-like system that contains only 2 of the original genes, i.e. an eccC homolog and one or more of the esx genes. In addition, there are usually a variable number of other (membrane) proteins required for secretion (discussed below). In contrast to the T7S systems in mycolic-acid containing bacteria, the T7S-like systems in Firmicutes are probably acquired relatively recently, as their composition and distribution is highly variable. For instance, Streptococcus agalactiae is the only species within the Streptococci family that has an intact T7S-like
system and the T7S-like system of Bacillus cereus is in composition and homology more similar to that of Listeria species than those of other Bacillus species. Most data on T7S-like systems in Firmicutes has been obtained for B. subtilis and S. aureus and therefore these systems will be discussed in more detail.

Although protein secretion in B. subtilis has been studied already for decades only recently a T7S-like system (Figure 1) was shown to be functional, which seems to be due to defective regulation in domesticated Bacillus strains (Baptista, Barreto, and São-José 2013). In undomesticated strains secretion could be readily observed in late log phase. The secreted substrate is an Esx-like protein known as YukE (Baptista, Barreto, and São-José 2013). Proteomic analysis showed that, under normal laboratory conditions, this is the only T7S-like substrate (Huppert et al. 2014). YukE is secreted as a dimer and, in analogy to the mycobacterial T7S substrates, the C-terminal tail is required for secretion. Unfortunately, the extracellular function of YukE is currently unknown. Detailed mutational analysis has shown that, in addition to the EccC-homolog (called YukB or YukBA), also YukCD and YueBC are required (Huppert et al. 2014; Baptista, Barreto, and São-José 2013). Apart from YukD, these proteins are all membrane proteins. Of these membrane proteins, YueB is the one with most potential TMDs (6). Because this protein probably forms a dimer (São-José, Baptista, and Santos 2004) this complex could be somewhat similar to EccD of mycobacteria. YueB-like proteins therefore could form a central component in T7S-like systems. In contrast to EccD, the YueB protein has a very substantial extracellular domain of more than 800 amino acids. This extracellular domain is a receptor for the bacteriophage SPP1 (São-José, Baptista, and Santos 2004). The only cytosolic protein that is essential for secretion, YukD, has a ubiquitin-like fold (van den Ent and Löwe 2005). Despite this structural similarity, YukD does not seem to be conjugated, neither to itself nor to other proteins (van den Ent and Löwe 2005). For the YukD homolog of S. aureus (SAUSA300_0281, also described as EsaB (Burts et al. 2005) (Figure 1)) it has been shown that this protein is involved in intracellular post-transcriptional regulation of a specific substrate of the T7S-like system (Burts, DeDent, and Missiakas 2008). Therefore, this ubiquitin-like protein perhaps functions as a specific chaperone or regulator.

The T7S-like system of S. aureus is called Ess and contains the same elements as the T7S-like system in B. subtilis, only the gene order is different (Figure 1). The only difference is that in S. aureus an additional gene is identified that might be required for secretion, i.e. esaD (Anderson et al. 2011). However, the effect of an esaD deletion on secretion is not absolute; a major reduction of protein secretion is observed when this gene is deleted, but not a complete blockade. Moreover, in other species esaD homologs are not always linked to this secretion system and the C-terminal domain of EsaD is predicted to form a nuclease/hydrolase. Therefore, more research is needed to substantiate this component. Another feature of T7S-like systems that is conserved between B. subtilis and S. aureus is the variability in secretion between strains. Both the amount and timing of secretion is highly strain dependent for S. aureus.
(Kneuper et al. 2014), which indicates that this system is controlled by a complex and variable regulation system. Some of these regulators have been identified; especially the production of the main substrate EsxA seems to be controlled by a number of different regulators (Schulthess, Bloes, and Berger-Bächi 2012; Anderson et al. 2013).

The most important difference between the T7S-like systems in S. aureus and B. subtilis is the number of substrates that are secreted. Whereas B. subtilis secretes only a single substrate, four different substrates have been identified thus far for S. aureus. Two Esx proteins are secreted and two proteins that do not officially belong to the EsxAB clan (Anderson et al. 2013), but are predicted to form a N-terminal helix-turn-helix motif. These new substrates are unfortunately called EsxC and EsxD. All these substrates are encoded by genes within the T7S-like locus.

Ess secretion is important for both colonization and persistence of S. aureus (Burts et al. 2005; Kneuper et al. 2014). Disruption of this T7S-like system or deletion of the substrates results in significantly reduced bacterial numbers and reduced pathology in a mouse model. Interestingly, a recent report indicates that this secretion system is involved in the intracellular survival of S. aureus in epithelial cells by blocking apoptosis and promoting escape (Korea et al. 2014). If substantiated, this would then be highly similar to the function of the ESX-1 system in pathogenic mycobacteria. In line with this function, close homologs of the S. aureus Ess system are present in other pathogens, such as Listeria monocytogenes and B. cereus.

CONCLUDING REMARKS

In approximately one decade of research on T7S, major advances have been made to elucidate the mechanism of T7S, and the role of T7S in pathogenesis. However, many questions remain on the functioning and structure of T7S systems. Is the T7S membrane channel in mycolic-acid containing bacteria a double membrane spanning complex, mediating transport in a one-step mechanism, or are there unidentified outer membrane components of T7S systems, responsible for transport over the specific outer membrane? Solving the structure of mycobacterial T7S membrane complex will answer many of these questions and would be a major breakthrough in this field. Another topic that needs to be addressed is the function of individual T7S substrates. Unfortunately, interdependence of substrates for secretion and complex regulatory and post-translational networks complicate research on this topic. The observed roles of T7S systems and their substrates in virulence and in nutrient acquisition show that these systems are promising targets for future drug development. Recently, the first T7S inhibitors have been identified and could form the front runners of new classes of drugs against pathogenic mycobacteria (Rybniker et al. 2014). A more detailed understanding of T7S systems and their substrates could lead to new concepts for treatment of mycobacterial diseases, but might also shed light on novel potential virulence factors secreted by other Actinobacteria and Firmicutes.
ESSENTIAL ROLE OF THE ESX-5 SECRETION SYSTEM IN OUTER MEMBRANE PERMEABILITY OF PATHOGENIC MYCOBACTERIA

Louis S. Ates¹*, Roy Ummels¹*, Susanna Commandeur¹, Robert van de Weerd¹, Marion Sparrius¹, Eveline Weerdenburg¹, Marina Alber², Rainer Kalscheuer², Sander R. Piersma³, Abdallah M. Abdallah⁴, Moataz Abd El Ghany⁴, Alyaa M. Abdel-Haleem⁴, Arnab Pain⁴, Connie R. Jiménez³, Wilbert Bitter¹,⁵ and Edith N.G. Houben⁵*

ABSTRACT

Mycobacteria possess different type VII secretion (T7S) systems to secrete proteins across their unusual cell envelope. One of these systems, ESX-5, is only present in slow-growing mycobacteria and responsible for the secretion of multiple substrates. However, the role of ESX-5 substrates in growth and/or virulence is largely unknown. In this study, we show that esx-5 is essential for growth of both *Mycobacterium marinum* and *Mycobacterium bovis*. Remarkably, this essentiality can be rescued by increasing the permeability of the outer membrane, either by altering its lipid composition or by the introduction of the heterologous porin MspA. Mutagenesis of the first nucleotide-binding domain of the membrane ATPase EccC5 prevented both ESX-5-dependent secretion and bacterial growth, but did not affect ESX-5 complex assembly. This suggests that the rescuing effect is not due to pores formed by the ESX-5 membrane complex, but caused by ESX-5 activity. Subsequent proteomic analysis to identify crucial ESX-5 substrates confirmed that all detectable PE and PPE proteins in the cell surface and cell envelope fractions were routed through ESX-5. Additionally, saturated transposon-directed insertion-site sequencing (TraDIS) was applied to both wild-type *M. marinum* cells and cells expressing *mspA* to identify genes that are not essential anymore in the presence of MspA. This analysis confirmed the importance of esx-5, but we could not identify essential ESX-5 substrates, indicating that multiple of these substrates are together responsible for the essentiality. Finally, examination of phenotypes on defined carbon sources revealed that an esx-5 mutant is strongly impaired in the uptake and utilization of hydrophobic carbon sources. Based on these data, we propose a model in which the ESX-5 system is responsible for the transport of cell envelope proteins that are required for nutrient uptake. These proteins might in this way compensate for the lack of MspA-like porins in slow-growing mycobacteria.

1Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, the Netherlands
2Institute for Medical Microbiology and Hospital Hygiene, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany
3Department of Medical Oncology, OncoProteomics Laboratory, VU University Medical Center, Amsterdam, the Netherlands
4Biological and Environmental Sciences and Engineering (BESE) division, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia
5Section Molecular Microbiology, Amsterdam Institute of Molecules, Medicine & Systems, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
* shared first authors
Author summary

Mycobacteria have a thick protective outer membrane that helps them to withstand adverse conditions both outside and within the host. However, in order to cause disease, the bacterium also needs to secrete proteins across this outer membrane. To achieve this, mycobacteria possess so-called type VII secretion systems. One of these systems, the ESX-5 secretion system, is only present in the group of slow-growing mycobacteria, which contains most pathogenic species. In this study, we show that the ESX-5 system is essential for growth of mycobacteria. We found that when we generated a ‘leaky’ outer membrane, by interfering in the construction of the outer membrane, or by introducing an outer membrane porin, the ESX-5 system was no longer essential for growth. We additionally show that ESX-5 mediates uptake of fatty acids, which suggests that ESX-5 substrates can form specific transport systems or pores in the outer membrane required for the uptake of crucial nutrients. Understanding the role of ESX-5 in outer membrane permeability helps us to understand a fundamental difference between fast-growing and slow-growing mycobacteria. Since most pathogenic mycobacteria are slow-growing this helps us to understand the mycobacterial requirements for pathogenesis in more detail.

INTRODUCTION

Mycobacterium tuberculosis is one of the most important bacterial pathogens; this pathogen has infected almost thirty percent of the world population and is responsible for 1.4 million deaths annually (World Health Organization 2013). A key characteristic that makes M. tuberculosis such a successful pathogen is the composition of its cell envelope. Mycobacteria and other families of the Actinobacteria belonging to the suborder Corynebacteriales (Goodfellow & Jones 2012) have a specialized outer membrane consisting of long chain (C_{50}-C_{90}) α-alkyl β-hydroxy fatty acids, known as mycolic acids. These mycolic acids are covalently linked to the arabinogalactan, which is in turn connected to the peptidoglycan matrix that is located in a periplasmic-like space. Mycolic acids, together with a wide range of other (glyco)lipids, such as lipooligosaccharides (LOSs) (Alibaud et al. 2013; P J Brennan and Nikaido 1995), phthiocerol dimycocerosates (PDIMs) (Camacho et al. 1999; Yu et al. 2012) and trehalose mycolates (Hunter et al. 2009), form a hydrophobic barrier that is organized as an outer membrane which is microscopically similar to that of Gram-negative bacteria (Hoffmann et al. 2008; Zuber et al. 2008). This mycobacterial outer membrane functions as a highly efficient permeability barrier and plays a major role in the persistent nature of mycobacterial infections. It allows the bacteria to survive inside host phagosomes due to an increased resistance to bactericidal host factors such as oxidative radicals and antimicrobial peptides (Purdy, Niederweis, and Russell 2009) and is also one of the main reasons for the antibiotic tolerance of mycobacteria (Gao et al. 2003).
In order to secrete virulence factors and other proteins over their specific cell envelope, mycobacteria have evolved the specialized type VII secretion (T7S) or ESX systems. T7S systems are found throughout the phylum of Actinobacteria and more distantly related systems are also present in Firmicutes. Thus far most information on T7S functioning and role in virulence has come from studying diverse mycobacterial species such as *M. tuberculosis*, the vaccine strain *Mycobacterium bovis* BCG, the closely related fish pathogen *Mycobacterium marinum* and the evolutionary more distant and avirulent *Mycobacterium smegmatis* (Stoop, Bitter, and van der Sar 2012; Simeone, Bottai, and Brosch 2009; E. N. G. Houben, Korotkov, and Bitter 2013). Pathogenic mycobacteria contain up to five different T7S systems named ESX-1 to ESX-5 (Bitter et al. 2009), which have probably evolved through duplication events (Gey van Pittius et al. 2006). These esx-loci are composed of several conserved genes, of which five encode for membrane components. Four of these membrane proteins, called EccB,C,D and E together form a large membrane complex, needed for protein transport (E. N. G. Houben et al. 2012). One of these proteins, EccC, is a putative FtsK/SpoIII-like ATPase with three nucleotide binding domains (NBDs) and is hypothesized to play a central role in substrate recognition (Champion et al. 2006). The fifth conserved membrane component of ESX systems is MycP, which is a subtilisin-like protease that is essential for secretion, although it is not part of the membrane complex.

ESX-1 was the first ESX system that was discovered (Pym et al. 2003) and is involved in the secretion of the important virulence factors EsxA (ESAT-6) and EsxB (CFP-10) as well as several other substrates (Ohol et al. 2010). Virulence of ESX-1 mutants is severely attenuated in macrophage cell lines and *in vivo* infection models, partially because they seem to be unable to escape the phagolysosome of macrophages (D. Houben et al. 2012; Simeone et al. 2012). Deletion of a large part of the ESX-1 genetic locus is also the major cause of the attenuation of the vaccine strain *M. bovis* BCG (Pym et al. 2003). The ESX-3 system seems to have a very different function, as it is involved in iron and zinc uptake (Siegrist et al. 2009; Serafini et al. 2009) and is therefore essential for growth of *M. tuberculosis*. ESX-5 is an intriguing system because it is only present in the slow-growing species of mycobacteria, which include most pathogenic species. This system is responsible for the secretion of many members of the large PE and PPE protein families in *M. marinum* (Abdallah et al. 2009). A major group of these ESX-5 substrates are the glycine-rich and repetitive PE_PGRS proteins, which have been postulated to be involved in virulence (Iantomasi et al. 2012; Chaturvedi et al. 2010) and immune evasion (Abdallah et al. 2008). Nevertheless, the precise role of ESX-5 and its substrates has not been elucidated yet.

To understand the role of ESX-5 in virulence and the mechanism of secretion by this system, we aimed to select a wide range of ESX-5 mutants in *M. marinum*. However, in previous screening assays we only identified transposon insertions in genes encoding the cytosolic chaperone EspG and the cytosolic EccA5 (van der Woude et al. 2012; Abdallah et al. 2009; Abdallah et al. 2006), but not in any of
the membrane components that make up the actual membrane transport machinery (E. N. G. Houben et al. 2012). The inability to find mutations in these genes suggests that these mutations are in fact lethal for the cell. Indeed, Di Luca and colleagues recently showed that one of the genes encoding an ESX-5 membrane component is essential in a strain of M. tuberculosis (H37Rv) (Di Luca et al. 2012; Bottai et al. 2012). However, this effect was not observed in another M. tuberculosis strain (CDC1551) (E. N. G. Houben et al. 2012). In this study, we show that ESX-5 membrane components are indeed essential for in vitro growth of M. marinum and M. bovis BCG. Strikingly, this essentiality can be circumvented by permeabilization of the outer membrane. Finally, we provide evidence that ESX-5 is involved in the uptake of nutrients and propose a model linking ESX-5 substrates to nutrient uptake and essentiality.

RESULTS

eccC_5 and mycP_5 are essential for growth of M. marinum
To investigate the role of individual esx-5 genes in secretion and virulence of M. marinum, these genes were deleted by allelic exchange using a specialized transducing mycobacteriophage (Bardarov et al. 1997). We initially focused on eccC_5 and mycP_5, as both these genes are coding for highly conserved membrane components of the ESX-5 secretion system (E. N. G. Houben et al. 2012), but EccC_5 is part of the membrane complex, whereas MycP_5 is not and must have a separate function in ESX secretion. All attempts to delete these two genes in M. marinum were unsuccessful. To test whether this was due to the essentiality of these genes, we first introduced an integrative plasmid containing a kanamycin resistance cassette and the eccB_5-eccC_5 operon (pMV-eccBC_5-kan) (Bottai et al. 2012), or mycP_5 (pMV-mycP_5-kan). Using these merodiploid strains, the endogenous eccC_5 gene (Table 1, upper two rows) or mycP_5 gene, respectively, could readily be deleted, indicating that these genes are indeed essential.

To confirm the essentiality of EccC_5 and MycP_5, a switching procedure of the complementation vectors was used in which we replaced the original construct with either a complementation plasmid (pMV-eccBC_5-hyg or pMV-mycP_5-hyg), or an empty version of this integrative plasmid (pMV361-hyg). Using these merodiploid strains, the endogenous eccC_5 gene (Table 1, upper two rows) or mycP_5 gene, respectively, could readily be deleted, indicating that these genes are indeed essential.

To confirm the essentiality of EccC_5 and MycP_5, a switching procedure of the complementation vectors was used in which we replaced the original construct with either a complementation plasmid (pMV-eccBC_5-hyg or pMV-mycP_5-hyg), or an empty version of this integrative plasmid (pMV361-hyg). These plasmids integrate at the same site, but contain a hygromycin resistance cassette instead of the kanamycin cassette (Pashley and Parish 2003). Introduction of the complementation vectors resulted in successful switching to hygromycin-resistant colonies. PCR analysis confirmed proper exchange of plasmids in these colonies. In contrast, introduction of the empty vector produced no colonies that were hygromycin-resistant and kanamycin-sensitive. The few colonies that appeared after hygromycin selection were all kanamycin resistant, indicative of illegitimate recombination. PCR analysis confirmed the presence of both plasmids in these bacteria. Finally, we tested whether possible polar effects on eccB_5 expression, which could be caused by the endogenous eccC_5 deletion were responsible for the observed phenotype. The effect of a stop codon in either the eccB_5 or eccC_5 gene in the integrative plasmid during the switching procedure,
was scored. Successful introduction of plasmids containing a stop codon in \( \text{eccC}_5 \), either in the middle or at the end of the gene, was not possible, whereas introduction of an integrative plasmid with a stop codon in the beginning of \( \text{eccB}_5 \) resulted in a high number of switch mutants (Table 1, rows 3-5). This experiment confirms that deletion of \( \text{eccC}_5 \) has no negative polar effects on \( \text{eccB}_5 \). Together, these data strongly suggest that the two ESX-5 genes, \( \text{eccC}_5 \) and \( \text{mycP}_5 \), are required for in vitro growth of \( M. \text{marinum} \).

**MycP\(_5\)** is essential for growth of \( M. \text{bovis} \) BCG

To determine whether ESX-5 essentiality is also observed for the live vaccine strain \( M. \text{bovis} \) BCG, we created both a conditional \( \text{mycP}_5 \) expression strain and a conditional \( \text{mycP}_5 \) depletion strain by introducing a heterologous promoter including a multicopy \( \text{tetO} \) cassette immediately upstream of the \( \text{mycP}_5 \) gene of \( M. \text{bovis} \) BCG Pasteur. This regulatable \( \text{mycP}_5 \) gene was tested both in combination with a Tet repressor (TetR) protein exhibiting high-binding affinity to the \( \text{tetO} \) sites in absence of the inducer anhydrotetracycline (ATc; for establishing a \( \text{mycP}_5 \) tet-on system) or in combination with a mutated TetR protein with reversed binding affinity to \( \text{tetO} \) sites upon binding of ATc (for establishing a \( \text{mycP}_5 \) tet-off system) (Klotzsche, Ehrt, and Schnappinger 2009). The c-\( \text{mycP}_5 \)-tet-on strain was unable to grow on 7H10 plates without ATc (Figure 1A), while full growth was only observed at a concentration of 10 μg/ml ATc. Inversely, when c-\( \text{mycP}_5 \)-tet-off was grown on 7H10 plates supplemented with 10 μg/ml ATc, i.e. the \( \text{mycP}_5 \) depleting condition, colony growth was suppressed

---

**Table 1.** Essentiality of \( \text{eccC}_5 \) and analysis of functional domains. Replacement of pMV-\( \text{eccBC}_5 \) by the input DNA was scored. Input DNA consisted of the pMV-361-hyg plasmid containing the indicated constructs. “+” indicates that more than 100 colonies were detected after electroporation with the indicated vector. “-” indicates between 0-20 colonies were found after electroporation. These latter colonies were shown by PCR to still contain the original vector, indicating illegitimate recombination or spontaneous antibiotic resistance. Results are representative data of three independent experiments.

<table>
<thead>
<tr>
<th>input DNA</th>
<th>merodiploid (WT + pMV-kan-eccBC(_5))</th>
<th>delinquent (Δ( \text{eccC}_5 ) + pMV-kan-eccBC(_5))</th>
</tr>
</thead>
<tbody>
<tr>
<td>empty vector</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \text{eccBC}_5)-WT</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>( \text{eccB}_5) P145 stop</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>( \text{eccC}_5) V57 stop</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \text{eccC}_5) R1365stop</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \text{eccC}_5) K506A (NBD1)</td>
<td>+/-(^*$)</td>
<td>-</td>
</tr>
<tr>
<td>( \text{eccC}_5) K879A (NBD2)</td>
<td>+/-(^*$)</td>
<td>+/-(^*$)</td>
</tr>
<tr>
<td>( \text{eccC}_5) R1181A (NBD3)</td>
<td>+/-(^*$)</td>
<td>+/-(^*$)</td>
</tr>
<tr>
<td>( \text{eccC}_5) R1181K (NBD3)</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

\(^*$\) The \( \text{eccC}_5\), NBD mutants appear to have a dominant negative effect on the functioning of endogenous EccC\(_5\).

\(^*$\) Colonies showed a strong growth defect, i.e. colonies were visible only after 17 days, compared to 10 days for the wild-type strain.
(Figure 1B), although some colonies were still visible, possibly due to revertants of the tet-off system.

In addition, growth of c-mycP5-tet-on and c-mycP5-tet-off in liquid culture was tested using a resazurin reduction assay as a correlate of growth. By growing an inoculum of bacteria in the presence of varying concentration of ATc, we showed that growth of c-mycP5-tet-on correlated with ATc-dependent expression of mycP5 (Figure 1C). In contrast, growth of c-mycP5-tet-off was clearly inhibited in an ATc dependent matter (Figure 1D). Together these results show essentiality of mycP5 expression, and/or the downstream genes eccE5 and eccA5, for in vitro growth and metabolic activity of M. bovis BCG.

Increasing the permeability of the mycobacterial outer membrane rescues the essentiality of ESX-5.

We reasoned that the ESX-5 system could be essential for slow-growing mycobacteria due to toxic accumulation of an ESX-5 dependent substrate(s). In order to identify this putative toxic substrate M. marinum transposon mutants were selected that tolerate the deletion of eccC5. First, we generated a transposon library in the M. marinum eccC5 mutant, complemented with the integrative vector pMV-eccBC5-

Figure 1. Expression of MycP5 is essential for growth of M. bovis BCG. A, B) The BCG-Pasteur c-mycP5-tet-on (A) and c-mycP5-tet-off (B) mutants were grown for 21 days on Middlebrook 7H10 agar plates containing the indicated ATc concentrations. Full growth of c-mycP5-tet-on was only observed at 10 μg/ml ATc, whereas this concentration of ATc did not completely abolish colony growth of c-mycP5-tet-off. C, D) Resazurin reduction is dependent on ATc-induced expression/repression of mycP5. Cells of the BCG-Pasteur c-mycP5-tet-on (C), or c-mycP5-tet-off (D) mutants were grown as liquid cultures in 96-well microtiter plates for 6 days at 37°C at the indicated ATc concentrations, after which 10% Alamar Blue was added and fluorescence (585 nm) was measured after 16 h incubation to determine metabolic activity as a correlate of growth. Values are means of triplicates; error bars represent the standard deviation.
kan. Subsequently, a switching procedure was performed with an empty vector. Two transposon mutants were identified, in which proper switching of the two plasmids had occurred. Interestingly, the transposons in these mutants were located in the genes mas and ppsD, both of which are reported to be involved in the biosynthesis of PDIMs and phenolic glycolipids (PGLs) (Camacho et al. 1999; Yu et al. 2012; Alibaud et al. 2013). Biochemical analysis showed that the mas::tn mutant indeed lacked PDIMs in its lipid extracts (Figure S1). To verify that the transposon insertions in these genes were indeed responsible for rescuing essentiality of eccC5, complementation experiments were carried out using the switching approach described above. We generated various complementation plasmids carrying only the mas gene, or mas together with either the wild-type eccBC5 operon or the eccBC5 operon with a stop codon in eccC5. These plasmids were introduced in the mas or ppsD transposon mutant containing the eccC5 deletion, or in the complemented eccC5 mutant with an intact mas gene. Subsequently, we scored legitimate switching events in these mutants (Table 2). Complementation of the mas mutation in the eccC5 deletion background was only tolerated when the introduced plasmid simultaneously complemented the eccC5 deletion. This confirms that the absence of Mas is responsible for rescuing the essentiality of EccC5 for growth.

Although the mutations affecting PDIM/PGL biosynthesis were clearly associated with ESX-5 essentiality they did not seem to be linked to putative lethal substrates. Therefore, we had to reevaluate our hypothesis. Interestingly, mutants in the PDIM/PGL biosynthesis locus of M. marinum are known to be more sensitive towards various antibiotics (Yu et al. 2012), indicating that the integrity of the cell envelope is affected. To test whether this was also true for the M. marinum mas::tn-ΔeccC5 strain we tested its resistance against a combination of ampicillin and the beta-lactamase inhibitor clavulanic acid via disc diffusion. Clavulanic acid is included in this assay because M. marinum contains a chromosomally encoded beta-lactamase. Since both ampicillin and clavulanic acid have their target in the periplasm, growth impairment is indicative for a compromised outer membrane. The mas::tn-ΔeccC5 mutant indeed showed increased ampicillin/clavulanic acid sensitivity as compared to wild-type M. marinum. This phenotype was not affected by complementation of eccC5 (Figure 2A) and was therefore probably due to the lack of PDIM production.

Bacterial outer membranes function as permeability barriers and therefore so-called porin proteins are usually present to allow passive diffusion of small hydrophilic molecules (Jones and Niederweis 2010; Stephan et al. 2005; Stahl et al. 2001). The most-studied mycobacterial porin is MspA, which was identified in the fast-growing, nonpathogenic species Mycobacterium smegmatis. Orthologues of this porin can be found in other fast-growing mycobacteria, but are generally not found or produced in slow-growing mycobacterial species, such as M. tuberculosis and M. marinum. We therefore hypothesized that introduction of MspA in M. marinum would lead to a more permeable outer membrane and could therefore possibly also rescue the essentiality of ESX-5. To this end, we introduced an mspa-expressing plasmid
Figure S1. 2D-TLC analysis of apolar lipids of M. marinum mas and eccC₅ mutants. Apolar lipids were extracted from the various M. marinum strains and analyzed for the presence of PDIMs by 2D-TLC. The first dimension consisted of PE:ethyl ether 98:2, the second dimension of PE:acetone 98:2. Wild-type strains M⁴¹ (A) and M⁴¹⁸ (C) showed the presence of PDIM indicated by black arrows. No PDIMs could be detected in an mas::tn mutant picked up in an independent transposon screen (B), in a targeted knock-out of eccC₅ combined with a transposon insertion in mas, picked up in our transposon screen (mas::tn-ΔeccC₅; D) and the complementant of this latter strain (mas::tn-ΔeccC₅-C; E).

Table 2. Complementation of mas::tn in the ΔeccC₅ mutant by replacement of the integrated pMV vector. Indicated strains were electroporated with the input DNA shown on the left. Input DNA consisted of the pMV-361-hyg plasmid containing the indicated constructs. Valid insertion of input DNA was scored as + or -, similarly as described for Table 1. Results are representative data of three independent experiments.

<table>
<thead>
<tr>
<th>input DNA</th>
<th>M. marinum wt</th>
<th>ΔeccC₅ + pMV-eccBC₅</th>
<th>ΔeccC₅ + pMV361 + ppsD::tn</th>
<th>ΔeccC₅ + pMV361 + mas::tn</th>
</tr>
</thead>
<tbody>
<tr>
<td>empty vector</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>eccBC₅</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>mas</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>eccBC₅-stop-mas</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>eccBC₅-mas</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
in the *M. marinum* ΔeccC<sub>5</sub> and ΔmycP<sub>5</sub> complemented strains. Introduction of this plasmid indeed increased antibiotic sensitivity in the ΔmycP<sub>5</sub> complemented strain (Figure 2B) and also resulted in increased uptake of ethidium bromide in wild-type *M. marinum* (Figure 2C), which shows that MspA is functionally expressed in these strains (Yu et al. 2012; Stephan et al. 2004). Subsequently, the switching procedure with empty vector and eccC<sub>5</sub> or mycP<sub>5</sub>-expressing plasmids was conducted as before. Strikingly, both the empty and complementation vectors resulted in successful switching, showing that introduction of MspA indeed alleviates the requirement of eccC<sub>5</sub> and mycP<sub>5</sub> for growth. The permeability of mspA-expressing strains was not affected by the presence of an intact mycP<sub>5</sub> (Figure 2B). Tolerance for ESX-5 mutations was not due to spontaneous mutations in PDIM biosynthesis genes, as PDIM levels of these mutants were comparable to wild-type levels (Figure S2). These findings confirm that increasing the permeability of the mycobacterial outer membrane rescues the essentiality of ESX-5 in *M. marinum*.

We previously failed to isolate transposon mutants in any of the genes encoding ESX-5 membrane components by screening transposon mutant libraries for secretion defects (Abdallah et al. 2009; van der Woude et al. 2012). To determine whether we

---

**Figure 2.** mas mutation or introduction of mspA lead to increased outer membrane permeability. A, B) Sensitivity to a combination of ampicillin and clavulanic acid of different M. marinum strains was measured by performing a disc diffusion assay on the indicated strains and measuring the surface of the growth inhibition zone. The mas transposon (mas::tn) mutants exhibit increased sensitivity, independent of the presence of an intact copy of eccC<sub>5</sub> (A). Similarly, introduction of pSMT3::mspA also leads to an increase in sensitivity independent on the presence of mycP<sub>5</sub> (B). Values are the means of triplicates; error bars indicate the standard deviation. C) Uptake of EtBr, measured by flow cytometric analysis. M. marinum wild-type (dark grey) and M. marinum::mspA (black) were incubated with 20 μM EtBr for 60 min and 20.000 events were analyzed for their fluorescence intensity at 585/540 nm. Light-grey lines indicate unstained samples. All measurements are depicted in duplicates and are representatives of three independent experiments. ΔeccC<sub>5</sub>-C and ΔmycP<sub>5</sub>-C refer to the complementation strains of the *M. marinum* ΔeccC<sub>5</sub> or the ΔmycP<sub>5</sub> mutants complemented with pMV::eccBC<sub>5</sub> or pMV::mycP<sub>5</sub> respectively.
could now isolate such transposon mutants by introduction of MspA, we repeated our original screens using a transposon library created in M. marinum expressing mspA. This transposon library of ~10,000 mutants was screened for the secretion of the ESX-5 dependent PE_PGRS proteins, using the previously described double filter assay (van der Woude et al. 2012). In total, eight transposon mutants were identified that showed completely abolished PE_PGRS secretion. All eight secretion mutants had transposon insertions in the ESX-5 region (Figure 3A), four of which were affected in genes encoding the membrane components eccB₅, eccD₅ and mycP₅. Secretion analysis confirmed that these ESX-5 transposon mutants showed strongly reduced expression and secretion of PE_PGRS proteins and lacked expression of the mutated components (Figure 3B, Figure S3). To show reversibility of the phenotype, the MycP₅ transposon mutant (LA9) was complemented (Figure 3B). In conclusion, the introduction of MspA allowed transposon insertions in the ESX-5 locus and only mutants within the ESX-5 gene cluster showed a complete lack of PE_PGRS secretion, underscoring the importance of this locus in this process.

The role of the nucleotide binding domains of EccC₅ in essentiality and secretion

Because ESX-5 essentiality can be rescued by introduction of MspA, we hypothesized that one or more of the ESX-5 substrates could be responsible for the essentiality. However, it is also possible that the presence of the ESX-5 membrane complex itself is responsible for this phenomenon. We reasoned that we could distinguish between these possibilities by further dissecting the role of the membrane component EccC₅. EccC₅ together with EccB₅, EccD₅ and EccE₅, forms a large ~1.5 MDa complex in the mycobacterial cell envelope (E. N. G. Houben et al. 2012) that likely constitutes the membrane channel through which substrates are transported. In addition, EccC is predicted to be an ATPase with three nucleotide binding domains (NBDs). These

![Figure S2. Expression of MspA has no effects on PDIM expression. M. marinum::mspA-ΔeccC₅::pMV::eccBC₅ (A), M. marinum::mspA-ΔeccC₅::pMV361 (B) and an independent PDIM negative ppsD::tn mutant (C) were analyzed for the presence of PDIMs by 2D-TLC. 1D = PE:ethyl ether 98:2, 2D = PE:acetone 98:2. PDIMs are indicated by black arrows.](image-url)
NBDs usually contain a characteristic lysine residue in the Walker A motif, essential for ATP binding. However, the third NBD of EccC₅ has an arginine at this position, which is a conserved feature for NBD3 of ESX-5 systems (Figure 4A). Based on analogy to other ATPases involved in secretion (Moncalián et al. 1999; Cascales and Christie 2004), the NBDs of EccC play either a role in substrate transport or in the assembly of the membrane complex. We reasoned that by mutating these domains we could

![Figure 3. Secretion analysis of ESX-5 mutant strains. A) A schematic representation of the ESX-5 region of *M. marinum* with the different ESX-5 mutations used in this study. Bars above the gene cluster indicate regions deleted by targeted knock-out mutagenesis. Arrows below indicate position and orientation of transposons (named LA1 to LA12) in mutants of the parental strain *M. marinum::mspA* defective in ESX-5 dependent secretion. B) Secretion analysis of *M. marinum::mspA* (WT::mspA), a *mycP₅* transposon mutant (*mycP₅::tn*, corresponding to LA9 in (A)) and the complemented version of this strain (*mycP₅::tn-C*). Secreted proteins (S) were separated from bacterial cells (P) by centrifugation. In addition, surface-associated proteins were enriched from the bacterial cells by extraction with 0.5% Genapol X-080 (GS) and separated from non-solubilized proteins (GP) by centrifugation. All fractions were analyzed for the presence of PE_PGRS proteins by immunoblotting. GroEL2 staining was used as a loading and lysis control. C) Expression of EccB₅ and EspG₅ was analyzed by immunoblotting of total cell lysates of wild-type *M. marinum* (WT), the Δesx-5::mspA mutant and the complemented Δesx-5::esx-5₅mut strain. D) The same strains as under (C) were analyzed for their ability to express and secrete PE_PGRS proteins following the same procedure as under (B).
determine which function (i.e. in secretion or complex formation) is essential for bacterial growth.

EccC<sub>5</sub> variants with point mutations in the 3 NBD domains were introduced in the eccC<sub>5</sub> knock-out strain using the switching procedure outlined previously. In these EccC<sub>5</sub> variants the conserved lysine residue in NBD1 or NBD2 was replaced by an alanine, whereas the arginine residue in NBD3 was replaced by either by an alanine (R1181A) residue or the preferred lysine residue (R1181K) (Table 1, rows 6-9). While valid plasmid exchange was observed for the NBD2, NBD3-R1181A mutants, no valid exchange was observed for the NBD1 mutant, indicating that this first NBD is crucial for EccC<sub>5</sub> functioning. Although the NBD2 and NBD3 R1181A mutations were tolerated, these mutants showed a significant growth defect on plate. In contrast, NBD3-R1181K did not show any growth inhibition. Strikingly, when the same plasmids were introduced in wild-type <i>M. marinum</i>, the same growth inhibition phenotype was observed, suggesting that these EccC<sub>5</sub> NBD mutants have a dominant negative effect on the functioning of endogenous EccC<sub>5</sub>.

Next, we studied the effect of the mutations in EccC<sub>5</sub> on PE_PGRS secretion via the ESX-5 system. For this, we carried out the plasmid switching procedure similarly as above, but now using the eccC<sub>5</sub> knock-out strain that additionally contained

---

**Figure S3. Phenotype of <i>M. marinum espG<sub>5</sub> and eccB<sub>5</sub> transposon mutants.** <i>M. marinum::mspA</i> and transposon mutants LA1 (<i>espG<sub>5</sub>::tn</i>) and LA2 (<i>eccB<sub>5</sub>::tn</i>) were analyzed for their expression and secretion of PE-PGRS proteins (A) and expression of EccB<sub>5</sub> and EspG<sub>5</sub> (B) by immunoblotting. Bacterial pellets were incubated with Genapol X-080, after which non-extracted material (GP) was separated by the solubilized material (GS) by centrifugation. Non-treated bacterial pellets (P) were additionally analyzed for the expression of EccB<sub>5</sub> and EspG<sub>5</sub> (B).
the MspA-expressing vector (Figure 4B). The NBD1 mutation completely abolished the presence and secretion of PE_PGRS proteins, while NBD2 and NBD3-R1181A mutations strongly reduced expression and secretion of these ESX-5 substrates. These data show a strong correlation between the level of secretion of ESX-5 substrates and the essentiality of EccC5 for growth. Importantly, the lack of PE_PGRS secretion was not due to instability of the various EccC5 mutants, as was shown by western blot analysis of isolated cell envelope fractions (Figure 4C). Furthermore, Blue Native PAGE analysis of detergent-solubilized membrane proteins showed that formation of the EccBCDE5 membrane complex was also not affected by these NBD mutations (Figure 4D). We conclude that none of the three NBDs of EccC5 is involved in assembly of the ESX-5 membrane complex, suggesting that they are dedicated to energize the transport of substrates over the cell envelope. In turn, this indicates that proteins transported by ESX-5 and not the system itself are responsible for the essentiality of ESX-5.

Figure 4. Role of NBDs domains of EccC5 in ESX-5 dependent secretion and membrane complex assembly. A) Predicted transmembrane domains (dark grey), and NBD (light grey) of EccC5 are indicated. The positions of relevant residues are depicted with a black bar. The numbers represent the position in amino acids. B) Secretion of PE_PGRS proteins in the different EccC5 mutant strains was analyzed by immunoblot of supernatants and cell pellets of wild-type (WT) M. marinum and the eccC5 deletion strain (ΔeccC5) complemented with various eccC5 mutated genes. GroEL2 staining was used as a control for lysis and equal loading. C) Immunoblot analysis of EccC5 expression in isolated membranes of indicated strains. D) Blue native PAGE and immunoblot analysis using an anti-EccD5 antibody of the ESX-5 membrane of M. marinum ΔeccC5::mspA, complemented either with an empty vector (-) or with various eccC5 mutated genes. For all samples that contained EccC5 variants the characteristic pattern of ESX-5 membrane complexes was observed, consisting of the largest ~1.5 MDa complex and two additional smaller subcomplexes (indicated by the three arrowheads).
Deletion of *M. marinum* ESX-5 and complementation with the ESX-5 region of *M. tuberculosis*

The identification of a strategy to generate viable ESX-5 deletion strains also allowed us to determine the secretome of an esx-5 null-mutant. A targeted knock-out of the complete ESX-5 gene cluster of *M. marinum*, spanning the genes eccB<sub>5</sub> to eccA<sub>5</sub> (Figure 3A) was created in the presence of MspA (Δesx-5::mspA). In addition, we introduced an integrative plasmid containing the esx-5 locus of *M. tuberculosis* (pMV::esx-5<sub>tub</sub>) (E. N. G. Houben et al. 2012) in wild-type *M. marinum*, after which the endogenous ESX-5 region was deleted as described above (Δesx-5::esx-5<sub>tub</sub>). Interestingly, this approach was possible in the absence of MspA, although this strain showed slightly slower growth than wild-type bacteria. These data show that the ESX-5-region of *M. tuberculosis* can, at least partially, take over the essential role of the ESX-5 system of *M. marinum*.

In order to assess whether introduction of the esx-5 locus of *M. tuberculosis* also leads to a fully functional complementation, we analyzed the expression and secretion of PE_PGRS proteins (Figure 3D) and expression of EspG<sub>5</sub> (Daleke, Woude, et al. 2012) and EccB<sub>5</sub> (E. N. G. Houben et al. 2012) (Figure 3C). While expression of both EspG<sub>5</sub> and EccB<sub>5</sub> was indeed restored, expression and secretion of PE_PGRS proteins were not (Figure 3D). This shows that the *M. tuberculosis* ESX-5 region is not able to fully complement its *M. marinum* orthologues.

Next, the secretome of these newly constructed strains was analyzed by mass spectrometry. Since we have previously shown that the majority of ESX substrates in *M. marinum* remains attached to the cell surface, rather than being secreted into the growth medium (Sani et al. 2010), we focused on cell-surface proteins in this analysis. As a control we also determined the proteome of the cell envelope fraction obtained after cell disruption. Cell surface proteins can be released by incubating intact bacteria with the detergent Genapol X-080 (Cascioferro et al. 2007). Because this procedure also results in the extraction of low amounts of cell envelope proteins (Sani et al. 2010), we compared these fractions with bona fide cell envelope fractions and selected for Genapol-enriched proteins. *M. marinum* wild-type, *M. marinum::mspA*, Δesx-5::mspA and Δesx-5::esx-5<sub>tub</sub> strains were grown in liquid culture, after which cell envelope fractions were isolated using cell disruption and centrifugation or surface proteins were isolated using Genapol-X080 extraction. Protein samples from two independent experiments were analyzed by LC-MS/MS and spectral counts were used to measure relative abundance of proteins across the different strains and fractions (File S1). First, cell surface enriched and cell envelope fractions of the wild-type strain were normalized and compared. Proteins that had four-time higher relative abundance in the cell surface enriched fractions as compared to the cell envelope fraction were defined as probable cell surface proteins. This created a list of 114 proteins that contained many known surface-associated proteins and a number of lipoproteins (File S1). This subset of putative surface proteins was compared between the strains *M. marinum::mspA* and Δesx-5::mspA, resulting in
a list of 30 proteins that were at least five-fold less abundant in the esx-5-null mutant (Table 3). Please note that half of these proteins are completely absent in the esx-5 mutant. As expected, the majority (24) of these putative ESX-5 substrates were PE and PPE proteins, of which 17 were PE_PGRS proteins and 7 PPE proteins. Besides the PE/PPE proteins that were detected, six proteins with no apparent link to ESX-5 are also in this list. Five of these six proteins are annotated as secreted or surface associated proteins and contain canonical N-terminal signal sequences, which makes it most-likely they are exported by the Sec machinery (Kapopoulou, Lew, and Cole 2011). Together, these results confirm and extend our previous results from mutants in individual ESX-5 genes (Abdallah et al. 2009), and confirm that ESX-5 is the major secretion pathway of PE and PPE proteins.

To quantify the extent of complementation by the *M. tuberculosis* esx-5 locus we analyzed which ESX-5-dependent surface-associated proteins were five times more abundant in Δesx-5::esx-5_tub compared to Δesx-5::mspA. In concordance with the data obtained by immunoblot, the secretion of only a limited number of ESX-5 substrates was restored by the introduction of esx-5_tub (Table 3). Among these few substrates were two PE/PPE proteins, namely PPE10 (MMAR_0761) and PE_PGRS50 (MMAR_2656 – PE_PGRS50). Previously, high density transposon mutagenesis analysis has shown that the genes encoding these proteins are not essential for *M. marinum* E11 (Weerdenburg et al. 2015), so either they are not the cause of the essentiality of ESX-5 or they are together responsible. The three other restored proteins are MMAR_4153, MMAR_3410 and MMAR_2586, all of which are not essential, do not have a T7S signal (Daleke, Ummels, et al. 2012) and have not been linked to ESX-5 secretion previously.

Because our analysis of surface-associated proteins could not pinpoint essential ESX-5 substrates, we hypothesized from the observed link with outer membrane permeability that the essential ESX-5 substrate could be integrally inserted and therefore more stably associated with the outer membrane. Therefore, we also compared the cell envelope fractions of the same set of strains (Table S1). As expected, the ESX-5 membrane components EccB_5 to EccE_5, and MycP_5 were detected in large amounts in both wild-type *M. marinum* strains (i.e. with or without episomal mspA) as well as in Δesx-5::esx-5_tub but could not be detected in Δesx-5::mspA. This analysis furthermore showed that there were indeed two putative ESX-5 substrates detectable in the cell envelope fraction. MMAR_1129 is a PPE protein with similarities to *M. tuberculosis* PPE64, whereas MMAR_1442 is a PE_PGRS protein similar to the PE_PGRS27 in *M. tuberculosis*. Both these proteins are shown to be cell envelope localized in an ESX-5 dependent manner, but are not complemented by the esx-5-locus of *M. tuberculosis*. Again, the genes encoding these proteins are non-essential in *M. marinum* (Weerdenburg et al. 2015). Interestingly, several proteins involved in lipid biosynthesis also seem affected by the ESX-5 mutation and are partially complemented in Δesx-5::esx-5_tub, suggesting that lack of a functional ESX-5 has implications for lipid metabolism. In summary, introduction of the ESX-5 region
of *M. tuberculosis* leads to rescue of essentiality of the ESX-5 region in *M. marinum*, which shows that the essential role of ESX-5 is conserved. However, the ESX-5 region of *M. tuberculosis* is only marginally able to restore ESX-5-dependent secretion in *M. marinum*, suggesting not only that substrate recognition is (partially) species specific.

Table 3. ESX-5-dependent surface proteins of *M. marinum*. Cell surface enriched and cell envelope proteins of *M. marinum::mspA* and Δesx-5::mspA strains were analyzed by LC-MS/MS. Proteins were classified as secreted proteins when the relative normalized abundance was four times higher in the cell-surface enriched fractions as compared to the cell envelope. Shown here are only proteins with at least a five-fold decrease in spectral counts in the Δesx-5::mspA strain. Grey highlighted rows indicate proteins with restored secretion in the presence of esx-5*tub*.

<table>
<thead>
<tr>
<th>Identified protein</th>
<th>Protein description</th>
<th>M.marinum::mspA</th>
<th>Δesx-5::mspA</th>
<th>Δesx-5::esx-5<em>tub</em></th>
<th>Fold Change</th>
<th>wild-type/Δesx-5::mspA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMAR_3105</td>
<td>PE-PGRS</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_4999</td>
<td>PE-PGRS</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_2459</td>
<td>PE-PGRS</td>
<td>54</td>
<td>0</td>
<td>2</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_2656</td>
<td>PE-PGRS550</td>
<td>54</td>
<td>0</td>
<td>11</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_3400</td>
<td>PE-PGRS</td>
<td>48</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>iiPb</td>
<td></td>
<td>46</td>
<td>0</td>
<td>2</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_1199</td>
<td>PE-PGRS</td>
<td>38</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_5135</td>
<td>PE-PGRS</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_1908</td>
<td>Membrane hydrolase</td>
<td>35</td>
<td>0</td>
<td>2</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_3316</td>
<td>PE-PGRS</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_2460</td>
<td>PE-PGRS</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_3728</td>
<td>PE-PGRS52</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_1130</td>
<td>PPE</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_4939</td>
<td>PE-PGRS</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_4153</td>
<td>Conserved secreted</td>
<td>11</td>
<td>0</td>
<td>33</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>MMAR_1442</td>
<td>PE-PGRS</td>
<td>180</td>
<td>3</td>
<td>0</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>MMAR_2235</td>
<td>PE-PGRS PE16</td>
<td>118</td>
<td>3</td>
<td>0</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>MMAR_5013</td>
<td>PE-PGRS</td>
<td>43</td>
<td>3</td>
<td>0</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>MMAR_3665</td>
<td>PPE</td>
<td>94</td>
<td>10</td>
<td>27</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>MMAR_0382</td>
<td>PE-PGRS</td>
<td>65</td>
<td>7</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>MMAR_0761</td>
<td>PPE10</td>
<td>62</td>
<td>7</td>
<td>51</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>MMAR_4187</td>
<td>PPE</td>
<td>29</td>
<td>3</td>
<td>0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>MMAR_1497</td>
<td>PPE28</td>
<td>57</td>
<td>7</td>
<td>0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>MMAR_2973</td>
<td>PE-PGRS</td>
<td>82</td>
<td>10</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>MMAR_1402</td>
<td>PPE</td>
<td>153</td>
<td>19</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>MMAR_5047</td>
<td>PPE</td>
<td>50</td>
<td>6</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>MMAR_1547</td>
<td>PE-PGRS</td>
<td>25</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>MMAR_2586</td>
<td>Conserved secreted</td>
<td>212</td>
<td>27</td>
<td>213</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>MMAR_3410</td>
<td>Pyruvate kinase</td>
<td>70</td>
<td>9</td>
<td>49</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>MMAR_0559</td>
<td>Conserved exported</td>
<td>43</td>
<td>6</td>
<td>11</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

* Average normalized spectral counts from two biological replicates
Table S1. ESX-5-dependent cell envelope proteins of *M. marinum*. Cell envelope proteins of *M. marinum::mspA*, Δesx-5::mspA and esx-5<sub>tb</sub> strains were analyzed by LC-MS/MS. Proteins were classified as ESX-5-dependent when the normalized spectral counts between *M. marinum::mspA* and Δesx-5::mspA were reduced at least 10-fold. Grey highlighted rows indicate the conserved ESX-5 components.

<table>
<thead>
<tr>
<th>Identified protein</th>
<th>Functional description</th>
<th>MS/MS Normalized spectral counts&lt;sup&gt;#&lt;/sup&gt;</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMAR_2676</td>
<td>EspG&lt;sub&gt;3&lt;/sub&gt;</td>
<td>17</td>
<td>9.8*</td>
</tr>
<tr>
<td>MMAR_2680</td>
<td>EccA&lt;sub&gt;3&lt;/sub&gt;</td>
<td>12</td>
<td>2.3*</td>
</tr>
<tr>
<td>MMAR_1442</td>
<td>PE_PGRS</td>
<td>8.8</td>
<td>∞</td>
</tr>
<tr>
<td>MMAR_3797</td>
<td>Lipid metabolism</td>
<td>8.5</td>
<td>∞</td>
</tr>
<tr>
<td>MMAR_1129</td>
<td>PPE</td>
<td>8.5</td>
<td>∞</td>
</tr>
<tr>
<td>CtpA_1</td>
<td>Cation transporter</td>
<td>7.5</td>
<td>∞</td>
</tr>
<tr>
<td>PapA3</td>
<td>Lipid metabolism</td>
<td>6.5</td>
<td>∞</td>
</tr>
<tr>
<td>FadD9</td>
<td>Lipid metabolism</td>
<td>6.2</td>
<td>∞</td>
</tr>
<tr>
<td>MMAR_3408</td>
<td>Oxidoreductase</td>
<td>6.2</td>
<td>∞</td>
</tr>
<tr>
<td>CobN</td>
<td>Cobalamin biosynthesis</td>
<td>5.2</td>
<td>∞</td>
</tr>
<tr>
<td>MMAR_2679</td>
<td>EccE&lt;sub&gt;3&lt;/sub&gt;</td>
<td>45</td>
<td>19*</td>
</tr>
<tr>
<td>FadD26</td>
<td>Lipid metabolism</td>
<td>12</td>
<td>34</td>
</tr>
<tr>
<td>MMAR_2677</td>
<td>EccD&lt;sub&gt;3&lt;/sub&gt;</td>
<td>60*</td>
<td>27</td>
</tr>
<tr>
<td>MMAR_1790</td>
<td>Cons. Hypothetical</td>
<td>9.8</td>
<td>26</td>
</tr>
<tr>
<td>MetE</td>
<td>Methionine biosynthesis</td>
<td>7.9</td>
<td>26</td>
</tr>
<tr>
<td>MMAR_2664</td>
<td>EccB&lt;sub&gt;3&lt;/sub&gt;</td>
<td>40*</td>
<td>21</td>
</tr>
<tr>
<td>MMAR_2665</td>
<td>EccC&lt;sub&gt;3&lt;/sub&gt;</td>
<td>162*</td>
<td>19</td>
</tr>
<tr>
<td>LeuC</td>
<td>Leucine biosynthesis</td>
<td>8.5</td>
<td>13</td>
</tr>
<tr>
<td>SigB</td>
<td>Sigma factor B</td>
<td>8.1</td>
<td>12</td>
</tr>
<tr>
<td>LeuS</td>
<td>Leucine tRNA ligase</td>
<td>7.2</td>
<td>11</td>
</tr>
<tr>
<td>OtSB1</td>
<td>Trehalose Biosynthesis</td>
<td>7.5</td>
<td>11</td>
</tr>
</tbody>
</table>

* Spectral counts for *M. marinum* and *M. tuberculosis* orthologues were combined into one value.
# Spectral counts were normalized in a two-way analysis between *M.marinum::mspA* and Δesx-5::mspA, or in a two-way analysis between *M. marinum* and Δesx-5::esx-5<sub>tb</sub>. Data for wild-type *M. marinum* are not depicted.
but also that the major portion of the identified ESX-5 substrates do not cause the essentiality of the secretion system.

Transposon directed insertion site sequencing
To directly identify ESX-5 substrates that cause the essentiality of ESX-5 we performed the genome-wide approach of transposon directed insertion site sequencing (TraDIS). This extensive technique allows the analysis of large libraries of random transposon insertion mutants and as such is able to identify which genes are essential under different conditions or in different genetic backgrounds (Griffin et al. 2011; Pritchard et al. 2014). We created >100,000 Mycomar transposon mutants of wild-type M. marinum and M. marinum::mspA, which number has been shown before to result in hitting 97% of all non-essential TA sites (Weerdenburg et al. 2015). Subsequently, transposon insertion sites were determined by Illumina sequencing, the data were normalized and the number of transposon hits per gene was established. By comparing the two libraries we could identify genes that are specifically enriched (>3 fold) in bacteria expressing mspA (File S2). As expected, the ESX-5 membrane components were found among the top hits (Table 4). For instance, transposon insertions in eccD5 were detected 1018-fold more in the mspA-expressing strain and similar patterns were seen for mycP5 (181x) eccB5 (62x), eccE5 (49x) and eccC5 (44x). Interestingly no ESX-5 substrates were identified among the top hits, although some genes encoding substrates were somewhat enriched. Insertions in ppe1 (mmar_0261), mmar_3290 (an M. marinum specific PE_PGRS) and ppe59 (mmar_4187) were 3-4 fold more common when mspA was expressed. This result indicates that not a single ESX-5 substrate, but multiple proteins together are responsible for the essentiality of ESX-5.

This transposon mutagenesis approach did reveal other effects of the presence of MspA. Overall, insertions in genes involved in lipid metabolism seem to confer a growth advantage in M. marinum::mspA, even though many of these genes are not essential. For instance, mutations in desA3 (mmar_1315), a stearoyl coenzyme A desaturase involved in the biosynthesis of oleic acid, are enriched 18-fold, reaching 0.12% of the amount of total insertions. A similar phenomenon can be observed for most genes of the mce1 locus. Although insertions in these genes are not enriched dramatically, they can be observed over the complete gene cluster. (yrbE1A (4x), mce1B (3.7x), mce1D (3x), mmar_0419 (4.7x) and mmar_0421 (3.4x). Another affected gene known to play a role in lipid metabolism is icl (mmar_1792) encoding for the isocitrate lyase enzyme (Eoh and Rhee 2014). Together these data show that when mspA is expressed in M. marinum there is a shift in lipid metabolism, which could be due to higher availability of simple carbon sources, such as glucose and glycerol, by the presence of the hydrophilic pore. However, the most dramatic change is the reduced essentiality of esx-5. As this effect could almost exclusively be observed for esx-5 genes and not for genes encoding ESX-5 substrates, the substrates that cause the essentiality of ESX-5 are likely redundant.
Table 4. *M. marinum* genes with enriched numbers of transposon insertions in *M. marinum* supplemented with MspA. Depicted are the number of transposon insertions detected by TraDIS in wild-type *M. marinum* or in a strain expressing MspA. The detected insertions are normalized for the total amount of reads per sample and genes are ranked based on the fold change between *M.marinum::mspA* divided by wild type *M.marinum*. The top ten hits are depicted. ESX-5 components are highlighted in grey.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene description</th>
<th>H37Rv orthologue</th>
<th>Gene length (bp)</th>
<th>Insertion counts</th>
<th>Normalized fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><em>M. marinum</em> wild-type</td>
<td><em>M. marinum::mspA</em></td>
</tr>
<tr>
<td>mmar_2677</td>
<td>eccD&lt;sub&gt;5&lt;/sub&gt;</td>
<td>rv1795</td>
<td>1512</td>
<td>30</td>
<td>1463</td>
</tr>
<tr>
<td>mmar_2221</td>
<td>ribC</td>
<td>rv1412</td>
<td>624</td>
<td>2</td>
<td>104</td>
</tr>
<tr>
<td>mmar_2678</td>
<td>mycP&lt;sub&gt;5&lt;/sub&gt;</td>
<td>rv1796</td>
<td>1743</td>
<td>36</td>
<td>840</td>
</tr>
<tr>
<td>mmar_5136</td>
<td>DNA polymerase III</td>
<td>Rv3664</td>
<td>1218</td>
<td>3</td>
<td>126</td>
</tr>
<tr>
<td>mmar_3746</td>
<td>Lipase/esterase</td>
<td>Rv2418</td>
<td>756</td>
<td>8</td>
<td>590</td>
</tr>
<tr>
<td>mmar_3941</td>
<td>Phage protein</td>
<td>-</td>
<td>336</td>
<td>13</td>
<td>105</td>
</tr>
<tr>
<td>mmar_2664</td>
<td>eccB&lt;sub&gt;5&lt;/sub&gt;</td>
<td>rv1782</td>
<td>1524</td>
<td>28</td>
<td>353</td>
</tr>
<tr>
<td>mmar_2679</td>
<td>eccE&lt;sub&gt;5&lt;/sub&gt;</td>
<td>rv1797</td>
<td>1212</td>
<td>12</td>
<td>124</td>
</tr>
<tr>
<td>mmar_2665</td>
<td>eccC&lt;sub&gt;5&lt;/sub&gt;</td>
<td>rv1783</td>
<td>4167</td>
<td>64</td>
<td>1443</td>
</tr>
<tr>
<td>mmar_1257</td>
<td>birA</td>
<td>Rv3279c</td>
<td>831</td>
<td>4</td>
<td>122</td>
</tr>
</tbody>
</table>
The ESX-5 system is involved in nutrient uptake

Since essentiality of ESX-5 could be rescued by increasing outer membrane permeability, we hypothesized that ESX-5 substrates could be involved in the uptake of nutrients that are essential for growth. Because TraDIS analysis did indicate that this effect was probably due to multiple substrates, we set out to test whether we could find specific carbon sources that were not transported in the ESX-5 mutant. One limitation in this analysis is of course that these mutants always contain the large hydrophilic MspA pore.

To test whether the ESX-5 system is involved in the uptake of nutrients, \( M. \text{marinum-} \Delta \text{mycP}_5::\text{mspA} \) and several control strains were grown in a modified 7H9 medium supplemented with different single carbon sources (Capyk et al. 2009). The mycP\(_5\) mutant grew almost as fast as the control strains in the presence of small hydrophilic carbon sources such as glucose, glycerol or acetate (Figure S4A). This was expected, as these strains contained the hydrophilic pore-forming MspA protein (Stephan et al. 2005; Stahl et al. 2001). However, when the strains were grown on medium with Tween-80 (Figure 5A) or Tween-40 (Figure S4B) as sole carbon source, only the mycP\(_5\) deletion strain showed strongly reduced growth. Mycobacteria are able to hydrolyze Tween and use the fatty acid components as carbon source. However, it is also known that free fatty acids can be toxic for mycobacteria. To discriminate between these two possibilities we added 0.2% glucose to the cultures after 8 days of growth on Tween-80. This resulted in normal outgrowth of the mycP\(_5\) mutant, indicating that the Tween-80 present in the medium did not specifically hamper growth of this strain, but that these cells were still viable and therefore probably starved (Figure S4C). This result indicates a role for the ESX-5 system in either the (extracellular) hydrolysis of Tween-80 or the uptake of released oleic acid. To examine whether ESX-5 secreted substrates are involved in the breakdown of Tween-80, co-culture experiments using wild-type \( M. \text{marinum} \) and \( M. \text{marinum-} \Delta \text{mycP}_5::\text{mspA} \) were performed. Growth of the mycP\(_5\) mutant strain was not rescued by the presence of wild-type bacteria (Figure S4D), indicating that factors secreted to the culture filtrate do not play a role in the observed growth defect. This was further confirmed by testing the role of the ESX-5 dependent lipase LipY in the ability of \( M. \text{marinum} \) to grow on Tween-80. LipY is the most active and abundant lipase secreted via ESX-5 (Daleke et al. 2011) and therefore a prime candidate for hydrolyzing Tween-80. However, an \( M. \text{marinum lipY} \) deletion mutant grew to a similar extent as the wild-type strain on medium with Tween-80 as a sole carbon source (Figure S4E), showing that LipY is not responsible for the ESX-5 dependent growth on Tween-80.

Next, we investigated in a more direct manner whether ESX-5 mutants are able to import fatty acids. Under hypoxic conditions mycobacteria incorporate fatty acids as intracellular triacylglycerol (TAG) in so-called lipid bodies (Daniel et al. 2011; Sirakova et al. 2006). These lipid bodies can be visualized by adding BODIPY-labeled fluorescent fatty acids to hypoxic cultures. \( M. \text{marinum-} \Delta \text{mycP}_5::\text{mspA} \) and the corresponding control strains were grown in the presence of a fluorescently
labeled fatty acid under hypoxic conditions. Uptake of this fatty acid was determined by measuring the fluorescence of the bacteria by FACS analysis (Figure 5B). The mycP<sub>5</sub>-deletion strain showed significantly reduced fluorescent intensity as compared to the complemented or wild-type strains (Figure 5C). This effect was confirmed by confocal microscopy (Figure 5D); while the wild-type and the complemented mycP<sub>5</sub>

![Graphs showing growth of M. marinum on different carbon sources](image)

**Figure S4. Growth of M. marinum on different carbon sources.** A) Growth of indicated M. marinum strains on 0.2% glucose, acetate or glycerol as a sole carbon source was assessed by measuring optical density at different time points. B) Growth of indicated strains at Tween-40. C) M. marinum ΔmycP<sub>5</sub>::mspA cannot grow on Tween-80 as a sole carbon source, but is able to grow normally when 0.2% glucose was added at day 8 (indicated by the arrow). D) M. marinum::pSMT3-crimson is not able to rescue growth of the mycP<sub>5</sub> mutant. A mixed culture of the two indicated strains on Tween-80 as a sole carbon source was performed. On the indicated time points bacteria were plated out on 7H10 plates with hygromycin and colony forming units (CFU) were quantified by counting fluorescent (WT) and non-fluorescent (ΔmycP<sub>5</sub>) colonies. E) M. marinum ΔlipY can grow normally on Tween-80 as a sole carbon source. In B-D, error bars depict the standard deviation over three independent cultures.
strain showed intracellular fluorescent lipid bodies, indicating that the fluorescent fatty acid was incorporated and stored in lipid bodies (Daniel et al. 2011), no fluorescent lipid bodies were observed in the mycP5 deletion strain. These data together show that ESX-5 or one of it substrates is involved in the uptake of fatty acids. Together these data show that ESX-5 substrates are involved in the efficient utilization of fatty acids.

**Figure 5. ESX-5 is involved in fatty acid uptake.** A) Growth of indicated *M. marinum* strains on Tween-80 as a sole carbon source was assessed by measuring optical density at different time points. Depicted is the average of three biological replicates. Error bars indicate standard deviations. B) Uptake of a fluorescently labeled fatty acid after 72 hours of hypoxic growth was measured by FACS analysis. 20,000 events gated for similar size were acquired for WT::mspA (black), ∆mycP5::mspA (light grey) or ∆mycP5-C::mspA (dark grey). C) Quantification of FACS analysis. Mean fluorescent intensity of three experiments per strain was acquired by FACS. Background staining, quantified by adding the fluorescent fatty acid to an unstained culture one hour before washing the cells, was deducted from the measured values. Error bars indicate the standard deviations and One-way ANOVA showed a statistical difference between the samples of \( p = 0.010 \). D) Uptake of the fluorescently labeled fatty acid and formation of lipid bodies was confirmed by confocal microscopy.
DISCUSSION

Previous results suggested that the ESX-5 system could be essential for growth of *M. marinum* (Abdallah et al. 2006; Abdallah et al. 2009; van der Woude et al. 2012). Here, we show that the membrane components EccC$_5$ and MycP$_5$ are indeed essential for this species. In addition, silencing of the *mycP$_5$* gene in *M. bovis* BCG results in significantly reduced growth. This phenotype is similar to the phenotype of an eccC$_5$ depletion strain of *M. tuberculosis* H37Rv (Di Luca et al. 2012). Strikingly, we were able to obtain mutants in the ESX-5 core components when the outer membrane permeability was increased, either by mutating PDIM biosynthesis genes or by introduction of the *M. smegmatis* outer membrane porin MspA.

One hypothesis that would explain this observation is that a defect in ESX-5 dependent secretion could cause the accumulation of toxic molecules in the periplasm, which are able to exit the cell via the MspA porin or through the more permeabilized outer membrane. However, it is difficult to envision how the absence of the ESX-5 protein secretion system could cause the periplasmic/cytoplasmic accumulation of such molecules. We therefore favor and tested the alternative hypothesis that the ESX-5 system is involved in the influx of nutrients and/or other metabolites that are crucial for growth. In support of the second hypothesis we could show that ESX-5 mutations strongly affected the ability of *M. marinum* to grow on different polysorbate detergents (i.e. Tween-40 and Tween-80), suggesting that ESX-5 facilitates the usage of fatty acids as a carbon source. In addition, the ESX-5 mutant is also impaired in the intracellular accumulation of fluorescent fatty acids. Notably, the ESX-5-dependent growth on polysorbate-like detergents and the uptake of fatty acids were observed in the presence of the MspA porin. Fatty acid uptake therefore seems to be relatively specific for ESX-5. Pathogenic mycobacteria accumulate lipid bodies, presumably as energy storage, during the dormant stage of infection. Our data suggest that ESX-5 plays a central role in this crucial process in infection. Interestingly, earlier studies have shown that an *espG$_5$* mutant, which shows strongly diminished ESX-5 dependent secretion, is in fact hypervirulent in adult zebrafish (Weerdenburg et al. 2012). It is possible that the inability of this mutant strain to take up fatty acids during infection prevents the bacterium to go into dormancy, resulting in outgrowth of the mutant in the host. Possibly, ESX-5 is also required for the utilization of other substrates. However, the uptake of hydrophilic substrates is more difficult to test, because we need the expression of the MspA porin when *esx-5* is deleted. Interestingly, also the ESX-3 secretion system of Mycobacteria is involved in nutrient import; this system is essential for the uptake of iron and zinc ions in *M. tuberculosis* (Siegrist et al. 2009). Unfortunately, the exact mechanism for metal ion uptake via ESX-3 is not known. Our TraDIS analysis shows that essentiality of this system cannot be alleviated by the introduction of MspA.

Introduction of MspA or defects in the biosynthesis of PDIM makes the outer membrane more permeable, which would allow passive diffusion of nutrients, thereby circumventing the requirement for the ESX-5-dependent nutrient uptake (Figure 6B).
Interestingly, while a compromised outer membrane by the absence of PDIM likely enables a more efficient influx of hydrophobic solutes, MspA is a water-filled channel and allows predominantly diffusion of small hydrophilic molecules. This would suggest that ESX-5 is involved in the influx of both hydrophobic and hydrophilic molecules. However, the presence of MspA has been shown to also increase the sensitivity of *M. smegmatis* for hydrophobic antibiotics such as erythromycin, from which it was hypothesized that this porin might additionally affect the integrity of the outer membrane (Stephan et al. 2004). This perhaps also explains why we and others (Purdy, Niederweis, and Russell 2009) observed an increase in uptake of ethidium bromide in the presence of MspA, while this molecule is theoretically too bulky to fit in the MspA channel. In addition, PDIM mutants have been shown, both in this study and by others (Yu et al. 2012; Camacho et al. 2001), to be more sensitive to both hydrophobic and hydrophilic antimicrobials.

Which proteins are responsible for the ESX-5-dependent uptake of fatty acids and possibly also other nutrients? Our analysis of the NBDs of EccC₅ indicates that it is not the presence of the ESX-5 membrane complex, but active secretion through this complex. The observed nutrient influx is therefore probably mediated by ESX-5 substrates that possibly form outer membrane porins or other types of outer membrane channels (Figure 6A). To identify these essential ESX-5 substrates, we analyzed the complete ESX-5 secretome of *M. marinum*. LC-MS/MS analysis showed that this mutant strain was deficient in secretion of all abundant PE and PPE proteins. The only PE/PPE proteins that have been shown previously to be independent on
ESX-5 are the ESX-1 dependent PE35/PPE68_1 protein couple that is associated with the ESX-1 gene cluster (Sani et al. 2010; Daleke, Ummels, et al. 2012). These proteins were not detected in sufficient amounts in this study to draw any conclusions. Interestingly, while the ESX-5 region of *M. tuberculosis* was able to complement the *M. marinum* ESX-5 system for growth, only a small number of relatively low-abundant *M. marinum* ESX-5 substrates were secreted by the ESX-5 system of *M. tuberculosis*. Possibly, the recognition of ESX-5 substrates is relatively species specific or species-specific chaperones are required for efficient secretion. This is not surprising as the PE/PPE proteins show a high variation between species (Sampson 2011), and especially the more species-specific PE/PPE proteins are the ESX-5 substrates that are most highly expressed in *M. marinum* (Abdallah et al. 2009). As *M. marinum* Δesx-5::esx-5 was viable, the partial complementation allowed us to generate a limited shortlist of PE and PPE proteins that could be responsible for the essential role of ESX-5 for growth. However, it should be noted that there were more (potential) substrates that did not reach the threshold levels, but showed a trend of partial complementation. In addition, PE and PPE proteins are notoriously difficult to identify using proteomics because of their atypical composition (i.e. many have extremely low number of potential trypsin digestion sites). A recent study showed that also other proteins, not belonging to the PE/PPE protein families, are predicted T7S substrates (Daleke, Ummels, et al. 2012). One of these predicted T7S substrates, Rv3903c or CpnT, has recently been shown to be an outer membrane porin involved in glycerol uptake (Danilchanka et al. 2014). This could be the first example of a T7S substrate that forms an outer membrane porin. However, we were not able to detect CpnT in our cell surface enriched protein extracts or in our cell envelope fractions of *M. marinum* (van der Woude et al. 2013), suggesting that this protein is not produced in significant amounts in this species. Alternatively, this outer membrane protein is somehow lost during our sample preparations, which could mean that we also miss other outer membrane channels in this analysis. Another recent study on the structure of the ESX-1 substrate EspB gives an interesting insight into possible channel formation of ESX substrates (Solomonson et al. 2015). EspB is organized as a ring-shaped heptamer with a central pore and with one side of the ring hydrophobic, suggesting that this protein could form membrane channels. Interestingly the EspB fold is highly similar to PE-PPE, which means that also PE-PPE proteins could form such ring-shaped pores. Our TraDIS analysis indicates that multiple (non-essential) substrates are responsible for the essential phenotype of ESX-5, which makes it difficult to identify them. This is not unexpected, since there are many ESX-5 substrates and they can be divided in major homology groups.

Although the ESX-5 system of *M. tuberculosis* can alleviate the essentiality of ESX-5 in *M. marinum*, the role of ESX-5 in *M. tuberculosis* does not necessarily have to be identical. There have been conflicting results concerning the essentiality of the ESX-5 system in *M. tuberculosis*. Di Luca et. al (Di Luca et al. 2012) have demonstrated the essentiality of EccB_5 and EccC_5 in *M. tuberculosis* H37Rv,
but in another study they reported that deletion of eccD and espG (Rv1794) (Bottai et al. 2012) was possible in this strain. In contrast, in M. tuberculosis CDC1551 transposon insertions in eccC and eccD that block ESX-5 functioning have been described (E. N. G. Houben et al. 2012). These results indicate that there are significant differences between different strains of M. tuberculosis. Based on the results in this study, it can be hypothesized that these differences could also be caused by differences in outer membrane permeability, for instance due to spontaneous mutations in PDIM biogenesis genes. Such mutations have been shown to occur with high frequency in M. tuberculosis, because they confer a growth advantage in culture (Kirksey et al. 2011; Domenech and Reed 2009). In order to investigate whether outer membrane permeability was affected in the CDC1551-derived ESX-5 mutant strains reported in Houben et al. (E. N. G. Houben et al. 2012), we tested the antibiotic sensitivity of these mutant strains and their complemented counterparts. Indeed, we observed a major increase in ampicillin sensitivity as compared to wild-type M. tuberculosis, in both the eccC and eccD mutant strains (Figure S5). Furthermore, the ESX-5 complemented strains still showed this increased susceptibility. However, this increased antibiotic sensitivity was not linked to the absence of PDIMs (Figure S6). Therefore, probably other factors are responsible for increased outer membrane permeability in these

![Figure S5. Ampicillin susceptibility of M. tuberculosis CDC1551 ESX-5 mutants.](image)

M. tuberculosis H37Rv and CDC1551 wild-type strains, CDC1551 derived transposon mutants in eccC (eccC::tn) or eccD (eccD::tn); and their respective complemented strains (eccC::tn-C and eccD::tn-C), were grown on 7H10 plates containing a combination of 30μg/ml ampicillin and 30μg/ml clavulanic acid (A), or without antibiotics (B). 5 μl of a 1.0 OD600 culture was spotted in serial dilutions (top row, undiluted; second row, 10x diluted; third row 100x diluted). In the absence of antibiotics, the H37Rv strain grew faster than the CDC1551 strains, while the eccC::tn strain showed a growth defect, which was alleviated upon complementation. Both ESX-5 transposon mutants and their respective complementants did not grow in the presence of 30μg/ml ampicillin and 30μg/ml clavulanic acid, indicating increased membrane permeability in these strains compared to the wild-type strain.
Figure S6. PDIMs are present in *M. tuberculosis* CDC1551 esx-5 mutants. Apolar lipids were extracted from *M. tuberculosis* CDC1551 (A), and transposon mutants eccC5::tn (rv1783, B) and eccD5::tn (rv1795, C) as described before and were analyzed for the presence of PDIMs by 2D-TLC. 1D = PE:ethyl ether 98:2, 2D = PE:acetone 98:2. PDIMs are indicated by black arrows. PDIMs are visible in all strains, although the eccC5 mutant appears to make higher amounts of PDIMs.
added at the following concentration: kanamycin, 25 mg ml\(^{-1}\), hygromycin, 50 mg ml\(^{-1}\) for mycobacteria and 100 mg ml\(^{-1}\) for *E. coli* and streptomycin 35 mg ml\(^{-1}\). For growth on defined carbon sources, growth medium was created as described by Capyk et al. (Capyk et al. 2009) and carbon sources were added to a final concentration of 0.2% w/v.

**Construction of plasmids**

Anchored primers (MunI for the 5′ primer and HindIII for the 3′ primer; for sequences see Table S1) were used to amplify both *eccB*\(_5\) and *eccC*\(_5\) from *M. marinum* M strain genomic DNA by PCR. Amplicons were cloned as MunI-HindIII fragments in EcoRI-HindIII digested pMV361 (Stover et al. 1991), resulting in pMV-eccBC\(_5\). Similarly, anchored primers (EcoRI for the 5′ primer and HindIII for the 3′ primer) were used to amplify *mycP*\(_5\) from *M. marinum* M strain genomic DNA. These amplicons were cloned as EcoRI-HindIII fragments in EcoRI-HindIII digested pMV361, resulting in pMV-mycP\(_5\). Subsequently, point mutations were introduced in *eccB*\(_5\) and *eccC*\(_5\) by a nested PCR approach using pMV-eccBC\(_5\) as template. In addition, the C-terminal 24 or 1332 amino acids of *eccC*\(_5\) and 363 amino acids of *eccB*\(_5\) were deleted using anchored 3′ primers containing a HindIII restriction site. To create hygromycin versions of the obtained plasmids, the kanamycin resistance cassette was exchanged by a hygromycin resistance cassette by digesting NheI-BcuI. For complementation of the *mas* transposon mutant with a deletion of *eccC*\(_5\), the *mas* gene was amplified from genomic DNA from *M. marinum* M using primers Mas-HindIII-F and Mas-R. This product was cloned into the pMV361 vector using HindIII and HpaI, creating pMV-mas. In addition, the *mas* gene was isolated from pMV-mas by restriction with HindIII and Nhel and ligated into pMV-eccBC\(_5\) or pMV-eccBC\(_5\)-stop, creating pMV-eccBC\(_5\)-mas and pMV-eccBC\(_5\)-stop-mas. For complementation of the *mycP*\(_5\) transposon mutant, pUCintCAT-mycP\(_5\) was created by isolating *mycP*\(_5\) from pMV-mycP\(_5\) using XbaI and Xhol and ligating it into pUCintCAT-empty (Abdallah et al. 2009) using the same enzymes. Finally, a MspA-expression vector was created by amplifying *mspA* from *M. smegmatis* chromosomal DNA using anchored primers (Nhel for the 5′ primer and BamHI for the 3′ primer) and cloning the obtained amplicon as a Nhel-BamHI fragment in pSMT3-LipY (Daleke et al. 2011), digested with the same enzymes. All constructs were checked by sequencing. All used plasmids are listed in Table S2, while all plasmids used in this study are listed in Table S4.

**Generating the eccC\(_5\), mycP\(_5\) and esx-5 region knock-outs in *M. marinum***

An *eccC*\(_5\) and *mycP*\(_5\) knock-out was produced in *M. marinum* M\(^{vu}\) and M\(^{USA}\) respectively, by first creating a merodiploid strain. pMV-eccBC\(_5\) or pMV-mycP\(_5\) was introduced in *M. marinum* strains by electroporation, after which endogenous *eccC*\(_5\) or *mycP*\(_5\), respectively, was deleted by allelic exchange using a specialized transducing mycobacteriophage (Bardarov et al. 1997). For deletion of *eccC*\(_5\), fragments bearing the 1164 and 1186 bp of flanking regions of endogenous *eccC*\(_5\) of *M. marinum*,
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EccC ko Lf</td>
<td>TTTTTTTTCCATAAATTGGCGCTGGACTTCAATCCACG</td>
</tr>
<tr>
<td>EccC ko Lr</td>
<td>TTTTTTTTCCATTTTGGCCGTTGGCAGAACGATGT</td>
</tr>
<tr>
<td>EccC ko Rf</td>
<td>TTTTTTTTCCATTTTGGCCGTTGGCAGAACGATGT</td>
</tr>
<tr>
<td>EccC ko Rr</td>
<td>TTTTTTTTCCATTTTGGCCGTTGGCAGAACGATGT</td>
</tr>
<tr>
<td>MycP5 LF</td>
<td>TTTTTTTTCCAGAAACTGATCTCTCAGCCCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5 LR</td>
<td>TTTTTTTTCCAGAAACTGATCTCTCAGCCCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5 RF</td>
<td>TTTTTTTTCCAGAAACTGATCTCTCAGCCCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5 RR</td>
<td>TTTTTTTTCCAGAAACTGATCTCTCAGCCCTGCTGCAATGT</td>
</tr>
<tr>
<td>EccB5MunI.fw</td>
<td>GATGCAATTCGCTAGGCTGAAACGGGCGGCTCAG</td>
</tr>
<tr>
<td>EccC5HindIII.rev</td>
<td>GATGCAATTCGCTAGGCTGAAACGGGCGGCTCAG</td>
</tr>
<tr>
<td>EccB stop 1</td>
<td>CGCGATTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>EccB stop 2</td>
<td>CGCGATTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>EccC stop 1</td>
<td>CATGAAAGGCGCCTGACCGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>BC Van91F</td>
<td>GACGACCGCAAGGAGTACAA</td>
</tr>
<tr>
<td>BC Van91R</td>
<td>GACGACCGCAAGGAGTACAA</td>
</tr>
<tr>
<td>BC Ncol F</td>
<td>CGAAGAACGCCCGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>BC XmnI R</td>
<td>CTGCGGAAACCGGGGTGGA</td>
</tr>
<tr>
<td>BC EcoRV R</td>
<td>GCCATGGTCCGCGGTCAACTA</td>
</tr>
<tr>
<td>BC K506A F</td>
<td>GTAACGCGCCGCTGCTGCTGCTGCTGCGGTGCA</td>
</tr>
<tr>
<td>BC K506A R</td>
<td>GTAACGCGCCGCTGCTGCTGCTGCTGCGGTGCA</td>
</tr>
<tr>
<td>BC R1181A F</td>
<td>GCCCGGCGCGCAAGGAGTACAA</td>
</tr>
<tr>
<td>BC R1181A R</td>
<td>GCCCGGCGCGCAAGGAGTACAA</td>
</tr>
<tr>
<td>BC R1181K F</td>
<td>GCCCGGCGCGCAAGGAGTACAA</td>
</tr>
<tr>
<td>BC R1181K R</td>
<td>GCCCGGCGCGCAAGGAGTACAA</td>
</tr>
<tr>
<td>BC R1365 stop R</td>
<td>GCCCGGCGCGCAAGGAGTACAA</td>
</tr>
<tr>
<td>Mas HindIII F</td>
<td>CTTAGCTCCTCCTCATGGCGGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MAS R</td>
<td>TGGTGCGAATCGGCTGCTGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>md_MMAR2678_F2</td>
<td>CCGGAATTCATATGCTGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>md_MMAR2678_R</td>
<td>CCGGAATTCATATGCTGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5S461A.fw</td>
<td>GCCGGGACTTATCTCGGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5S461A.rev</td>
<td>GCCGGGACTTATCTCGGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5dTm.rev</td>
<td>GCCGAAGCGCTGCTGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MspA-Nhel-F</td>
<td>GCCGAATTCATATGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MspA-BamHI-R</td>
<td>GCCGAATTCATATGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>pMycO1</td>
<td>GCCCGGCGCGCAAGGAGTACAA</td>
</tr>
<tr>
<td>T7_1</td>
<td>CCGGGAATTCATATGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>T7_2</td>
<td>CCGGGAATTCATATGCTGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>pSalg</td>
<td>GCCGGGACTTATCTCGGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5KIUUpFw</td>
<td>TTTTTTCCATTTTGGCGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5KIUUpRv</td>
<td>TTTTTTCCATTTTGGCGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5KIDownFw</td>
<td>TTTTTTCCATTTTGGCGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
<tr>
<td>MycP5KIDownRv</td>
<td>TTTTTTCCATTTTGGCGCTGCTGCTGCTGCAATGTCCACCCCTAGGCGGCTGCTGCAATGT</td>
</tr>
</tbody>
</table>
Table S2. (continued).

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>TetRFw</td>
<td>TTTTTTGAATTCTAGATGTCTAGATTAGATAAAAG</td>
</tr>
<tr>
<td>TetRRv</td>
<td>TTTTTTAAGCTTAAAGCCACCTTTTTCATTTAAG</td>
</tr>
<tr>
<td>RevTetRFw</td>
<td>TTTTTGAATTCTAGACGACGATCCGGGTCACCATC</td>
</tr>
<tr>
<td>RevTetRRv</td>
<td>TTTTTAAAGCTTAAAGCCACCTTTTTCATTTAAG</td>
</tr>
<tr>
<td>Exs-5 ko Lf</td>
<td>TTTTTTTTCCATAAATTGGATCGCCTCGACACGGAAGGTA</td>
</tr>
<tr>
<td>Exs-5 ko Lr</td>
<td>TTTTTTTTCCATAAATTGGATCGCCTCGACACGGAAGGTA</td>
</tr>
<tr>
<td>Exs-5 ko Rf</td>
<td>TTTTTTTTCCATAAATTGGATCGCCTCGACACGGAAGGTA</td>
</tr>
<tr>
<td>Exs-5 ko Rr</td>
<td>TTTTTTTTCCATAAATTGGATCGCCTCGACACGGAAGGTA</td>
</tr>
</tbody>
</table>

which resulted in a deletion of 96% of the gene, were synthesized by PCR (primer set EccC ko Lf and EccC ko Lr for the 5’ region and primer set EccC ko Rf and EccC ko Rr for the 3’ flanking region). For deletion of mycP\(_5\), the same procedure was performed with primers MycP5 LF, MycP5 LR, MycP5 RF and MycP5 RR. Amplicons corresponding to 5’ or 3’ flanking regions were digested with AlwNI and cloned into the Van91I digested p0004s plasmid that contains a hygromycin resistance cassette and the sacB gene to be able to select for sucrose sensitivity. This allelic exchange substrate was introduced into the Pacl site of plasmid phAE159 and electroporated into \(M.\ \text{smegmatis}\) mc\(^2\)155 to obtain high titers of phage phAE159 according to Bardarov et al. (Bardarov et al. 1997). Subsequently, the \(M.\ \text{marinum}\) strain containing pMV-eccBC\(_5\) or pMV-mycP\(_5\) was incubated with high titers of corresponding phage to create eccC\(_5\) and mycP\(_5\) delinquents. Colonies in which endogenous eccC\(_5\) or mycP\(_5\) was deleted were selected on hygromycin plates and verified for sucrose sensitivity. The deletions were confirmed by PCR analysis and sequencing. Using a temperature sensitive phage encoding the \(\gamma\delta\)-resolvase (TnpR) (a kind gift from Apoorva Bhatt, University of Birmingham, UK), the resistance genes were removed, generating an unmarked deletion mutation. Finally, the pMV-eccBC\(_5\) or pMV-mycP\(_5\) vectors present in the delinquent strains were exchanged by the plasmids containing the various mutated eccB\(_5\), eccC\(_5\) and mycP\(_5\) genes or no ESX-5 gene (empty vector) and the hygromycin resistance cassette by a switching procedure (Pashley and Parish 2003). In addition, pSMT3-mspA, which provides hygromycin resistance, was introduced in the eccC\(_5\) and mycP\(_5\) delinquent strains, after which the complementing pMV-eccBC\(_5\) and pMV-mycP\(_5\) vectors were exchanged with pSM128 (Timm, Lim, and Gicquel 1994) that integrates at the same site as pMV361 and contains a streptomycin resistance cassette. Colonies that showed proper exchange of the plasmids were subsequently used to introduce kanamycin-resistance conveying plasmids with mutated versions of eccBC\(_5\) and mycP\(_5\).

In order to create \(\Delta\text{esx-5::mspA}\) and \(\Delta\text{esx-5::esx-5}_\text{tub}\) another mycobacteriophage was created as described above. For creation of this phage primers esx-5 KO Lf, esx-5 KO Lr, esx-5 KO Rf and esx-5 KO Rr were used to amplify the regions upstream
of eccB and downstream of eccA respectively. These products were cloned, as described above, in a mycobacteriophage vector bearing an apramycin resistance cassette and which lacks the sacB counterselection cassette (a kind gift from Apoorva Bhatt, University of Birmingham, UK). We used this newly constructed phage to infect either M. marinum::mspA or the above described eccC delinquent strain in which pMV-eccBC was first swapped with pMV-esx-5. Deletions were confirmed by PCR and sequencing.

Generation of the conditional BCG-Pasteur c-mycP-tet-on and tet-off mutants.
For establishing regulated expression of the mycP gene, a synthetic gene cassette (hyg-Pmyc1-4XtetO) comprising a hygromycin resistance gene and the Pmyc1 promoter from M. smegmatis engineered to contain four tetO operator sites, which are the DNA binding sites for the cognate repressor protein TetR, was inserted immediately upstream of the mycP start codon in M. bovis BCG-Pasteur (Figure S7A). Targeted gene knock-in was achieved employing the temperature-sensitive mycobacteriophage as described above. For generation of allelic exchange constructs for site-specific insertion in M. bovis BCG-Pasteur of the hyg-Pmyc1-4XtetO cassette, upstream and downstream DNA regions flanking the mycP start codon were amplified by PCR employing the primers listed in Table S1. Subsequently, the upstream and downstream flanks were digested with the indicated restriction enzymes, and ligated with Van91I-digested plasmid arms. The resulting knock-in plasmid was then linearized cloned and packaged into the temperature-sensitive phage phAE159. The resulting mycP knock-in phage was propagated in M. smegmatis and allelic exchange in M. bovis BCG-Pasteur was carried out as described above (see also Figure S7A). The obtained BCG-Pasteur knock-in mutant c-mycP was verified by Southern analysis of MluI digested genomic DNA using a probe as shown in Figure S7B. For achieving controlled gene expression of the target gene mycP, either the E. coli Tn10 tetR gene encoding a repressor protein exhibiting high-binding affinity to tetO sites in absence of the inducer tetracycline (for establishing a mycP tet-on system) or a synthetic gene (rev-tetR) derived from Tn10 tetR encoding a mutated TetR protein with reversed binding affinity to tetO sites upon binding of tetracycline (Klotzsche, Ehrt, and Schnappinger 2009) (for establishing a mycP tet-off system) was heterologously expressed in the knock-in mutant. For this, the tetR gene was amplified by PCR employing the oligonucleotide primer pair TetRFw and TetRRv (Table S1) using an irrelevant tetR-harboring plasmid as a template and cloned using the restriction enzymes EcoRI and HindIII into the episomal E. coli-mycobacterium shuttle plasmid pMV261-RBS-E, which is a derivative of plasmid pMV261 (Stover et al. 1991) harboring a mutated ribosome binding site.

The rev-tetR gene was amplified by PCR employing the oligonucleotide primer pair RevTetRFw and RevTetRRv, using the plasmid pTC-28S15-0X (Addgene plasmid 20316, kindly provided by D. Schnappinger) as a template and cloned using the restriction enzymes EcoRI and HindIII into the episomal shuttle plasmid.
pMV261-RBS-F. The resulting plasmids pMV261::tetR-RBS-E and pMV261::rev-tetR-RBS-F, respectively, providing constitutive gene expression from the HSP60 promoter in mycobacteria were transformed by electroporation into the M. bovis BCG-Pasteur c-mycP₅ knock-in mutant using solid medium containing 50 mg l⁻¹ hygromycin and 20 mg l⁻¹ kanamycin for selection. This yielded the conditional mutant BCG-Pasteur c-mycP₅ pMV261::tetR-RBS-E (referred to as BCG-Pasteur c-mycP₅-tet-on) allowing silencing of the mycP₅ gene in absence of the inducer anhydrotetracycline (ATc) or the conditional mutant BCG-Pasteur c-mycP₅ pMV261::rev-tetR-RBS-F (referred to as BCG-Pasteur c-mycP₅-tet-off) allowing silencing of the mycP₅ gene in presence of ATc. Due to the operonic organization of mycP₅ and the downstream genes eccE₅

Figure S7. Construction of an inducible MycP₅ expression system in M. bovis BCG. Construction of the M. bovis BCG-Pasteur c-mycP₅-Tet-on mutant. A) Organization of the mycP₅ locus in M. bovis BCG-Pasteur wild-type (WT) and the mycP₅ knock-in mutant. The sizes of relevant fragments as well as the location of the probe used for Southern blot analysis are indicated. Hyg, hygromycin resistance gene; Pmyc1-4×tetO, modified Pmyc1 promoter harboring 4 tetO sites. The same method was used to construct c-mycP₅-Tet-off, for which 4 synthetic tet-repressor sites (tetR) were introduced instead of the tetO sites. B) Southern blot analysis of MluI-digested genomic DNA using a probe hybridizing to the position indicated in (A), showing knock-in of the promoter cassette in c-mycP₅-Tet-on.
(Rv1797) and eccA5 (Rv1798) with transcription likely driven by a single promoter, silencing of mycP5 probably also concomitantly downregulates gene expression of eccE5 and eccA5.

Viability determination using Alamar Blue
To measure the viability of the BCG-Pasteur c-mycP5-tet-on and -tet-off mutants in liquid culture, metabolic activity was quantified using Alamar Blue dye (Life Technologies) as a correlate of growth in microtitre plates. Cells of the BCG-Pasteur c-mycP5-tet-on and c-mycP5-tet-off mutant were washed and precultured under mycP5-depleting conditions, i.e. in medium containing no ATc for the tet-on strain and in medium containing 10 µg ml\(^{-1}\) ATc for the tet-off strain. Subsequently, cultures (total volume 100 µl per well in 96-well plates) containing 50 mg l\(^{-1}\) hygromycin, 20 mg l\(^{-1}\) kanamycin and increasing concentrations of ATc (0-20 µg ml\(^{-1}\)) were inoculated 1% (v/v) from the precultures and incubated for 6 days at 37°C. Subsequently, 10% (v/v) Alamar Blue dye solution was added and cells were incubated for a further 16 h at 37°C. Finally, cells were fixed at room temperature for 30 minutes by addition of formalin (5%, v/v, final concentration) and fluorescence was measured using a microplate reader (excitation 560 nm, emission 590 nm).

PERMEABILITY ASSAY
Sensitivity of *M. marinum* towards a combination of ampicillin and clavulanic acid was measured by a disc diffusion assay. Bacterial cultures grown to an OD\(_{600}\) of 1.0 were diluted 10 times in 0.5% agar, which was kept fluid at 37°C. Subsequently, the bacteria-agar suspension was transferred onto 7H10 plates (10 ml per plate) and was let solidify at room temperature. Pills containing 200 μg ampicillin and 25 μg clavulanic acid were placed in the middle of the top-agar and plates were incubated at 30°C until a continuous deck of bacterial growth was observed. The surface of growth inhibition of bacteria was measured in cm\(^2\).

The uptake of ethidium bromide was determined using an adapted protocol as described previously (Danilchanka, Maiaender, and Niederweis 2008). Bacteria were grown to an OD\(_{600}\) of 0.6-1.1. Pellets were collected by centrifugation and resuspended in uptake buffer (50 mM sodium phosphate [pH 7.0], 5 mM magnesium sulfate and 0.05% Tween-80) to an OD\(_{600}\) of 0.55 and pre-energized with 25 mM glucose for 5 minutes at room temperature. A final concentration of 20 µM ethidium bromide was added to each sample containing 27 µl of the bacterial suspension and incubated for 60 minutes at room temperature. Samples were washed with PBS containing 0.05% Tween-80 and subsequently acquired on a BD Accuri C6 flow cytometer (BD biosciences) equipped with a 488 nm laser and 585/40 nm filter. 20,000 gated events were collected per sample and data was analyzed using BD CFlow software.
Transposon mutagenesis, double-filter assay and spotblot analysis

To select for mutations that could rescue the essentiality of eccC₅, a transposon library of the eccC₅ deletion strain containing the complementation plasmid pMV-eccBC₅ was generated using the mycobacterial specific phage phiMycoMarT7 containing the mariner-like transposon Himar1 (Sassetti, Boyd, and Rubin 2003). Subsequently, transposon mutants, in which eccC₅ could be deleted, were selected by exchanging pMV-eccBC₅ with empty vector pSM128 as described above (Pasley and Parish 2003; Timm, Lim, and Gicquel 1994). The removal of pMV-eccBC₅ was confirmed by PCR analysis. To establish the chromosomal location of the transposon insertion, ligation-mediated PCR was used as described by Abdallah et al. (Abdallah et al. 2006). Two mutants were identified with transposon insertions at positions 5266 bp and 4072 bp from the transcription start sites of mas and ppsD, respectively.

Transposon mutagenesis and double filter assays to find PE_PGRS secretion mutants were performed as described in van der Woude et al. (van der Woude et al. 2012). In short, strains M₅VU and M₅USA were transformed with pSMT3-mspA and transposon mutant libraries were created as described above (Sassetti, Boyd, and Rubin 2003). Libraries were plated out on a nitrocellulose filter (Millipore HATF08250) and grown at 30°C on selective 7H10 plates until colonies were visible. A fresh filter was placed between the original filter and the plate and incubated overnight. The second filter was analyzed on PE_PGRS by antibody labeling, visualized by HRP reduction of 4-Chloronaphtol/3, 3-diaminobenzidine (DAB/CNPO). Colonies on the plates were compared to the stained filters and colonies that did not exhibit staining were rechecked in the double-filter assay. Mutants that exhibited no PE_PGRS secretion in both double-filter assays were tested for surface associated PE_PGRS proteins. For this, 20 mg ml⁻¹ wet weight of bacteria was suspended from plate in 0.5% Genapol X-080, vortexed for 1 min and spun down for 5 min at 5000 rpm in a tabletop centrifuge. 2 μl of the detergent supernatant was spotted on a nitrocellulose filter and stained for PE_PGRS proteins as explained above. Only colonies that showed no PE_PGRS proteins on the surface were further analyzed. The transposon integration sites of negative mutants were identified by ligation mediated-PCR as described before (Abdallah et al. 2006; van der Woude et al. 2012).

Lipid analysis of phthiocerol dimycocerosates (PDIMs)

M. tuberculosis and M. marinum strains were cultured to an OD₆₀₀ of 0.7-1.0 and 50 OD-units biomass was collected. The mycobacterial apolar lipids were extracted according the guidelines published by Minnikin and colleagues (Minnikin, Dobson, and Parlett 1985). Concisely, the bacterial biomass was treated with a biphasic mixture of methanolic saline and petroleum ether (PE, 60-80°C) and mixed for 1 hour. The biphasic mixture was separated by centrifuging for 10 min at 2000xg and the upper PE layer was collected. Subsequently, the lower hydrophilic layer was mixed with PE for a second apolar lipid extraction and the PE layers were combined. The PE layer was N₂-gas evaporated and subsequently the apolar lipids were re-dissolved in
250 µl dichloromethane. The PDIM analysis was performed with 2D-TLC as described earlier (Besra 1998). Briefly, equal amounts of the apolar lipids were spotted on silica TLC plates and PDIM lipids were separated by 2D-TLC solvent system A. The first TLC dimension was performed in solvent PE and ethyl-acetate (98:2) and the second dimension in PE and acetone (98:2). Subsequently, the TLC-plates were air dried and the PDIM lipids were visualized by using 5% ethanolic molybdophosphoric acid (MPA) coloring agent and TLC-plate charring at 150°C for 10 minutes.

**Protein secretion and western blot analysis**

Secretion analysis of *M. marinum* was performed as described earlier (Daleke et al. 2011). Briefly, bacterial cultures were grown until mid-logarithmic phase in 7H9 broth supplemented with Middlebrook ADC supplement, 0.05% Tween-80 and appropriate antibiotics. Cells were washed and inoculated in 7H9 without ADC, supplemented with 0.2% dextrose and 0.05% Tween-80 at a starting OD$_{600}$ of 0.35 and incubated overnight, harvested and washed in PBS, while supernatants were filtered through a 0.45 µm filter (Millipore) and TCA precipitated (S). Pellets were resuspended directly in SDS sample buffer (P), or in 100 µl 0.5% Genapol X-080 detergent and incubated for one hour of head-over-head rotation. Cells were spun down and pellets were washed with PBS and resuspended in SDS sample buffer (Genapol pellet (GP)), while 80 µl of the detergent phase was dissolved in 5x concentrated sample buffer (Genapol supernatant (GS)). Western blot of SDS-PAGE gels were stained with polyclonal rabbit sera against EccB$_S$ (E. N. G. Houben et al. 2012), EspG$_S$ (E. N. G. Houben et al. 2012), anti-GroEL2 (kind gift from J. Belisle (Colorado state University and the NIH, Bethesda, MD, USA)) and mouse anti-PE_PGRS (7C4.1F7) (Abdallah et al. 2006). Isolation of cell envelope fractions of *M. marinum*, detergent solubilization of these fractions and subsequent Blue Native PAGE and immunoblot analysis were carried out as described previously (E. N. G. Houben et al. 2012).

**LC-MS/MS**

Cell surface proteins of *M. marinum* strains MVU, MVU::mspA, MVU::Δesx-5::mspA and MVU::Δesx-5::exx-5nub, were isolated using Genapol X-080 essentially as described above. 100 OD-units of PBS-washed bacteria were resuspended in 10 ml 0.5% Genapol X-080 detergent in PBS and incubated one hour at RT with head-over-head rotation. Bacteria were spun down by low speed centrifugation. Supernatants were collected and extracted proteins were concentrated by TCA precipitation. Cell envelope fractions were isolated from another 100 OD-units of PBS-washed bacteria were isolated as previously described (E. N. G. Houben et al. 2012). Samples were analyzed by SDS-PAGE and CBB staining. Total protein lanes were excised in 5 fragments per lane and analyzed by LC-MS/MS (Piersma et al. 2013). Peptides were separated with a 20 cm x 75 µm ID fused silica C18 column (DrMaisch GMBH, Ammerbuch-Entringen, Germany). Peptides were trapped on a 10 mm x 100 µm ID trap column and separated at 300 nl/min in a 8–32% ACN in 0.5% HAc gradient in 60 min (90 min inject-to-
Eluting peptides were ionized at a potential of +2 kV into a Q Exactive mass spectrometer (Thermo Fisher, Bremen, Germany). Intact masses were measured at resolution 70,000 (at m/z 200) in the orbitrap. MS/MS spectra (top-10 precursors) were acquired at resolution 17,500 (at m/z 200). For protein identification, MS/MS spectra were searched against the Uniprot M. marinum complete proteome (ATCC BAA-535M) (downloaded March 2013; 5418 entries) using MaxQuant 1.3.0.5. (Cox and Mann 2008). Additionally, for analysis of the Esx-5<sub>mut</sub> complementation strain, the FASTA file was supplemented with an M. tuberculosis ESX-5 locus FASTA file (Rv1782-Rv1798). Enzyme specificity was set to trypsin and up to two missed cleavages were allowed. Cysteine carboxamidomethylation was treated as fixed modification and methionine oxidation and N-terminal acetylation as variable modifications. Peptide precursor ions were searched with a maximum mass deviation of 6 ppm and fragment ions with a maximum mass deviation of 20 ppm. Peptide and protein identifications were filtered at an FDR of 1% using the decoy database strategy. Proteins were (label-free) quantified by spectral counting (Liu, Sadygov, and Yates 2004; Pham et al. 2010), i.e. the sum of all MS/MS spectra for each identified protein. For quantitative analysis across cell-envelope samples, spectral counts were normalized to the sum of the spectral counts per biological sample. Differential analysis of samples was performed using the beta-binominal test, which takes into account within- and between-sample variations, giving fold-change values and associated p-values for all identified proteins (Pham et al. 2010). Cell-surface associated samples were defined by comparing normalized spectral counts of the Genapol X-080 samples with those of cell-envelope fractions followed by selecting proteins that were at least four-fold enriched in the Genapol X-080 extracted material. This subset of proteins was then normalized for all samples using ten secreted proteins with a relative stable presence in all samples (pckG, mpt64, fbpA, EsxB_1, MMAR_2949, MMAR_1179, MMAR_0722, MMAR_1553, rpiB & MMAR_2047). This step is required because the overall abundance of proteins in the esx-5 mutant samples is highly divergent due to the absence of all PE/PPE proteins.

**TraDIS library preparation and sequencing**

Construction of TraDIS libraries and sequencing were carried out essentially as described previously (Langridge et al. 2009). Briefly, about two micrograms of genomic DNA was sheared to an average size of 300 bp. DNA was purified using QiaQuick PCR purification kit (Qiagen) according to the manufacturer’s recommendations, and subsequently Illumina DNA fragment library preparation was performed using NEBNext® DNA Library Prep Reagent Set for Illumina® (New England BioLabs Inc) following the manufacturer’s instructions. Ligated fragments were run in 2% agarose gel, and fragments corresponding to an insert size of 250–350 bp were excised. DNA was extracted from the gel slice using QiaQuick gel extraction kit (Qiagen). To amplify the transposon insertion sites, 22 cycles of PCR were performed using a transposon-specific forward primer and a custom Illumina reverse primer (see Table 1B).
Amplified libraries were finally purified with AMPure beads (Beckman Coulter) as per the manufacturer’s instructions. A small aliquot (2 µl) was analyzed on Invitrogen Qubit and Agilent Bioanalyzer DNA1000 chip, following the manufacturer’s instructions. The amplified DNA fragment libraries were sequenced on single end Illumina flow cells using an Illumina Genome Analyzer Ilx sequencer for 105 cycles of sequencing, using a custom sequencing primer and 2× hybridization buffer. This primer was designed such that the first 10 bp of each read recognizes transposon sequence. Data were processed with the Illumina Pipeline Software v1.82. The TraDIS reads were analysed using the Bio-TraDIS pipeline (https://github.com/sanger-pathogens/Bio-Tradis). The Bio-Tradis pipeline first filters the reads that match the transposon tags. Filtered reads with transposon tags removed are then mapped to M. marinum M-strain reference genome using SMALT (https://www.sanger.ac.uk/resources/software/smalt) short read mapper. Mapped reads were then sorted using samtools. Insertion sites plots were analyzed by TraDIS essentiality R script to obtain a list of essential genes. All primers used for the TraDIS analysis are listed in Table S3.

**Fatty acid uptake experiments**

Indicated M. marinum strains were grown to mid-logarithmic phase, washed and 0.5 OD units were inoculated in 2 ml screw caps containing 1.5 ml 7H9 medium containing 0.05% Tween-80. To measure fatty acid uptake, 4µg/ml fluorescently labeled fatty acid Bodipy 558/568 C₁₂ (Life Technologies) was added to three independent cultures. Two cultures were incubated without fluorescently labeled fatty acid. All cultures were grown for 72 hours under hypoxic conditions, after which Bodipy-C₁₂ was added to one of the control cultures and incubated one more hour as a negative control. 1 ml of all cultures was washed with PBS, and acquired after gating for similar particle size on a BD Accuri C6 flow cytometer (BD biosciences) as described above. Mean fluorescent intensity was calculated for all samples and adjusted for negative controls. One-way ANOVA was performed to analyze statistical differences between the groups. 0.5 ml of the cultures were additionally washed with PBS and fixed in 4% paraformaldehyde. Bacteria were dried on glass slides and analyzed by confocal microscopy (Leica TCS SP8). Imaging was performed using Leica confocal software with identical settings for each strain.
Table S3. Primers used for TraDIS experiments.

<table>
<thead>
<tr>
<th>Name</th>
<th>Adapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind_Ad_T</td>
<td>ACACTCTTTCCTACACGAGGCTCTTCCGATCT*T</td>
</tr>
<tr>
<td>Ind_AD_B</td>
<td>pGATCCAGAAAGACGGTTTCAGCAGGAATGCCGGAGACCGATCTC</td>
</tr>
</tbody>
</table>

**Primers for qPCR**

<table>
<thead>
<tr>
<th>Name</th>
<th>Adapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>TopA_qPCR</td>
<td>CTTTCCCTACACGAGGCTCTTCC</td>
</tr>
<tr>
<td>BotA_qPCR</td>
<td>ATCCCTGCTGAAACCGCTCTTCC</td>
</tr>
</tbody>
</table>

**qPCR for PCR - amplified libraries**

<table>
<thead>
<tr>
<th>Name</th>
<th>Adapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syb_FP5</td>
<td>ATGATACGGCGACCCACCGATCTCACAC</td>
</tr>
<tr>
<td>Syb_RP7</td>
<td>CAACACAGAAGACCGGCTACAC</td>
</tr>
</tbody>
</table>

**Mycomar P5 primers**

<table>
<thead>
<tr>
<th>Name</th>
<th>Adapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mycomar_P5_</td>
<td>AATGATACGGCGACCCACCGATCTACAC TCTGCGGTAATACGACTC</td>
</tr>
<tr>
<td>PCR3pr1</td>
<td></td>
</tr>
<tr>
<td>Mycomar_3pr_seq</td>
<td>CACTTATAAGGTCTACAGACGGACCGCTTAC</td>
</tr>
<tr>
<td>Mycomar_p5_</td>
<td>AATGATACGGCGACCCACCGATCTACAGAGGCTTACACGAGACCGATCTC</td>
</tr>
<tr>
<td>PCR5pr1</td>
<td>TCCATCGTTGC</td>
</tr>
<tr>
<td>Mycomar_5pr_seq</td>
<td>CTATAGGGATCTAGAGACCGGGGACTTA</td>
</tr>
</tbody>
</table>

**Transposon index read primer**

<table>
<thead>
<tr>
<th>Name</th>
<th>Adapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>RInvSeq</td>
<td>AGATCGGAAGACGCGTCGTTAGGAAAGAGTGT</td>
</tr>
</tbody>
</table>

**Multiplexing primers RInv4**

<table>
<thead>
<tr>
<th>Name</th>
<th>Adapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>RInV 4.1</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RInV 4.2</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RInV 4.3</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RInV 4.4</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RInV 4.5</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RInV 4.6</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RInV 4.7</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RInV 4.8</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RInV 4.9</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RInV 4.10</td>
<td>CAACAGAAGACGCCATAAGAGATCGGTACAAGCTA</td>
</tr>
<tr>
<td></td>
<td>ACACCTCCTTCCCTACACGAGGCTCTTCCGATCT</td>
</tr>
</tbody>
</table>
Table S3. (continued)

<table>
<thead>
<tr>
<th>Name</th>
<th>Multiplexing primers Rlnv4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RlnV 4.11</td>
<td>CAAGCAGAAGACGGCATACGAGATCGGTCATCAAAGT</td>
</tr>
<tr>
<td></td>
<td>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</td>
</tr>
<tr>
<td>RlnV 4.12</td>
<td>CAAGCAGAAGACGGCATACGAGATCGGTCATCAAAGT</td>
</tr>
<tr>
<td></td>
<td>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</td>
</tr>
</tbody>
</table>

**ACKNOWLEDGEMENTS**

We thank M. Daleke, M. Kok, L. van Leeuwen, G. Koningstein, A. Hol Horeman and L. Bockel for technical assistance. We thank J. Knol and T. Pham for data processing and statistical analysis. We would like to thank J. Luirink, C. Vandenbroucke-Grauls, A. van der Sar and B. Appelmelk for valuable discussions.
SUPPORTING INFORMATION

Table S4. List of plasmids used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Features</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMV361-hyg</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(49)</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-kan</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, kan&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-myC&lt;sub&gt;5&lt;/sub&gt;-kan</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, kan&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-hyg</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-myC&lt;sub&gt;5&lt;/sub&gt;-hyg</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-P145 stop</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-V57 stop</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-R1365stop</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-K506A (NBD1)</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-K879A (NBD2)</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-R1181A (NBD3)</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-R1181K (NBD3)</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-mas</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-stop-mas</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-ecBC&lt;sub&gt;5&lt;/sub&gt;-mas</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pMV-exs&lt;sub&gt;5&lt;/sub&gt;-&lt;sub&gt;ub&lt;/sub&gt;</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pSMT3-mpA</td>
<td>hsp60 promoter, pAL5000 origin, oriE, hyg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pUCintCat-myc&lt;sub&gt;5&lt;/sub&gt;</td>
<td>ag85 promoter, L5 attP &amp; aph, oriE, cat&lt;sup&gt;a&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pSM128</td>
<td>hsp60 promoter, L5 attP &amp; aph, oriE, strep&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(51)</td>
</tr>
<tr>
<td>pMV261::tetR-RBS-E</td>
<td>hsp60 promoter, OriM, tetR</td>
<td>This study</td>
</tr>
<tr>
<td>pMV261::rev-tetR-RBS-F</td>
<td>hsp60 promoter, OriM, rev-tetR</td>
<td>This Study</td>
</tr>
</tbody>
</table>

File S1. LC-MS/MS results of cell surface and cell envelope fractions.

File S2. TraDIS results of genes that have more transposon insertions upon expression of mspA.
THE ESX-5 SYSTEM OF PATHOGENIC MYCOBACTERIA IS INVOLVED IN CAPSULE INTEGRITY AND PHAGOSOMAL RUPTURE THROUGH ITS SUBSTRATE PPE10

Louis S Ates¹, Aniek D van der Woude¹,², Jovanka Bestebroer¹, Gunny van Stempvoort², René JP Musters³, Juan J Garcia-Vallejo⁴, Daisy I Picavet⁵, Massimiliano Maletta⁶, Coenraad P Kuijl¹, Nicole N van der Wel⁵, Wilbert Bitter¹,²

Manuscript Submitted.
ABSTRACT

Mycobacteria produce a capsule layer, which consists of glycan-like polysaccharides and a number of specific proteins. In this study, we show that, in slow-growing mycobacteria, the type VII secretion system ESX-5 plays a major role in the integrity and stability of the capsule. We have identified PPE10 as the ESX-5 substrate responsible for this effect. Mutants in esx-5 and ppe10 both have impaired capsule integrity, with reduced amounts of surface localized proteins and glycolipids. Since capsular proteins secreted by the ESX-1 system are important virulence factors, the effect of the capsular defect on virulence mechanisms was assessed. Both esx-5 and ppe10 mutants of Mycobacterium marinum were shown to be impaired in ESX-1-dependent hemolysis and rupture of the phagosomal membrane. The observed defects in capsular integrity resulted in a reduced recruitment of ubiquitin in cellular infections and intermediate attenuation in zebrafish embryos, confirming the importance of capsular integrity in the early stages of infection. These results provide a pivotal role for PPE10, a substrate of ESX-5, in the capsular integrity of pathogenic mycobacteria. These findings open up new roads for research on the mycobacterial capsule and its role in virulence and immune modulation.
Author Summary

Mycobacteria are well protected from effectors of the immune system and from antibiotics by their cell envelope. The recently discovered mycobacterial capsule constitutes the outer layer of this cell envelope. This capsule consists of glucan-like polysaccharides, proteins and glycolipid molecules and is thought to interact with the immune system of the host. In this study, we show that one of the protein secretion systems of slow-growing mycobacteria, called ESX-5, is important in maintaining the structure of this capsule. Furthermore, we identified PPE10, a protein secreted via ESX-5, as the main protein responsible for capsular integrity. We show that disturbed capsule integrity affects the ability of the mycobacteria to rupture the membrane of the phagosome. The bacterial mutants in esx-5 or ppe10 were also attenuated in a zebrafish model for tuberculosis, indicating that the capsular integrity is important for mycobacteria to cause disease. Knowledge on the mycobacterial capsule is only recently emerging and these results could open up new avenues in tuberculosis vaccine or treatment design.

INTRODUCTION

Mycobacteria cause a wide range of diseases in humans, such as tuberculosis, Buruli ulcer and leprosy [1]. Mycobacteria are characterized by their unique mycolic acid-containing outer membrane (MOM). As the name implies, this outer membrane mainly consists mainly of long-chain (C60-C90) fatty acids known as mycolic acids, which are partially covalently linked to the periplasmic peptidoglycan/arabinogalactan layer (Patrick J Brennan and Crick 2007; P J Brennan and Nikaido 1995) and partially linked to trehalose molecules. In addition, this membrane also contains a number of unusual and specific (glyco)lipids. The MOM is extremely impermeable and thereby confers high intrinsic antibiotic resistance and provides protection against many harmful host factors. Although the MOM is in composition very different from the outer membrane of Gram-negative bacteria, electron-microscopy has shown that the form and thickness of the membranes are very similar (Hoffmann et al. 2008; Zuber et al. 2008). Recent EM analysis also showed that there is a capsular layer surrounding the MOM (Sani et al. 2010; Daffé and Etienne 1999). This capsule is loosely attached to the cell-surface and consists of different (lipo)glycans, such as alpha-glucan and lipoarabinomannan (LAM) (Lemassu and Daffé 1994). This structure probably plays an important role in the interaction with the host, although the exact role of the capsule is difficult to determine, as there are no mutants identified yet with a complete loss of the capsule (Sambou et al. 2008; Appelmelk et al. 2008). The capsule of Mycobacterium marinum, has also been shown to contain secreted proteins, mainly the substrates of the type-VII secretion (T7S) systems (Sani et al. 2010). Pathogenic mycobacteria have up to five different T7S systems, named ESX-1 to ESX-5 (Bitter et al. 2009). Substrates of the ESX-1 system are required for the escape of mycobacteria from the host cell phagosome (D. Houben et al. 2012; van der Wel et al. 2007; Simeone et al. 2012).
It is hypothesized that secreted ESX-1 substrates disrupt membrane integrity of the mycobacteria-containing vacuole (Garces et al. 2010; Smith et al. 2008; de Jonge et al. 2007) and thereby allow translocation of the bacteria to the cytosol (van der Wel et al. 2007; D. Houben et al. 2012; Simeone et al. 2012; C. A. Collins et al. 2009). Upon translocation, the mycobacteria are able to proliferate and cause cell necrosis [13,20], allowing cell-to-cell spread of the bacteria. DNA released by cytosolic bacteria is recognized by the cGAS receptor leading to binding of ubiquitin followed by ubiquitin polymerization (Watson et al. 2015; Wassermann et al. 2015; A. C. Collins et al. 2015). This leads to the sequestration of bacteria to autophagic vacuoles (C. A. Collins et al. 2009; A. C. Collins et al. 2015), which is an important process in the protection against mycobacterial infection.

Another T7S system is the ESX-5 system, which is only present in slow-growing species of mycobacteria. This system is involved in the secretion of the majority of PE and PPE proteins (Abdallah et al. 2009; Ates et al. 2015), which are named after a conserved motif in the N-termini (Gey van Pittius et al. 2006). PE proteins consist of a conserved domain of approximately 100 amino acids, which is necessary for secretion. In many cases this PE domain is fused to large C-terminal domains that are not involved in the secretion process. A large group of the pe genes (more than 60 in Mycobacterium tuberculosis and more than 130 in M. marinum) contain polymorphic GC-rich sequences (PGRS), which are coding for glycine-rich repeats that are postulated to play a role in immune evasion. Another group of ESX-5 substrates are the PPE-proteins. These are defined by a conserved N-terminal domain of approximately 180 amino acids, which is also required for secretion. Like the PE proteins, large C-terminal domains can be attached to this PPE domain (Gey van Pittius et al. 2006). Based on the paradigm substrates PE25 and PPE41 and the genomic co-localization of many pe and ppe genes (Gey van Pittius et al. 2006), it has been suggested that PE and PPE proteins are secreted as folded dimers (Korotkova et al. 2014; Strong et al. 2006). Although defined functions for a limited number of ESX-5 substrates have been described (Deb et al. 2006; Daleke et al. 2011; Delogu et al. 2004), most of the PE and PPE proteins have not been studied individually.

In this study, we show that the ESX-5 system is involved in the integrity and stability of the mycobacterial capsule. This effect is dependent on the ESX-5 substrate PPE10. Additionally, impairment of this process by genetic disruption of esx5 or ppe10 results in reduced phagosomal rupture and attenuated virulence in the early stages of infection through a reduction in surface localized ESX-1 substrates.

RESULTS
Identification and characterization of a transposon mutant in ppe10
In an earlier study, set out to identify ESX-5 secretion mutants in M. marinum E11 (van der Woude et al. 2012), ~12,000 transposon mutants were screened in a double-filter assay for defective PE_PGRS secretion. For this screen, we used a monoclonal antibody recognizing part of the PGRS repeat unit and therefore interacting with
many (putatively all) PE_PGRS proteins (Abdallah et al. 2009). The identified secretion
negative mutants were described previously (van der Woude et al. 2012), but we
also identified eighteen transposon mutants that secreted more PE_PGRS proteins
as compared to the wild-type (i.e. supersecretion). Most of the identified transposon
insertions were localized upstream of pe_pgrs-genes and were therefore discarded,
because the supersecretion phenotype was likely due to overexpression of those
specific pe_pgrs genes. However, three of the mutants had a transposon insertion
in the ppe10 gene (mmar_0761), which is not next to a pe_pgrs-gene. PPE10 was
earlier identified to be a substrate of the ESX-5 system (Ates et al. 2015; Abdallah
et al. 2009), without any described function. Repeated double filter analysis of
the M. marinum-ppe10::tn strain confirmed the supersecretion phenotype of PE_-
PGRS proteins. Importantly, this phenotype could be complemented to wild-type
levels when the ppe10 gene was reintroduced using plasmid pSMT3::mmar_0761
(Figures 1A & 1B). In addition to the secretion phenotype, M. marinum-ppe10::tnalso
exhibited an altered colony morphology (Figure 1B), indicating that the cell-wall
composition could be altered. The mutant had a “smooth” colony morphology, similar
to what was found previously in an espG::tn mutant in the same background strain
(Abdallah et al. 2009). This colony morphology was reverted upon complementation
and the complemented strains showed an even more pronounced rough colony
morphology as compared to wild-type M. marinum.

The PE_PGRS supersecretion phenotype of ppe10::tn and complementation of
the mutant was further confirmed by immunoblot analysis of culture filtrate fractions
(CF in Supplemental Figure S1). As a control, we also included the pellet fraction
and a fraction enriched in capsular proteins by treating whole cells with 0.5% of
the mild detergent Genapol X-080, as described previously (Sani et al. 2010).
Typically, a substantial amount of PE_PGRS proteins ends up in this capsule layer of
M. marinum(Ates et al. 2015; Sani et al. 2010). Multiple PE_PGRS proteins were present
in higher amount in the culture filtrate of ppe10::tn, which indicates that this mutation
has a general effect on PE_PGRS proteins. The PE_PGRS supersecretion phenotype
of ppe10::tn was completely reverted upon introduction of pSMT3::mmar_0761,
or the plasmid pSMT3::mmar_0761-HA containing an HA-tagged version of PPE10
(Supplemental Figure S1). As a control for secretion analysis, we included both
the cellular protein GroEL2 and the ESX-1 substrate EspE, which is the main capsule
protein of M. marinum (Sani et al. 2010). The GroEL2 control protein was only seen
in the pellet fraction, as expected. Surprisingly however, the capsular localization
of EspE was completely lost in the ppe10::tn strain. This effect could partially be
complemented by the introduction of pSMT3::mmar_0761 (Supplemental Figure S1).
In conclusion, the ppe10 mutation seemed to affect both ESX-1 and ESX-5 substrates
of the capsule layer.

Previous proteomic analyses by our group indicated that PPE10 is secreted to
the cell-surface of M. marinum in an ESX-5 dependent manner (Ates et al. 2015). To
examine this in more detail we studied the behavior of HA-tagged PPE10 (PPE10-HA)
in a wild-type strain, an ESX-5 secretion mutant (espG::tn) (Abdallah et al. 2009) and the ppe10::tn mutant (Fig. 1C). This analysis showed that a processed form of PPE10-HA of approximately 15 kDa is secreted to the culture filtrate of both wild-type bacteria and ppe10::tn. In contrast, in the ESX-5-deficient strain espG::tn this 15kDa product was not observed in the supernatant and full-length PPE10-HA accumulated in the pellet fractions, which confirms the ESX-5 dependent secretion of this protein. Surprisingly, only relatively low amounts of PPE10-HA were detected in the Genapol X-080 fraction of wild-type bacteria, although previous proteomic analysis did show the presence of substantial amounts of PPE10 in the capsule enriched fraction (Ates et al. 2015). Therefore, perhaps only the C-terminus, containing the HA-tag, is cleaved and shed to the culture filtrate. The C-terminal cleavage of M. marinum PPE10 upon secretion corresponds to what was previously observed for M. tuberculosis PPE10 (Rv0442c) (Abdallah et al. 2009).
Surface localization of capsular proteins is impaired in esx-5 and ppe10 mutants. Since the surface localization of both the ESX-1 substrate EspE and ESX-5 dependent PE_PGRS substrates seemed to be affected by the transposon insertion in ppe10, we investigated the capsule layer of various mutant strains, when grown in the presence or absence of the detergent Tween-80. Detergents such as Tween-80 or Tyloxapol are generally used in the liquid cultures of mycobacteria to avoid bacterial clumping, as in the experiments described above. However, it was recently demonstrated that the presence of these mild detergents also disrupts the capsule layer of mycobacteria and thereby alters the localization of capsular protein and capsular (lipo)polysaccharides (Sani et al. 2010). For instance, upon culturing with detergent, the PE_PGRS proteins are present in increased amounts in the supernatant. On the other hand, the capsular protein EspE is also strongly reduced at the cell surface upon growth in the presence of detergents, but does not accumulate in the supernatant, possibly due to proteolytic degradation (Sani et al. 2010), Figure 2]. When secretion of PE-PGRS proteins of *M. marinum*-ppe10::tn was analyzed by immunoblotting, it was evident that more
PE_PGRS proteins were present in the culture filtrate of the mutant as compared to the wild-type (Figure 2A). However, this PE_PGRS supersecretion phenotype was only observed when ppe10::tn was grown with the detergent Tween-80. In cultures grown without Tween-80, the majority of PE_PGRS proteins remained surface localized, similar to the wild-type strain. These data therefore indicate that ppe10::tn does not produce more PE_PGRS protein, but instead this strain releases more PE_PGRS proteins from its surface, when grown in the presence of detergent. To analyze whether this was a more general phenotype, the localization of EspE was analyzed in the same secretion experiment. Western blot analysis confirmed that EspE is indeed localized in the capsule-enriched (i.e. Genapol supernatant) fraction, when bacteria were grown without Tween-80 (Figure 2B). However, lower amounts of EspE were present in the capsule-enriched fraction of the ppe10::tn and espG::tn mutants as compared to wild-type M. marinum. Similarly, when M. marinum was grown in medium containing Tween-80, lower amounts of EspE were detected in the capsule fraction.
ROLE OF ESX-5 IN MYCOBACTERIAL CAPSULE INTEGRITY

(Figure 2B & Supplemental Figure S1). These data indicate that also the EspE capsular protein is indeed more loosely attached to the cell surface in a ppe10-mutant. We also tested the localization of another ESX-1 substrate, i.e. EsxA (ESAT-6). There was no clear difference in the culture filtrate levels of this protein in the wild-type and the ppe10-mutant strain (Figure 2D).

Capsule integrity is impaired in ESX-5 mutants of pathogenic mycobacteria

To confirm a differential localization of capsular proteins in esx-5 and ppe10 mutant strains, we surface-labeled bacteria with an anti-EspE antibody, which was subsequently visualized using a FITC-labeled secondary antibody, followed by flow cytometric analysis. Wild-type M. marinum E11 showed high levels of EspE surface labeling (Figure 3A), whereas the negative control strain mutated in the ESX-1 secretion system (eccCb1::tn) had only residual staining. As expected, surface labeling of the wild-type cells was reduced when the cells were grown in the presence of Tween-80, although the levels were still significantly higher than the negative control strain. The espG5::tn mutant showed EspE surface labeling similar to the wild-type bacteria when cultured without Tween-80. However, the presence of Tween-80 reduced EspE surface labeling to values only marginally higher than the ESX-1 mutant. The phenotype of the ppe10::tn mutant was even stronger, showing intermediate EspE surface labeling in the absence of Tween-80, while EspE surface labeling was almost completely abolished when Tween-80 was present. Another striking observation in the flow cytometry analysis was that both the espG5::tn and ppe10::tn mutant showed an overall smaller particle size (Supplemental Figure 2A). Calibration beads were used to estimate particle size of analyzed bacteria and showed that a majority of the particles in E11 wild-type cultures grown in the presence of Tween-80 were similar or smaller in size than the 1μm calibration beads (Supplemental Figure S2B), which could represent single-cell particles. However, when E11 was grown in the absence of Tween-80, much larger particles could be observed, ranging in sizes up to 15μm. This confirms the formation of clumps of multiple mycobacterial cells under normal growth conditions without detergent. In contrast, ppe10::tn did not show any bacterial clumping in the presence or absence of Tween-80, indicating that the deficiency in the mycobacterial capsule also affects the clumpy phenotype of M. marinum. This behavior seems to correlate with the smooth colony morphology of this mutant. The ESX-1 mutant showed a more clumpy phenotype as compared to wild-type M. marinum (Supplemental Figure 2A), with approximately 30% of the particles being larger than single cells even in the presence of detergent (Supplemental Figure S2C). This shows that ESX-1 substrates are not responsible for the single-cell phenotype.

Tween-80 dependent surface localization of EspE in the ESX-5 mutant espG5::tn was further confirmed by surface immunolabeling of whole M. marinum cells followed by transmission electron microscopy. First the EspE antibody specificity was determined by analysing M. marinum strain Musa wild-type and M. marinum ΔRD1 (Δmmar_3871-9 (McLaughlin et al. 2007)) (Supplemental Figure S3). As expected [6], EspE surface
Figure 3. Surface labelling of capsular protein EspE and capsular glycans is dependent on ESX-5.
A) Parental strains of wild-type M. marinum E11, or isogenic mutant strains espG₅::tn, eccCb₁::tn and ppe10::tn expressing pSMT3::mCherry, were grown in the presence or absence of Tween-80 and labeled with the α-EspE antibody and a FITC labeled secondary antibody. Subsequently, cells were analyzed by flow cytometry analysis. High levels of surface expression of EspE could be detected in the absence of Tween-80 in both wild-type M. marinum and the espG₅::tn mutant strain, while ppe10::tn showed an intermediate phenotype. When grown in the presence of Tween-80, surface labeling of EspE was almost completely lost and reduced to similar levels of the ESX-1 mutant strain eccCb₅::tn. B) Immuno-electron microscopy analysis of M. tuberculosis CDC1551, the esx-5 mutant Mtb-eccC₅::tn and the mutant complemented with an integrative plasmid containing the complete esx-5 region; pMV-esx-5. Cells were grown in liquid culture with 0.05% Tween-80 (lower row) or without Tween-80 (upper row) and labeled with a monoclonal antibody directed against PIM₆-LAM capsular glycolipids and a gold-labeled secondary antibody. C) Gold labeling of the different Mtb strains depicted in B was quantified by counting the number of gold particles per cell after growth in the presence or absence of Tween-80 (error bars indicate the standard deviation). D) The capsule morphology of wild-type M. marinum and espG₅::tn was analyzed after plunge freezing of bacteria grown in the absence of Tween-80 by cryo-electron microscopy. The black bar indicates the mycobacterial capsular layer.
Figure S2. Mutations affecting capsular integrity influence cellular aggregation of *M. marinum*. 
A) Flow cytometry plots showing forward scatter (FSC – x-axis) and side scatter (SSC – y-axis) of indicated *M. marinum* mutants in the presence or absence of 0.05% Tween-80. Gate P2 was set as a measurement of mycobacterial ‘clumping’, with the percentage of gated cells depicted in red. The presence of Tween-80 reduces clumping in all mutants. However, espG5::tn and ppe10::tn show considerably less clumping compared to the wild-type, while the eccCb::tn shows larger aggregates. B) Size estimation of analyzed particles. Black arrows indicate sizes of calibration beads used as a reference for particle sizes. C) Quantification of gated cells of plots depicted in [A].
labelling of the ESX1 secretion mutant was significantly reduced as compared to the M. marinum wild-type strain (Supplemental Figure S3B). Similar to the results obtained by flow cytometry and our immunoblot analysis, EspE surface labeling of the ESX-5-deficient strain espG<sub>5</sub>:\tn was completely dependent on Tween-80, i.e. in the absence of this detergent surface labeling was similar to wild-type levels, while surface labeling was completely lost in the presence of Tween-80 (Supplemental Figure S3B). This result therefore confirms that the capsule layer of M. marinum is more loosely attached to the cell surface when the ESX-5 secretion system is not functional.

The effect of ESX-5 on capsular integrity was also examined for another mycobacterial pathogen, M. tuberculosis. As M. tuberculosis EspE is not recognized by our antibodies (Sani et al. 2010), we used antibodies directed against glycolipid components of the M. tuberculosis capsule, i.e. an anti-PIM6 antibody (Driessen et al. 2009) and an anti-ManLAM antibody(Kolk et al. 1984). Wild-type M. tuberculosis CDC1551, an isogenic eccC<sub>5</sub>:\tn transposon mutant and the eccC<sub>5</sub>:\tn strain complemented with the complete esx-5 locus (E. N. G. Houben et al. 2012), were grown in the presence of Tween-80 after which a sample was taken and fixed. Subsequently, the strains were grown in 7H9 medium without Tween-80 for six days and both samples were analyzed by immunogold labeling and electron microscopy. As described before (Sani et al. 2010), efficient surface labeling of wild-type M. tuberculosis was observed with these antibodies, which was confirmed in our experiments (Figure 3B, upper frame and Supplemental Figure 3D, upper frame). However, when these strains were grown in the presence of Tween-80, surface labeling was strongly reduced (Figures 3B &3C; Supplemental Figures 3C& 3D). Even in the absence of Tween-80, surface labeling with both antibodies was completely abrogated in the eccC<sub>5</sub>:\tn strain (Figure 3B, middle frame; supplemental figure 3D, second frame). Complementation of eccC<sub>5</sub>:\tn resulted in an increase in surface labeling with anti-PIM<sub>6</sub> compared to the wild-type strain. Together, these data confirm that the ESX-5 system also plays an important role in the capsular integrity of M. tuberculosis.

Next, we analyzed the morphology of the capsule of wild-type M. marinum and espG<sub>5</sub>:\tn by cryo-electron microscopy. The mycobacterial capsule can be visualized as a defined electron-dense layer around the outer membrane (Sani et al.2010). This was clearly visible in wild-type M. marinum grown without detergent (Figure 3D, left frame). A capsular layer was also present in espG<sub>5</sub>:\tn (Figure 3D, right frame), but with a clearly different morphology. The capsule of the espG<sub>5</sub>:\tn mutant was 32±3 nm (measured at 5 sides on 12 bacteria), which is not significantly different from the wild-type (33.27 ± 0.67 (Sani et al. 2010)). However, unlike the defined smooth structure of the M. marinum wild-type(Figure 3D, left panel), the espG<sub>5</sub>:\tn mutant showed ruffled, uneven structures, which could be linked to the instability of the capsular layer under these conditions (Figure 3D, right panel). Together, these data demonstrate that the capsule integrity of pathogenic mycobacteria is dependent on a functional ESX-5 system. Our data obtained in M.
Figure S3. Electron microscopy surface labeling of capsular components. A) Quality and specificity control of the anti-EspE polyclonal antibody. *M. marinum* M45 or the RD1 deletion strain Δmar_3871-9 (McLaughlin et al. 2007) (ΔRD1) were labeled by the anti-EspE serum or by pre-immune rabbit serum followed by a gold-labeled secondary antibody. Specific surface labeling was only observed in the wild-type bacteria labeled with the anti-EspE serum, indicating that this antibody specifically labels EspE. B) Quantification of EspE surface labeling of *M. marinum* strains. EspE surface labeling could be detected in wild-type *M. marinum* irrespective of the presence of Tween-80. EspE surface labeling was reduced to levels of the negative control eccCb5::tn when espG5::tn was grown with Tween-80. C) Quantification of electron microscopy surface labeling of *M. tuberculosis* wild-type strain CDC1551 or an isogenic ESX-5 mutant strain eccC5::tn by an anti-Mannose-capped-lipoarabinomannan (ManLAM) antibody. Surface labeling of ManLAM is reduced in CDC1551 in the presence of Tween-80 and is markedly less in the eccC5::tn strain irrespective of the presence of Tween-80. D) Transmission electron microscopy images of representative images from the dataset depicted in C. The length of the black scale bars represents 500 nm.
*Marinum* suggest that the ESX-5 substrate PPE10 plays a large role in this phenomenon. Different components of the mycobacterial capsule have different susceptibility to resist detergent extraction, but all tested components are more loosely attached in esx-5 and ppe10 mutant strains compared to wild-type bacteria.

**Reduced capsule integrity of esx-5 and ppe10 mutants leads to deficient phagosomal rupture.**

Because the esx-5 and ppe10 mutants have a reduced capsule integrity, we hypothesized that this might affect the virulence of mycobacteria. For instance, reduced capsule integrity results in lower amounts of surface localized ESX-1 substrates, which could affect the membrane-disrupting capacities of these mutants (Kennedy et al. 2014; D. Houben et al. 2012; Gao et al. 2004). Therefore, the membrane-disrupting potential of different strains was tested by quantifying the contact-dependent lysis of erythrocytes in a hemolysis assay (Kennedy et al. 2014; Gao et al. 2004). In line with published data, hemolytic activity was observed for the wild-type strain, independently of the presence of detergent in the culture medium (Figure 4A). As expected, the ESX-1-mutant strain eccCb1::tn did not show hemolytic activity, grown either with or without detergent. In line with our hypothesis, hemolytic activity of the espG5::tn mutant was only observed for cells grown without detergent in the culture medium, while in the presence of Tween-80 hemolytic activity was completely abrogated. The ppe10::tn strain exhibited an even stronger phenotype and showed hardly any hemolytic activity when grown with or without Tween-80. This effect was reversed upon complementation. These experiments show that there is a clear correlation between the PPE10-dependent surface localization of ESX-1 substrates and the hemolytic potential of the mycobacterial strains.

Next, we assessed the potential of these mutants to rupture the phagosomal membrane. Different *M. marinum* strains expressing pSMT3::mCherry (Meijer et al. 2008) were cultured overnight in the presence or absence of Tween-80 and subsequently used to infect differentiated THP-1 cells. After two hours of infection and one hour of incubation, the cells were fixed and stained with the FK2 antibody, which recognizes mono- and poly-ubiquitinylated proteins but not free ubiquitin. As ubiquitin is known to accumulate around cytosolic bacteria (D. Houben et al. 2012), ubiquitinylation can be used as a marker for cytosolic contact of bacteria (C. A. Collins et al. 2009; Watson, Manzanillo, and Cox 2012; Lerena and Colombo 2011). Co-localization of bound ubiquitin and fluorescent mycobacteria was quantified using imaging flow cytometry (complete data in Supplemental Figure S4). For this, first the infected cells were selected and subsequently the relative co-localization of ubiquitin with the intracellular bacteria was determined. When wild-type bacteria were grown without Tween-80, 33.3% of the detected bacteria co-localized with ubiquitin, while the amount of co-localization was only 9.2% for the negative control strain eccCb1::tn (Figure 4B). An intermediate phenotype was seen for espG5::tn (20.3%), while ppe10::tn (7.15%) showed comparable levels to the negative control. When
Both the \textit{esx-5} and \textit{ppe10} mutants have reduced hemolytic activity and show reduced phagosomal rupture upon infection of host cells. A) Contact-dependent hemolysis of red blood cells (RBCs) by \textit{M. marinum}. \textit{M. marinum} E11 wild-type cells and isogenic transposon mutants disrupted in \textit{espG}, \textit{eccCb}, and \textit{ppe10} and the complemented \textit{ppe10} mutant were grown in the presence or absence of Tween-80. Subsequently, washed cells were used for the hemolysis assay. B) Quantification of differentiated THP-1 cells that were infected with \textit{M. marinum}. Only cells gated as positive for co-localization of the bacteria with ubiquitin were chosen for further analysis (full data in Supplemental Figure S4). The \textit{ppe10::tn} mutant showed similar levels of ubiquitin co-localization as \textit{eccCb::tn} indicating that this mutant cannot escape the phagosome in the early stages of infection. C) Images obtained by imaging flow cytometry of wild-type \textit{M. marinum} and \textit{ppe10::tn} three hours after infection of differentiated THP-1 cells. Bacteria express \textit{mCherry} (Red), while ubiquitin is visualized by FK-2 antibody against poly-ubiquitin (Green). D) Super-resolution confocal microscopy images of wild-type \textit{M. marinum} and \textit{ppe10::tn} illustrating differential bacterial clustering and ubiquitin labeling of bacteria. Bacteria express \textit{mCherry} (Red), while ubiquitin is visualized by FK-2 antibody against poly-ubiquitin (Green). A 3-dimensional view of these images can be found in supplementary Movies S1 and S2.
Figure S4. Analysis of co-localization of ubiquitin with bacteria by imaging flow cytometry. THP-1 macrophages infected with the indicated strain of *M. marinum* expressing mCherry were stained with the FK2 antibody recognizing poly-ubiquitin and were analyzed by imaging flow cytometry. Bacteria were pre-cultured in the presence (Red lines) or absence (Blue lines) of Tween-80. Relative co-localization of green and red fluorescence was quantified per particle (X-axis). Cells within gate R1 (green line) were seen as positive for co-localization of ubiquitin and bacteria for further analyses. Data of two independent experiments were pooled and analyzed together.

the strains were pre-cultured with Tween-80, reduced co-localization of ubiquitin and bacteria was observed for wild-type *M. marinum* (22.7%) and espG₂::tn (15.9%). Values for eccCb₁::tn (10.7%) and ppe10::tn (9.19%) remained similar, indicating that these values are likely the background level of this assay. Representative images of these analyses are shown in Figure 4C, where it is illustrated how co-localization of bacteria and ubiquitin is indeed reduced in the ppe10::tn mutant strain (Figure 4C). High resolution confocal microscopy was used to illustrate this phenomenon in more detail for the wild-type *M. marinum* and ppe10::tn strains.
pre-cultured without Tween-80 (Figure 4D). Surface labeling of bacteria by FK2 could be observed on many clusters of wild-type \textit{M. marinum} (Figure 4D, upper panel; Supplemental Movie S1), but was generally not observed for \textit{ppe10::tn} cells (Figure 4D, lower panel; Supplemental Movie S2).

Another observation from both imaging flow cytometry and confocal microscopy analyses was that \textit{ppe10::tn} bacteria are almost exclusively seen as single bacteria in infected macrophages whereas wild-type \textit{M. marinum} was often seen in clusters (Figure 4D, Supplemental Figure S5C, Supplemental Movies S1 & S2), which is in line with our flow cytometry results. One potential confounding factor was the trend of increased ubiquitin co-localization with larger cluster sizes for wild-type \textit{M. marinum} (Supplemental Figure S5A). However, the \textit{eccCb1::tn} mutant showed decreased ubiquitin recruitment, despite having the largest cluster sizes, suggesting cluster size is not the primary cause of ubiquitin recruitment. To exclude the possibility that the amount of bacteria per cell is a cause of ubiquitin recruitment, we also assessed the mean fluorescent intensity of FK2 co-localization of cells infected with one bacterium (Supplemental Figures S5A & S5B). THP-1 macrophages infected by a single \textit{M. marinum} bacterium showed the same trends of phagosomal rupture as observed in the whole population (Supplemental Figure S5B), suggesting that the observed phenotypes are causes by differences in capsule integrity and not differences in bacterial cluster size (Figure S5C). Together, these data show a clear effect on the virulence of \textit{M. marinum} in the early stages of infection through a defect in phagosomal escape in \textit{ppe10::tn} and to a lesser extend in \textit{espG5::tn}. These data correlate very well with the amount of detected capsular ESX-1 substrates and with the previously observed delay in phagosomal escape of the ESX-5 mutant as observed by electron microscopic analysis (Abdallah et al. 2011).

**Zebrafish experiments**

A previous report described the temporary attenuation of the \textit{espG5::tn}mutant in zebrafish embryos (Weerdenburg et al. 2012). We set out to test whether this phenotype of delayed infection was due to reduced capsule integrity and defective secretion of PPE10. Zebrafish embryos were infected with wild-type \textit{M. marinum}, \textit{espG5::tn}, \textit{eccCb1::tn}, \textit{ppe10::tn} and \textit{ppe10::tn-C} and 5 days after infection embryos were homogenized and colony forming units (CFUs) were quantified (Figure 5A). As expected, the ESX-1 mutant strain \textit{eccCb1::tn} showed strong attenuation, which was observed as an approximate 100-fold reduction in recovered CFUs after five days. As previously described, the \textit{espG5::tn} strain exhibited a significant reduction in recovered CFUs (log transformed mean difference compared to wild-type 0.6649, p<0.001). The \textit{ppe10::tn} mutant was attenuated to a similar extend as \textit{espG5::tn} (log transformed mean difference compared to wild-type 0.8275, P<0.001). The difference between \textit{espG5::tn} and \textit{ppe10::tn} was non-significant, while the complemented \textit{ppe10::tn} strain was comparable to wild-type \textit{M. marinum}. Although both the \textit{espG5} and \textit{ppe10} mutants were attenuated, they showed
Figure S5. Co-localization of ubiquitin and bacteria is influenced by bacterial cluster size, but also by differences in capsular ESX-1 substrates. THP-1 macrophages were infected by the indicated strains of *M. marinum* expressing mCherry and were pre-cultured in the presence or absence of Tween-80. Infected cells were stained with the FK2 antibody and a FITC-labeled secondary antibody. A) Cells were analyzed by imaging flow cytometry and were sorted for the amount of bacteria per cell (Y-axis) and the intensity of FK2 staining (X-axis). Color coding indicates the density of analyzed particles with indicated fluorescent intensities. The blue rectangle indicates the gate used to analyze macrophages infected by a single bacterium. B) Analysis of the mean fluorescent intensity (MFI) of FK-2 staining measured by the FITC signal (Y-axis) of gated cells that contain only a single bacterium. Relative MFI data correlated very well with data obtained in other assays and analyses (Figure 4A, 4B) indicating that the number of bacteria per macrophage is not the driving force behind the observed phenotype of phagosomal rupture. C) Average number of bacteria per infected macrophage. The *ppe10::tn* strain had markedly fewer bacteria per macrophage compared to wild-type *M. marinum* irrespective of the presence of detergent. The amount of bacteria per macrophage was highly dependent on the presence of Tween-80 for the *espGΔ::tn* strain.
Figure 5 The *M. marinum* ppe10::tn mutant is attenuated in zebrafish embryo infections. A) *M. marinum* E11 wild type strain and the isogenic mutants espG::tn, eccCb::tn and ppe10::tn, as well as the complemented ppe10::tn strain (PPE10-C), were pre-cultured in liquid medium containing Tween-80. 50-100 CFUs of bacteria were injected in the bloodstream of zebrafish embryos at 28 hours post fertilization. Embryos were homogenized five days post infection and plated to establish the number of CFU per embryo. Three independent experiments of six embryos per group were performed and the data were pooled. B) Visualization of *M. marinum* infection of zebrafish embryos. Wild-type *M. marinum* or ppe10::tn containing the plasmid pSMT3::mCherry was injected in the bloodstream of zebrafish embryos as described above. After 5 days of infection the embryos were examined by brightfield or fluorescence microscopy.

**DISCUSSION**

In this study, we demonstrate that the ESX-5 secretion system plays a major role in capsular integrity of pathogenic mycobacteria and that the ESX-5 substrate PPE10 seems to be a key player in this phenomenon. The existence of the mycobacterial capsule has already been postulated a long time ago, but remained controversial.
(Daffé and Etienne 1999; CHAPMAN, HANKS, and WALLACE 1959) until the recent visualization of intact, plunge-frozen mycobacteria by cryo-EM (Sani et al. 2010). These experiments showed that detection of the capsule was only possible when the bacteria were grown without detergent, indicating that this capsule is a labile structure, prone to disruption (Sani et al. 2010). Moreover, different components of the mycobacterial capsule have different susceptibilities to the presence of detergent [6, this study] or for extraction by chemical methods (Ortalo-Magné et al. 1996). For instance, whereas EspE, the major capsular protein component of *M. marinum*, is only present on the cell surface in high amounts when cells are grown without Tween-80, PE_PGRS proteins can be identified in the capsule-enriched fraction in both conditions. These data correspond with earlier reports (Ortalo-Magné et al. 1996; Sani et al. 2010) suggesting that the mycobacterial capsule might consist of several layers with different susceptibilities for disruption. In addition, there is also a clear difference in capsule stability and composition between different mycobacterial species [6]. Here we show that most capsular components of *M. marinum* that were examined were more susceptible to disruption by detergents in strains defective in ESX-5 secretion or in the ESX-5 substrate PPE10. This is confirmed by cryo-EM analysis, which showed that the esx-5 mutant has a ruffled morphology of the capsule. This suggests that PPE10 and the ESX-5 system are not directly involved in the biogenesis of the capsule, but play a direct or indirect role in the anchoring of the capsule. Our data also show that capsular defects result in reduced hemolytic activity, a delay in rupturing of the phagosomal membrane and in vivo attenuation in zebrafish embryos. Throughout these experiments, the ppe10 mutant showed more pronounced phenotypes than the ESX-5 mutant. A possible explanation for this could be that some residual secretion of native PPE10 might occur in the ESX-5 mutant that was used (espG<sub>5</sub>::tn), as this mutant is strongly reduced but not completely devoid of secretion (Abdallah et al. 2009; Daleke, Woude, et al. 2012). Please note that a completely negative ESX-5 secretion mutant is lethal, as we and others have recently reported for different mycobacterial strains/species (Ates et al. 2015; Di Luca et al. 2012).

The ESX-1-dependent membrane disruption and phagosomal escape of pathogenic mycobacteria has been extensively studied in recent years (D. Houben et al. 2012; van der Wel et al. 2007; Simeone et al. 2012; Wassermann et al. 2015; Simeone et al. 2015). The importance of the ESX-1 system in this process could be confirmed using imaging flow cytometry, which measured the co-localization of ubiquitin with intracellular *M. marinum*. This new procedure allows a rapid quantitative approach to study access of mycobacteria to the host cytosol. Interestingly, the co-localization of *M. marinum* with ubiquitin correlated very well with surface exposed EspE, as detected by our various extraction and surface labeling methods. Although EspE is the most abundant capsular protein in *M. marinum*, it is not known whether this ESX-1 substrate is responsible for the observed phenotypes in *in vivo* experiments. Other studies have implicated EsxA as the membrane disrupting factor of the ESX-1 secretion system (De Leon et al. 2012; Ma, Keil, and Sun 2015). In our experiments,
surface localization of EsxA did not correlate with membrane disrupting potential of the strains, indicating that possibly other ESX-1 secreted substrates are involved in this phenotype.

Although our data show that capsular defects result in attenuation of M. marinum through the differentially localized ESX-1 substrates, there could also be other effects on virulence. For instance, the capsular polysaccharide α-glucan (Geurtsen et al. 2009) and the glycolipid ManLAM (Geijtenbeek et al. 2003) were previously shown to interact with pathogen recognition receptors and have immune-modulating properties (Geurtsen et al. 2009; Geijtenbeek et al. 2003). One study identified a ppe10 transposon mutant of M. bovis BCG in a screen for mutants that were defective in avoiding phagosome acidification (Stewart et al. 2005). This phenotype led to an attenuation of BCG-ppe10::tn compared to wild-type BCG after 4 and 6 days of infection. Since BCG does not possess a functional ESX-1 secretion system, these results show that this defect in the arrest of phagosome acidification is not coupled to ESX-1 substrates. Another interesting observation is the effect we observed of capsule integrity on cell clumping. Both the espG5::tn and ppe10::tn mutant showed a single cell phenotype, indicating that the capsule is an important component in cell clumping. Mycobacterial clumping has also been associated with trehalose dimycolate (cord factor) [Reviewed in 52] and a defect in lipooligosaccharide (LOS) biosynthesis (van der Woude et al. 2012). Possibly the ESX-5 system and/or PPE10 play a role in the biogenesis of extracellular (glyco)lipids, leading to the observed phenotypes on capsular morphology and cell aggregation.

What is the molecular mechanism for the role of PPE10 in capsular integrity? As discussed above, PPE10 could have a role in the biogenesis of a lipid or carbohydrate moiety important for the integrity of the capsule. Alternatively, PPE10 could be a structural component of the capsule by binding both the cell-surface and other capsular components. PPE10 is detected on the cell-surface of M. marinum in relatively high amounts by mass-spectrometry (Ates et al. 2015), even though sequence coverage of the protein is relatively poor due to the paucity of trypsin cleavage sites. This indicates that the protein is probably present in high amounts on the bacteria. PPE10 is also expressed in vivo, as it is detected in the lungs of infected guinea pigs at both 30 and 90 days post infection (Kruh et al. 2010), and expression levels seem to be relatively constant at different time points in a mouse infection model (Soldini et al. 2011). There are close orthologues of ppe10 present in other pathogenic mycobacteria such as M. tuberculosis, M. ulcerans, and M. bovis (Gey van Pittius et al. 2006), which likely have the same function in these species. PPE10 orthologues of pathogenic mycobacteria contain pentapeptide repeat domains (pfam01469 (Finn et al. 2014)) and are therefore grouped in the PPE subfamily of major polymorphic tandem repeat (MPTR) proteins (Hermans, van Soolingen, and van Embden 1992; Cole et al. 1998), of which PPE10 is thought to be the most ancient (Gey van Pittius et al. 2006). Unfortunately, no role for PPE-MPTR proteins has been described. The strong phenotype of the ppe10::tn strain suggests that there is no
redundancy in the function of PPE10 and other PPE-MPTR proteins of *M. marinum*. Therefore, it is possible that the MPTR domains are only structural domains not directly involved in protein function. It is intriguing that PPE10, which is only present in slow-growing pathogenic mycobacteria, is so important for capsular integrity, since fast-growing mycobacteria also contain a capsule (Sani et al. 2010). Other pentapeptide repeat containing proteins of fast-growing mycobacteria could possibly perform a similar function. For instance the MYCSM_05776 protein of *M. smegmatis* strain JS623 contains a penta-peptide repeat with 40-50% identity compared to PPE10. However, this protein contains a classical signal sequence instead of a PPE domain. A similar situation is observed for LipY, which is a PE or PPE protein in slow-growing mycobacteria but a signal-sequence containing protein in fast-growing mycobacteria (Daleke et al. 2011).

Finally, our data illustrate that the different phenotypes attributed to the different T7S systems of pathogenic mycobacteria can be interdependent. We have shown that reduced amounts of surface localized ESX-1 substrates in esx-5 and ppe10 mutants lead to a delayed phagosomal rupture in cell-culture experiments and attenuation in zebrafish embryos. These data help to elucidate previously-described but unexplained results (Abdallah et al. 2009; Weerdenburg et al. 2012; Stewart et al. 2005). These new results open the way to study the role of the mycobacterial capsule in virulence by new approaches and might uncover multiple other effects on virulence that are suggested by literature (Stewart et al. 2005; Geurtsen et al. 2009; Geijtenbeek et al. 2003). The surface localization of PPE10, combined with the multiple possible effects on mycobacterial pathogenesis of this protein, make this an interesting candidate for further analysis to unravel virulence mechanisms of mycobacterial pathogens.

Supplemental Movies S1 and S2
The *ppe10::tn* mutant has reduced co-localization with ubiquitin and a single cell phenotype in cell-infection. Wild type *M. marinum* E11 (Supplemental Movie S1) or the *ppe10::tn* mutant (Supplemental Movie S2) expressing mCherry (Red) were used to infect THP-1 cells and stained with the FK-2 antibody (Green) as described in Figure 4D. Images were obtained by STED super-resolution confocal microscopy and Z-stacks were combined using Huygens Professional Software (SVI). Movies show a three-dimensional view of the analyzed bacterial clusters and a reduction and restoration of the threshold of green fluorescence to visualize the levels of FK-2 labelling.

MATERIALS AND METHODS
Strains and growth conditions
All *M. marinum* strains that were used were derived from the wild-type strain E11 (van der Sar et al. 2004; Weerdenburg et al. 2015), except for the quality control of the anti-EspE antibody for which MUSA and its ΔRD1 mutant were used. The ESX-5
secretion mutant espG<sub>5</sub>:tn is previously described as transposon mutant 7C1 and has a transposon insertion in mmar_2676 (espG<sub>5</sub>) (Abdallah et al. 2009). The mutant eccC<sub>5</sub>:tn was identified previously as an ESX-1 mutant (Stoop et al. 2011). Strain CDC1551 and the isogenic transposon mutant in eccC<sub>5</sub> (JHU1783-2086) were described previously (E. N. G. Houben et al. 2012) and were kindly provided by BEI-resources (Lamichhane et al. 2003). All strains were cultured on Middlebrook 7H10 plates or in Middlebrook 7H9 medium (Difco) containing ADC supplement and, when required, 0.05% Tween-80 and the appropriate antibiotic selection markers (25 μg/ml kanamycin (Sigma) and/or 50 μg/ml hygromycin (Roche)). <i>M. marinum</i> was incubated at 30 °C whereas <i>M. tuberculosis</i> strains were cultured at 37°C.

**Transposon mutagenesis**
Transposon mutagenesis and secretion analysis by double filter assay was performed as described by van der Woude et al. (van der Woude et al. 2012). Briefly, <i>M. marinum</i> strain E11 was infected with the mycobacteriophage φMycomarT7, containing the mariner-like transposon Himar1 (Sassetti, Boyd, and Rubin 2003). The transposon library was plated on 7H10 plates covered by a nitrocellulose filter. After colonies were formed, a second filter was placed underneath and was incubated overnight. The second filter was stained by an anti-PE_PGRS antibody (7C4.1F7) (Abdallah et al. 2009) and colonies were screened for increased secretion. Positive colonies were rescreened on a second double-filter assay, before the localization of the transposon insertion was determined using ligation-mediated PCR (Abdallah et al. 2006; Prod’hom et al. 1998).

**Construction of plasmids**
The mmar_0761 (ppe10) gene was amplified from <i>M. marinum</i> E11 genomic DNA using primers LA_0761-fw (TTTGCTAGCGTGCAAACCCGCATT) containing a NheI restriction site and LA_0761-rv (AAAGATATCAGCATAATCAGGAACATCATACGGATACTACTCCGTGCAGCAGCGCA) containing an EcoRV restriction site and a stop codon before the HA sequence. For the construct with an HA-tag, reverse primer LA_0761-HA-Rv (TGCCGCTGCGACGGAGTATCCGTATGAGTTTCCGTATTGCTTGGATATC) was used. PCR products were cloned into the pSMT3 vector (Daleke et al. 2011) by replacing lipYtub with the indicated constructs after restriction with EcoRV and NheI. All constructs were checked by sequence analysis.

**Immunoblot and secretion analysis**
Bacteria were grown until mid-logarithmic phase and were washed and cultured overnight in 7H9 medium containing glycerol and dextrose, but without bovine serum albumin (BSA) (Sigma). Depending on the experimental conditions, 0.05% Tween-80 was added to the cultures. After overnight culture, supernatants were filtered using a 0.02 μm filter, concentrated by TCA precipitation and washed with acetone.
Bacteria were harvested and washed once with PBS. Aliquots were taken for whole cell lysates and for Genapol X-080 extraction of capsular proteins. Genapol extraction was performed by incubating bacteria 30 minutes in 0.05% Genapol X-080 in PBS under head-over-head rotation at room temperature. Genapol X-080 extractable fractions were diluted in 5X concentrated solubilization/denaturation (S/D) buffer. Whole cell lysates and Genapol X-080 extracted cell residues were solubilized in S/D buffer and sonicated to lyse the cells. All samples were boiled for 10 minutes at 95 °C before loading on SDS-PAGE. After SDS-PAGE, proteins were transferred to nitrocellulose membranes by western blot, which were subsequently incubated with different antibodies followed by electrochemiluminescence. Primary antibodies used were: anti-PE_PGRS antibody (7C4.1F7) (Abdallah et al. 2009); anti-GroEL2 (CS44, Colorado state university); anti-EsxA (Hyb76-8) (Harboe et al. 1998); anti-influenza hemagglutinin (HA) epitope (HA.11, Covance) and polyclonal anti-EspE (This study, produced as described in (Carlsson et al. 2009)). Secondary goat-anti-mouse (American Qualex) or goat-anti-rabbit (Rockland) horseradish peroxidase labeled antibodies were used and visualized by enhanced chemiluminescence (ECL prime, Amersham).

**Bacterial flow cytometry analysis and surface labeling**

*M. marinum* strains expressing pSMT3::mCherry (Meijer et al. 2008), were pre-cultured in 7H9 medium with Tween-80 and incubated overnight in the presence or absence of Tween-80. Cells were harvested by centrifugation and were incubated with an anti-EspE polyclonal antibody (Carlsson et al. 2009), followed by a FITC-labeled secondary goat anti-rabbit antibody (BD bioscience). Bacterial cells were acquired on a BD AccuriC6 flow cytometer (BD biosciences). Particle size estimation was performed by comparing it to beads from the flow cytometry size calibration kit (Life Technologies). 20,000 red fluorescent (610/20 nm) events were acquired in a gate on side scatter and forward scatter that corresponded to particle sizes of single bacteria. Green fluorescence (530/30 nm) of the gated cells was quantified and plotted.

**Electron microscopy**

For immunogold electron microscopy, *M. marinum* strains were inoculated from an exponentially growing culture to 7H9 medium with or without detergent at 0.35 OD$_{600}$/ml and incubated without agitation. After overnight incubation, 5 OD-units of bacteria were collected by centrifugation and fixed for 24 hours in 0.2M PHEM (final concentrations: 120mM PIPES, 50mM HEPES, 4mM MgCl$_2$ 20mM EGTA) buffer with 4% paraformaldehyde and 0.4% glutaraldehyde. *M. tuberculosis* was grown until mid-logarithmic phase in the presence of 0.05% Tween-80. At this point, samples were collected and fixed as described above, while cells were also re-inoculated in a medium lacking Tween-80 and incubated for 6 days after which samples were collected and fixed. Fixed bacteria were incubated on carbon coated formvar grids
for 5 minutes. Grids were stained with different antibodies: anti-EspE polyclonal rabbit serum (produced as described in (Carlsson et al. 2009), quality control in Supplemental Figure S3); anti-PIM\(_6\) (F183-24) (Driessen et al. 2009; Sani et al. 2010); anti-ManLAM (55.92.1A1) (Appelmelk et al. 2008; Sani et al. 2010). Gold-labelled secondary antibodies (Utrecht University) were used and surface labelling was visualized on a Tecnai 12 electron microscope (FEI, Eindhoven, the Netherlands). Plunge freezing and Cryo-EM was performed as described in Sani et al. 2010 (Sani et al. 2010).

Hemolysis

Hemolysis experiments were performed mostly as described by Smith et al. [17]. Shortly, M. marinum strains were grown in a liquid pre-culture in 7H9 medium supplemented with 0.05% Tween-80 until mid-logarithmic phase was reached. All strains were washed once in 7H9 medium without Tween-80 and inoculated in 7H9 with- or without Tween-80 at 0.35 OD\(_{600}\)/ml. After overnight culture, bacteria were collected by centrifugation, diluted in PBS, and quantified by absorbance measurement at OD\(_{600}\). It was previously estimated that 1 ml of 1.0 OD\(_{600}\) of bacteria contain 2 * 10\(^8\) bacteria. In parallel, defibrinated sheep erythrocytes (Oxoid – Thermo Fisher, the Netherlands) were washed 5 times and diluted in fresh DMEM medium without phenol red (Gibco, Life technologies). Erythrocytes were counted using a Bürker-Türk hemocytometer and 8.3*10\(^8\)/ml erythrocytes and 2.08*10\(^9\)/ml bacteria were added in a round-bottom 96 well-plate and gently centrifuged for 7 min. at 600 RPM in a swing out-plate centrifuge and incubated at 30°C for 3 hours. After incubation, cells were brought into suspension and centrifuged. Supernatant was transferred to a flat-bottom 96-wells plate and absorbance at 405nm was measured to quantify hemoglobin release.

Cell infections and ubiquitin co-localization

M. marinum E11 and transposon mutants eccCb\(_1\)::tn, espG\(_5\)::tn and ppe10::tn strains expressing pSMT3::mCherry (Meijer et al. 2008), were cultured overnight with or without 0.05% Tween-80 as described above. THP-1 cells were maintained in RPMI 1640 GlutaMAX supplemented with 10% FCS at 37°C and 5% CO\(_2\). Monocyte differentiation was started 5 days prior to infection by incubation with 10ng/ml PMA for 48 hours. For infection studies, THP-1 cells were seeded in 6 well plates at a density of 2*10\(^6\) cells per well. After differentiation cells were infected at an MOI of 10 for 2 hours. Subsequently, cells were washed 3 times with warm PBS and incubated for an additional hour in complete RPMI media. To analyze ubiquitin recruitment to the bacteria, cells were trypsinized and fixed in 2% PFA for 15 minutes. All staining procedures of the fixed THP-1 cells occurred in the presence of 0.1% Saponin. Cells were blocked for 45 minutes with 1% BSA and stained for ubiquitin with the FK2 antibody (Millipore, 1:200 dilution) in the presence of 1% BSA for 1 hour. Cells were subsequently washed twice and labeled with a secondary antibody Goat anti Mouse
Alexa488 (Life technologies, 1:200 dilution) for 30 minutes followed by 2 washes with PBS. Cells were post-fixed with 2% PFA before analysis.

**Imaging flow cytometry**

Cells were analyzed on the ImageStreamX100 (Amnis-Merck Millipore) imaging flow cytometer as previously described (J J García-Vallejo et al. 2014; Juan J García-Vallejo et al. 2015). A minimum of 15,000 cells were acquired per sample. For standard acquisition, the 488 nm laser line (for FK2) was set at 100 mW and the 561 nm laser line (for mCherry) was set at 200mW. Single staining controls were used to generate compensation files that allowed correction of spectral overlap. Co-localization scores were calculated as previously described (J J García-Vallejo et al. 2014) using the bright detail similarity R3 feature in the Ideas software. This feature corresponds to the logarithmic transformation of Pearson’s correlation coefficient of the localized bright spots with a radius of 3 pixels or less within the whole cell area in the two input images. Bacterial counts were calculated using the peak mask in combination with the spot count feature as previously described (Juan J García-Vallejo et al. 2015).

**STED super-resolution microscopy**

Stimulated Emission Depletion microscopy (STED) was performed on a Leica TCS SP8 STED 3X microscope, Leica Microsystems, (Wetzlar, Germany). Samples were irradiated with a pulsed white light laser at wavelengths 488 nm and 587 nm with a power of 20%, which translates to 84mW. A continuous wave STED laser line at a wavelength of 660nm and 210mW power (20%), Leica Microsystems, was used for depletion of the 488nm fluorophore, reaching a lateral resolution of ~70nm. The signal was detected using a gated Hybrid Detector (HyD), Leica Microsystems, with a gain of 120% and a range of 503 – 569 nm for the green channel. The red channel was set at a gain of 142% and a range of 603 – 651 nm. STED images were acquired using a dedicated oil objective with 100x magnification and a numerical aperture of 1.4, Leica Microsystems. A Z-stack was made with a step size of 200nm and pixel size of 24nm x 24nm, optimized using Nyquist Calculator (SVI Scientific Volume Imaging, Hilversum, the Netherlands). Finally, deconvolution was performed with Huygens Professional Software (SVI).

**ZEBRAFISH EXPERIMENTS**

Zebrafish (Danio rerio) embryo experiments were performed as described previously (Stoop et al. 2011). Briefly, M. marinum strains were grown to an OD_{600} of 1.0, washed in PBS and declumped in 0.5% Tween-80. The amount of colony-forming units (CFU) was verified by plating on 7H10 plates. Zebrafish embryos were injected 28 hours post fertilization, with 100-200 CFUs per embryo. Five days post infection, embryos were sedated after which the embryos were homogenized in 100µl of 5%SDS solution in PBS. Homogenized embryos were treated with Mycoprep (BBL), to
decontaminate the sample and different concentrations were plated on 7H10 plates without antibiotic additives. 10 days after plating, colony-forming units were counted. The experiment was performed on three individual time-points with six embryos per group per experiment.

ETHICS STATEMENT
All procedures involving embryonic zebrafish (Danio rerio) were performed in compliance with local animal welfare laws. All zebrafish were maintained according to standard protocols (zfin.org) and breeding of adult dish was authorized by the Animal Experimental licensing Committee (DEC) of the VU University Medical centre. All experiments followed the international guidelines specified by the EU animal protection directive 86/609/EEC, which allows use of embryonic zebrafish until a free-living stage (approximately 5-7 days after fertilization) and were therefore approved by the DEC of the VU University Medical centre (Amsterdam, the Netherlands).

ACKNOWLEDGEMENTS
We thank Marion Sparrius, Susanna Commandeur, Jeroen Kole, Roy Ummels and Janneke Maaskant for technical assistance. We thank Sander Piersma for data analysis. We thank Eric Brown, Astrid van der Sar, Ben Appelmelk and Edith Houben for useful discussions.
HYPERVIRULENT MYCOBACTERIUM TUBERCULOSIS ISOLATES DO NOT SECRETE PE_PGRS PROTEINS DUE TO MUTATIONS IN THE PPE38 LOCUS

Louis S Ates¹, Roy Ummels¹, Anzaan Dippenaar², Sander R Piersma³, Aniek D van der Woude¹, Marit Kok¹, Connie R Jiménez⁴, Edith NG Houben⁴, Rob M Warren², Rogelio Hernández-Pando³, Wilbert Bitter¹.⁴

Manuscript in preparation
1 Department of Medical Microbiology and Infection Prevention, VU University Medical Center, Amsterdam, the Netherlands.
2 DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa
3 Department of Medical Oncology, OncoProteomics Laboratory, VU University Medical Center, Amsterdam, the Netherlands
4 Department of Molecular Microbiology, VU University, Amsterdam, the Netherlands.
5 Experimental Pathology Section, Department of Pathology, National Institute of Medical Sciences and Nutrition “Salvador Zubirán”, México City, 14000, México.
SUMMARY

The ESX-5 secretion system of pathogenic mycobacteria is responsible for the secretion of dozens of substrates classified as the PE and PPE proteins. The largest sub-group of PE proteins are the PE_PGRS proteins, which are suggested to be involved in immune modulation. In this study, we show that the ESX-5 substrate PPE38 is required for the secretion of all detected PE_PGRS and all detected PPE-MPTR proteins. Hypervirulent M. tuberculosis strains of the Beijing lineage that have natural disruptions of the ppe38 locus also have a concomitant defect in PE_PGRS and PPE-MPTR secretion. Introduction of an intact ppe38 locus restored this secretion defect. Furthermore, complementation of the ppe38 locus resulted in decreased virulence, thereby linking PE_PGRS secretion deficiency with hypervirulence. In agreement with this, disruption of the intact ppe38 locus in a classical M. tuberculosis isolate resulted in a secretion defect and in increased lung bacillary loads. Together, these results show that PPE38 is an important modulating factor for virulence in tuberculosis.

INTRODUCTION

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), causes nine-million new cases each year and is currently the most deadly bacterial pathogen (World Health Organization 2013). Antibiotic resistant strains of M. tuberculosis are increasing in occurrence, posing a threat for further disease control measures (World Health Organization 2013). Furthermore, strains from specific lineages, such as the Beijing lineage, are reported to be hypervirulent and hypertransmissible in human populations. These Beijing strains seem to have an evolutionary advantage even in vaccinated individuals, indicating that these strains are adapting to spread more efficiently (Luo et al. 2015; Hanekom et al. 2011; Abebe and Bjune 2006). Although many different hypotheses have been put forward to explain this hypervirulence phenotype, they do not fully explain the observed strain differences (Hanekom et al. 2011; Sinsimer et al. 2008).

Mycobacteria possess a highly impermeable cell-envelope consisting of both an inner and an outer membrane (Zuber et al. 2008; Hoffmann et al. 2008), conferring high innate resistance to the hosts immune system and to antibiotics. To secrete proteins across this unique cell envelope, mycobacteria have up to five specialized TypeVII secretion (T7S) systems, called ESX-1 to ESX-5 (Bitter et al. 2009). These T7S systems are essential for the secretion of substrates belonging to the Esx, PE and PPE families. Substrates secreted by the ESX-1 system are involved in the rupture of the phagosome by mycobacteria (van der Wel et al. 2007), a process required for the virulence of pathogenic mycobacteria (Pym et al. 2002). ESX-5 is the most-recently evolved T7S system and is only present in slow-growing mycobacteria, which include most pathogenic mycobacterial species (Gey van Pittius et al. 2006). Interestingly, the ESX-5 system is also essential for growth of specific mycobacterial strains (Di Luca et al. 2012; Ates et al. 2015), probably through its role in outer membrane
permeability (Ates et al. 2015). In several mycobacterial pathogens, including *M. tuberculosis* and *Mycobacterium marinum*, the genes encoding the putative ESX-5 substrates have expanded significantly through duplication events. The genome of *M. tuberculosis* contains 99 pe and 69 ppe genes (Gey van Pittius et al. 2006), the vast majority of which have been shown to be secreted via the ESX-5 system (Abdallah et al. 2009; Ates et al. 2015). Biological functions have only been found for a few ESX-5 substrates, such as the lipase LipY (Deb et al. 2006), or PPE10, which we recently showed to be important for capsular integrity (Chapter 4). A large sub-group of ESX-5 substrates are the PE_PGRS proteins, named for the polymorphic GC-rich-repetitive sequences in their genes (Poulet and Cole 1995). These PGRS domains can be highly extended and encode glycine-rich repetitive domains (typically GGA and GGN repeats). Because of their variability and redundancy, these proteins are suggested to be involved in immune evasion of mycobacteria (Reviewed in Sampson, 2011 and Ates et al., in press). An ESX-5-depleted strain of *M. marinum* was shown to be hypervirulent in adult zebrafish (Weerdenburg et al. 2012), concurring with this hypothesis. However, non-essential ESX-5 mutants of *M. tuberculosis* were shown to be attenuated in a mice model of TB (Sayes et al. 2012; Bottai et al. 2012). The large amounts of ESX-5 substrates, their high level of homology and therefore their possible redundancy have obstructed detailed analysis of these groups of proteins.

In this study, we show that ppe38 mutants of *M. marinum* and *M. tuberculosis* are deficient in PE_PGRS secretion. Furthermore, natural mutations in the genomic ppe38-71 locus occur in clinical *M. tuberculosis* strains (McEvoy, van Helden, et al. 2009; McEvoy, Warren, et al. 2009) and we show that these mutation give rise to large variations in production and secretion of PE_PGRS and a subset of PPE proteins.

**RESULTS**

**Disruption of ppe38 results in reduced PE_PGRS secretion.**

In a previous study, aimed to identify transposon insertion mutants unable to secrete PE_PGRS proteins, we identified multiple insertions in the esx-5 locus (Ates et al. 2015). In addition, we also identified a transposon mutant in ppe38 (LA10), but this mutant was not included in the previous study since it was not part of the esx5-locus (Figure 1C) (Ates et al. 2015). PPE38 is in itself also an ESX-5 substrate (Ates et al. 2015) and the ppe38 gene is part of an operon which also encodes two Esx proteins (EsxP_3 & EsxN_4) (Dong et al. 2012). A transposon mutant of ppe38 in *M. marinum* was previously reported to have an altered colony morphology (Dong et al. 2012), a phenotype that could also be observed for LA10. In our mutant library we had a second independent transposon insertion mutant in ppe38 with an inversed orientation of the transposon as compared to LA10. Immunoblot analysis showed that also this mutant did not secrete PE_PGRS proteins into the culture filtrate and no PE_PGRS proteins were present in the Genapol-X080 extractable cell-surface enriched fraction (Figure 1A). Staining of GroEL2 was used as a loading
Figure 1. ppe38 mutants of M. marinum show a PE_PGRS secretion defect. A) M. marinum strain M and the isogenic transposon mutant ppe38::tn were analyzed for the secretion of PE_PGRS proteins by immunoblot. Complementation was performed by introducing derivatives of pSMT3 containing different parts of the ppe38 locus. Expression and secretion of PE_PGRS proteins in the ppe38::tn mutant was completely lost in the whole cell lysates (P), Genapol X-080 treated material (GP), the Genapol X-080 extractable cell-surface enriched fraction (GS), or the culture supernatant (S). Full complementation was achieved by reintroduction of the complete ppe38 operon (38-PN). Introduction of only ppe38 (38) or esxPN (PN) did not (fully) complement PE_PGRS secretion. Staining for GroEL2 was used as a loading and lysis control. Secreted PE_PGRS proteins of M. marinum usually show a more diffuse running behaviour on SDS-PAGE for unknown reasons.

B) HA-tagged PPE38 (38-HA) is efficiently secreted to the culture supernatant of M. marinum - ppe38::tn, but does not complement the secretion phenotype. C) Genapol-extracted material of M. marinum strain M expressing mspA, the ppe38::tn mutant in this background and mycP5 and eccD5 transposon mutants (described in Ates et al. 2015), was spotted on a nitrocellulose filter and stained with the anti-PGRS antibody.
the entire operon on a plasmid (pPPE38-PN) resulted in restored colony morphology and in restored expression and secretion of PE_PGRS proteins (Figure 1A). However, introduction of only the ppe38 gene (pPPE38) did not restore these phenotypic characteristics; only PE_PGRS secretion into the culture filtrate, but not to the cell surface, could be observed. Introduction of only esxP_3 and esxN_4 (pPN) did not result in complementation of any of the observed phenotypes. Therefore, the entire operon seems to be required for full complementation.

To examine the localization of PPE38, gene variants coding for HA-tagged PPE38 (PPE38-HA), or PPE38-HA together with the EsxP_3 and EsxN_4 proteins (PPE38-HA-PN) were expressed in M. marinum-ppe38::tn. In both cases a protein with an apparent molecular weight of 50 kDa was detected in high amounts in the culture filtrate by an anti-HA antibody (Figure 1B). This apparent molecular weight is higher than the predicted 37kDa, suggesting post-translational modifications. Surprisingly, no HA-tagged protein was detected in the Genapol-X-080 extractable fraction, indicating that the majority of PPE38 is secreted and not cell-surface associated. Neither the plasmid encoding PPE38-HA nor PPE38-HA-PN were able to complement the PE_PGRS secretion defect of the M. marinum-ppe38::tn strain, indicating that the C-terminal HA-tag might interfere in the molecular function of PPE38, even though these modified proteins are efficiently secreted.

To exclude the possibility that the PE_PGRS negative phenotype of ppe38::tn was due to a regulatory effect responsible for loss of PE_PGRS expression, we examined the effect of heterologous expression of different HA-tagged PE_PGRS proteins. The genes encoding M. tuberculosis PE_PGRS33-HA and PE_PGRS45-HA were placed under control of the 19kDa lipoprotein promoter (De Smet et al. 1999) and introduced in both wild-type M. marinum and ppe38::tn. As described previously (Abdallah et al. 2009), a processed form of both PE_PGRS proteins was observed in the Genapol-X080 extractable fraction in wild-type M. marinum, showing that these proteins are surface localized (Figure 2A). This processed form (~35kDa) of the PE_PGRS proteins could not be detected in the Genapol-X080 extractable fraction of ppe38::tn, although low amounts of the full-length protein (40-43kDa) could be detected in this Genapol supernatant. PE_PGRS33-HA and PE_PGRS45-HA could also be detected in the culture filtrate of wild-type M. marinum, but not in ppe38::tn. Next, a PE_PGRS gene from M. marinum (mmar_3581) was expressed under control of the hsp60 promoter. Introduction of this plasmid in the wild-type strain resulted in high amounts of a processed form (~53kDa) of MMAR_3581-HA in the culture filtrate of wild-type M. marinum. This band was not observed in the ppe38::tn strain (Figure 2B). Together, these results show that PPE38 is not required for the regulation of PE_PGRS proteins, but that this protein needs to be produced to efficiently secrete PE_PGRS proteins.
HYPERVIRULENCE AND SECRETION DEFECTS BY PPE38 MUTATIONS

Figure 2. Secretion and processing of heterologously expressed PE_PGRS proteins is impaired in the ppe38::tn mutant of M. marinum. A) Expression of HA-tagged PE_PGRS proteins of M. tuberculosis, i.e. PE_PGRS33-HA (33) or PE_PGRS45-HA (45), in wild-type M. marinum (WT) or the isogenic ppe38::tn mutant. As a control, we also used cells without a plasmid [-]. PE_PGRS levels were determined in whole cells, cells treated with genapol, genapol supernatant fraction enriched for surface proteins and the culture supernatant. B) Processing of the M. marinum PE_PGRS protein MMAR_3581-HA requires PPE38. Immunoblots stained with an anti-HA or anti-PGRS antibody of wild-type M. marinum or ppe38::tn expressing MMAR_3581-HA. Full-length MMAR3581-HA is detected in pellet fractions of both strains as an ~70kDa protein product. In wild-type M. marinum, MMAR_3581 is processed (~55kDa) and secreted in high amounts to the culture filtrate, while in the ppe38::tn strain a full-length product could be detected.

The PPE38 locus is required for PE_PGRS secretion in M. tuberculosis

In M. tuberculosis H37Rv, ppe38 (Rv2352c) is annotated as a ppe gene that is not flanked by esx genes (Cole et al. 1998). However, this was later shown to be incorrect; most M. tuberculosis strains, including laboratory strains H37Rv and CDC1551, contain a ppe38 locus of four genes (McEvoy, van Helden, et al. 2009). The outer two genes are both ppe genes (ppe38 and ppe71), which are almost identical in most M. tuberculosis isolates. These two genes flank a pair of esx-genes (mt2420-21, Figure 3A). In order to test whether the M. tuberculosis orthologue(s) of PPE38 perform the same function as in M. marinum, we created three different mutations of this locus in M. tuberculosis strain CDC1551 by homologous recombination (Supplemental materials). The first mutation, Δppe38-71, deleted the complete locus (Figure 3A). The second deletion, ΔBeijing, truncates both ppe-genes and deletes both esx genes. This mutation was made, because some clinical isolates belonging to the Beijing lineage have a similar deletion caused by insertion of an IS6110 mobile genetic element and subsequent recombination events (McEvoy, van Helden, et al. 2009). The third deletion strain,
Figure 3. PPE38 is required for PE_PGRS secretion in M. tuberculosis. A) Schematic representation of the ppe38-71 genomic region in M. tuberculosis CDC1551 and the deletions (black lines) that were made in this strain by homologous recombination. Genes ppe38 (mt2419) and ppe71 (mt2422) (both depicted in grey) are highly homologous (100% identity, with a 7 amino acid insertion in PPE38). The esx-genes (white) have no official gene nomenclature and are therefore referred to by their gene number (mt2421 and mt2420 respectively). B) Secretion of PE_PGRS proteins, but not PPE41, is lost in strains carrying mutations deleting or truncating the ppe38 locus. Whole cell lysates (pellet) and culture filtrate of the indicated mutant strains and the complemented, indicated with [–C], Δppe38-71 and ∆”Beijing” strains were analyzed by immunoblot with anti-GroEL2, anti-PPE41 or anti-PGRS antibodies.

Δesx, deleted the two esx genes, but left both ppe38 and ppe71 intact (Figure 3A). In line with what we observed in M. marinum, deletion or truncation of the whole ppe38 locus resulted in complete loss of PE_PGRS secretion (Figure 3B). However, in contrast to what was observed in M. marinum, the esx genes do not seem to be required
for PE_PGRS secretion, since Δesx secreted similar amounts of PE_PGRS proteins as wild-type *M. tuberculosis* (Figure 3B). Secretion of the ESX-5 substrate PPE41 was not affected by these mutations, indicating that the secretion defect is specific for a sub-group of ESX-5 substrates. PE_PGRS secretion was completely restored in both Δppe38-71 and ΔBeijing strains upon complementation (Fig. 3B).

**Clinical *M. tuberculosis* isolates with mutated ppe38-loci are deficient in PE_PGRS secretion**

As mentioned above, the *ppe38 locus* in *M. tuberculosis* is hypervariable due to genetic insertion of a IS6110 genetic element, combined with recombination events between the nearly identical *ppe* genes (McEvoy, van Helden, et al. 2009; McEvoy, Warren, et al. 2009). These recombination events are most frequent in strains belonging to the Beijing lineage and in some cases lead to truncations of these genes. We hypothesized that strains that have truncated or deleted copies of the *ppe38* locus would also be deficient in PE_PGRS secretion. To test this, three clinical isolates from a cohort in South Africa (Hanekom et al. 2007), all belonging to the Beijing lineage, were tested. These strains have different genomic organization of the *ppe38-71*-locus (McEvoy, van Helden, et al., 2009) (Figure 4B). Although all strains belong to the Beijing lineage, there seems to be high variation in their virulence and transmission. The strain with an intact *ppe38* locus (SAWC_2088) was only isolated from a single patient and shown to be hypovirulent and lowly transmittable in a mouse model of tuberculosis (Aguilar et al. 2010). On the other hand, isolate SAWC_2135 belongs to a strain with a disrupted *ppe38* locus (Family F29 or Beijing sublineage 7), was highly transmitted in the Cape Town cohort and was isolated from 221 different patients (Hanekom et al. 2007). In this strain, both *esx* genes are deleted and *ppe38* and *ppe71* are both truncated due to recombination events (Figure 4B) (McEvoy, van Helden, et al., 2009). Finally, strain SAWC_2701 (family F27, Beijing sublineage 2) was isolated from eight different patients and this strain has the same genetic structure of the *ppe38-71* locus as SAWC_2135, but with an inversed orientation of the IS6110 element (Figure 4B) (McEvoy, van Helden, et al., 2009).

Secretion analysis of these strains showed that the isolate with an intact *ppe38* locus (SAWC_2088) secreted PE_PGRS proteins (Figure 4A). In contrast, the two other strains, in which all four genes of the *ppe38* locus were truncated or deleted, were completely deficient in PE_PGRS secretion (Figure 4A). Interestingly, introduction of the genomic locus *ppe38* via the integrative plasmid pMV::*ppe38*-71 completely restored PE_PGRS secretion in these isolates (Figure 4A). As a control we also examined the ESX-1 substrate EsxA and the cytosolic protein GroEL2. EsxA was secreted in slightly lower amounts in strain SAWC_2701, but was unaffected by reintroduction of the *ppe38-71* genes (Figure 4A). These results show that mutations in the *ppe38* locus in clinical *M. tuberculosis* isolates result in a loss of PE_PGRS secretion and that this phenotype can be reversed by introducing an intact copy of the *ppe38* locus.
Proteomic analysis of Beijing isolates of *M. tuberculosis* reveals that both PE_PGRS and PPE_MPTR proteins require PPE38 for secretion.

We used a proteomic approach to investigate the secretome of the different clinical isolates of *M. tuberculosis* in more detail and to determine the PPE38-dependently secreted proteins in these strains. LC-MS/MS analysis was performed on the culture filtrate of SAWC_2088, SAWC_2135, SAWC_2701 and the PPE38-complemented strains SAWC_2135-C and SAWC_2701-C (Full results in Supplemental information). The relative abundances of known T7S substrates and a subset of nine T7S-independent extracellular proteins of *M. tuberculosis* were compared between the three Beijing isolates (Figure 5). The most abundantly secreted substrates in all strains included...
HYPERVIRULENCE AND SECRETION DEFECTS BY PPE38 MUTATIONS

Figure 5. An intact ppe38 locus is required for PE_PGRS and PPE_MPTR secretion. Graphical summary of LC-MS/MS analysis of secreted proteins in strains with a mutated ppe38 locus: SAWC_2135 (left) and SAWC_2701 (right). The number of detected spectral counts was plotted and compared between the strains without ppe38 (Y-axes) and SAWC_2088 with an intact ppe38 copy (X-axes). Proteins with restored secretion upon introduction of pMV::ppe38-71 are indicated with an asterisk and all belonged to the PE_PGRS or PPE_MPTR subclasses. The number of spectral counts represents the sum of two independent biological replicates. Full data are available in the supplemental information.

the ESX-1 substrates EsxA, EsxB, EspA and EspB and the non-T7S secreted proteins MPT64, MPT63, FbpA and FbpC. Secretion levels of these substrates showed only moderate strain variation (Figure 5) and were not significantly affected by the introduction of pMV::ppe38-71 in SAWC_2135 and SAWC_2701, affirming that presence of the ppe38 locus does not affect the secretion of these substrates (Supplemental Results). Although only low levels of PE_PGRS proteins could be detected by this proteomic analysis, these proteins were only detected in complemented strains SAWC_2135-C and SAWC_2701-C and in strain SAWC_2088, which has a functional ppe38 locus, but were completely absent in strains SAWC_2135 and SAWC_2701 (Figure 5, Table 1). These proteomic data thereby support our findings obtained by immunoblotting and suggest that PE_PGRS secretion is fully dependent on PPE38. Surprisingly, also a number of PPE proteins were shown to be PPE38-dependent. Apart from the PPE38 protein, which was detected in our samples of both the complemented strains, five PPE proteins were only detected in the complemented strains (Table 1). Of these five PPE proteins, (i.e. PPE40, PPE10, PPE53, PPE33 and PPE52), four were also detected in strain SAWC_2088 and were therefore considered to be PPE38 dependent. As expected, there were also a number of PPE proteins that were clearly not affected by the presence or absence of PPE38, such as PPE18, which was detected in high amounts in all five strains examined. Finally, three PPE proteins (i.e. PPE36, PPE51, PPE65) were detected in SAWC_2135, but not in the other clinical
Table 1. Proteins restored in secretion upon PPE38 complementation. Depicted are the spectral counts of proteins detected in LC-MS/MS analysis of the culture filtrate of two clinical isolates of M. tuberculosis with mutations in the ppe38 locus (SAWC_2135 and SAWC_2701) and these isolates complemented with the ppe38 locus on a plasmid (pMV::ppe38-71). Only the PE, PPE, Esx, ESX-1 associated (Esp) and nine major secreted substrates of M. tuberculosis were analyzed and proteins were depicted when a two-fold increase in detected spectral counts occurred in both complemented strains compared to the original isolates.

<table>
<thead>
<tr>
<th></th>
<th>SAWC_2135</th>
<th>SAWC_2135- pMV::ppe38-71</th>
<th>SAWC_2701</th>
<th>SAWC_2701- pMV::ppe38-71</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPE40</td>
<td>0.0</td>
<td>4.1</td>
<td>0.0</td>
<td>10.5</td>
</tr>
<tr>
<td>PPE38</td>
<td>0.0</td>
<td>4.1</td>
<td>0.0</td>
<td>15.3</td>
</tr>
<tr>
<td>PE_PGRS39</td>
<td>0.0</td>
<td>4.1</td>
<td>0.0</td>
<td>7.3</td>
</tr>
<tr>
<td>PPE10</td>
<td>0.0</td>
<td>3.0</td>
<td>0.0</td>
<td>9.2</td>
</tr>
<tr>
<td>PPE53</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>4.0</td>
</tr>
<tr>
<td>PE_PGRS38</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>4.0</td>
</tr>
<tr>
<td>PE_PGRS24</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>2.7</td>
</tr>
<tr>
<td>PPE33</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>PE_PGRS29</td>
<td>0.0</td>
<td>0.9</td>
<td>0.0</td>
<td>2.5</td>
</tr>
<tr>
<td>PPE52</td>
<td>0.0</td>
<td>6.1</td>
<td>0.0</td>
<td>9.3</td>
</tr>
</tbody>
</table>

isolates and were not affected by complementation of PPE38 in this strain. These differences were therefore considered as regular strain-specific variations that were not related to the presence of PPE38. Analysis of the PPE38 dependent PPE substrates revealed that they all were part of the Major Polymorphic Tandem Repeat (PPE_MPTR) subclass of PPE’s and inversely, none of the PPE38-independent PPEs belonged to this subclass (Gey van Pittius et al. 2006). These data suggest, that in addition to the PE_PGRS proteins, also the PPE_MPTR proteins are dependent on PPE38 for their secretion.

Loss of PE_PGRS and PPE_MPTR secretion due to mutations in the ppe38 locus leads to hypervirulence of M. tuberculosis in a mouse model.

The highly transmitted M. tuberculosis isolate SAWC_2135 (Hanekom et al. 2007), with a disrupted ppe38 locus, corresponds to the reported hypertransmissibility and hypervirulence of certain Beijing isolates (Aguilar et al. 2010). We have previously reported that reduced secretion via ESX-5 in M. marinum also results in hypervirulent phenotype in adult zebrafish (Weerdenburg et al. 2012). Therefore, we hypothesized that hypervirulence of M. tuberculosis strains could be (partially) caused by impaired PE_PGRS or PPE_MPTR secretion due to mutations in the ppe38 locus. To test this hypothesis, clinical isolates SAWC_2135 (no PE_PGRS/PPE_MPTR secretion) and the closely related strain SAWC_2088 with an intact ppe38 locus were tested in a mouse model of M. tuberculosis. In addition, we also included SAWC_2135-pMV::ppe38-71 with restored PE_PGRS/PPE_MPTR secretion. Similar levels of colony forming units
(CFUs) were isolated from the organs of all infected mice in the first weeks of infection (Figure 6A). At 21 days post infection, a significant (P<0.001) increase of CFUs could be detected for strain SAWC_2135 as compared to SAWC_2088, which increased up to an approximate 500-fold difference at 28 days post infection. Importantly, the same isolate complemented with plasmid pMV::ppe38-71 did not show increased CFUs as compared to SAWC_2088, indicating that the increase in CFU is caused by the loss of PPE38 and therefore probably by the loss of PE_PGRS and/or PPE_MPTR secretion (Figure 6A). We also observed an increased mortality of mice infected with strain SAWC_2135 compared to SAWC_2088 (Figure 6B). Also this phenotype was partially restored in SAWC_2135-pMV::ppe38-71. To confirm whether loss of PE_PGRS secretion leads to hypervirulence, we also tested the effect of deleting the ppe38 locus in a strain normally secreting PE_PGRS proteins. To do so, the virulence of laboratory strain CDC1551, its isogenic Δppe38-71 mutant and the complemented strain were tested in the same mouse model. Disruption of ppe38 indeed resulted in a significant increase in lung bacillary loads at later time points as compared to the wild-type and complemented strains (Fig 6C). However, this increase in bacterial numbers did not result in increased mortality, probably due to the more mild increase in virulence. Together, these results show that mutations of the ppe38 locus and the concomitant loss of PE_PGRS/PPE_MPTR secretion affect bacillary loads in a mouse model of pulmonary tuberculosis and therefore could play a role in the hypervirulence and hypertransmission observed for strains of the Beijing lineage (Aguilar et al. 2010).

DISCUSSION

In this study, we have shown that PPE38 is required for the secretion of all detected PE_PGRS proteins in both M. marinum and M. tuberculosis. Furthermore, we have shown that variations in the ppe38 locus in clinical M. tuberculosis isolates result in large differences in the number of secreted PE_PGRS proteins. Proteomic analysis showed that the secretion defect also included the PPE_MPTR subclass of ESX-5 substrates. Finally, complementation of this phenotype resulted in reduced virulence in a mouse model of M. tuberculosis. Although the mechanism of action by which PPE38 is required for PE_PGRS secretion remains uncertain, our data have given several indications of its possible function. In M. marinum, the phenotype of the ppe38::tn mutant strain could only be complemented by the complete operon, encompassing ppe38 as well as esxP_3 and esxN_4. This indicates that the ESX proteins are important for the function of PPE38 in secretion. However, in M. tuberculosis deletion of the two esx genes did not result in abrogation of PE_PGRS secretion. A possible explanation for this is, that there are multiple close homologues of the EsxP and EsxN proteins in the M. tuberculosis genome that could compensate for the loss of these genes. Another hypothesis is that this phenotype is due to a fundamental difference in the localization of PE_PGRS proteins between M. marinum and M. tuberculosis; i.e. M. tuberculosis, in contrast to M. marinum, has no PE_PGRS proteins that can
Figure 6. ppe38 mutations in M. tuberculosis lead to an increased virulence in a mouse model of tuberculosis and increased lung bacillary loads. A) Bacterial loads in CFU per lung of BALB/c mice intratracheally infected with 2.5 \( \times \) 10^5 CFUs of the indicated strains. Mice were sacrificed at the indicated time-points, one lung per mouse was homogenized and serially diluted and plated to count CFUs. Strain SAWC_2135 (white) had significantly (indicated with *) higher bacillary loads from 21 days of infection and onwards, compared to SAWC_2088 (black). B) Kaplan-Meier survival curves of mice infected with the indicated strains. Mice infected with SAWC_2135 (light grey line) died significantly faster compared to strain SAWC_2088 (black line), or compared to SAWC_2135-pMV::ppe38-71 (dark grey line). C) Lung bacillary burden of mice infected with CDC1551 (black), the Δppe38-71 strain (white) and the complemented strain Δppe38-71-C (grey). A significantly higher bacterial burden was detected 60 days post infection for mice infected with Δppe38-71 compared to the wild-type or complemented strains. D) Survival of mice infected with the indicated M. tuberculosis strains. CDC1551 wild-type (black line) and Δppe38-71 (light grey line) showed similar levels of survival, whereas all mice infected with the complemented strain (dark grey line) survived until 120 days post infection when the experiment was terminated. Statistical analysis of bacillary burden was performed by two-way ANOVA followed by the Tukey test for multiple comparisons. Differences were deemed significant when P<0.001. Statistical analysis of Kaplan-Meier curves was performed by the Mantel-Cox log rank test.

be detected after extraction from the cell surface by Genapol X-080 (Sani et al., 2010; Ates and Ummels unpublished results). Therefore, only the observed effects in the culture filtrate can be analysed. Overall, it is clear that the strongest role in PE_PGRS secretion is played by PPE38 and that the esx-genes play at most a supporting role in this process. Surprisingly, HA-tagged PPE38 constructs were unable to
complement the ppe38::tn-mutant, even though the protein products were efficiently secreted. This suggests that C-terminal part of this protein plays an important role in PPE38 functioning. Furthermore, these data indicate that, although PPE38 is required for PE_PGRS secretion, there is no mutual co-dependence, as was observed previously for ESX-1 substrates (Fortune et al. 2005). One possible role of PPE38 is that it is involved in the processing of PE_PGRS proteins upon secretion. However, BLAST analysis, or Phyre protein structure prediction (Kelley and Sternberg 2009) of the C-terminal domain of PPE38 did not reveal any known protease domains or other suggested functions. Perhaps PPE38 functions as a chaperone for PE_PGRS protein and is required for extracellular stability. Current efforts in our lab are focussing on identifying the interaction partners of PPE38. Our proteomic and immunological analyses suggest that the genetic loss of the ppe38 locus in M. tuberculosis results in a loss of secretion of all produced PE_PGRS proteins and a subset of PPE proteins known as PPE_MPTRs, but that other ESX-5 substrates remain unaffected. Interestingly, both these affected subfamilies have only recently evolved and some of the encoding genes form an operonic structure (Gey van Pittius et al. 2006). This seems to indicate that PPE_MPTR and PE_PGRS proteins could have a shared function or co-dependence, which further analyses will have to confirm. The fact that ppe38 locus with the two esx genes is required only for the secretion of a subset of ESX-5 substrates is not without precedent. Recent published data showed that another pe-ppe-esx locus (Shah et al. 2015), is required for the secretion of a different set of ESX-5 substrates.

A previous virulence study of M. tuberculosis ESX-5 secretion mutants showed an attenuation of virulence (Bottai et al. 2012; Sayes et al. 2012). This attenuation was correlated with a loss of cell-wall integrity and increased membrane permeability. Later, we and others have shown that the ESX-5 system is usually essential for growth (Ates et al. 2015; Di Luca et al. 2012), unless outer membrane permeability was compromised (Ates et al. 2015). Since ppe38 mutants have been readily identified in different transposon screens of M. marinum (This study, Dong et al., 2012), clinical strains of M. tuberculosis (McEvoy, van Helden, et al., 2009) and were shown to be non-essential by saturated transposon mutagenesis of two different M. marinum strains (Weerdenburg et al. 2015; Ates et al. 2015) and M. tuberculosis (Griffin et al. 2011), ppe38 is clearly not essential for in vitro growth of mycobacteria. This provides an opportunity to study the biological role of PPE38-dependent ESX-5 substrates. For instance, our data suggest that the PPE38-dependent substrates of ESX-5, i.e PE_PGRS and PPE_MPTR proteins, are not the substrates responsible for the essentiality of this secretion system. Inversely, the PPE38 dependent ESX-5 substrates do seem to be responsible for the tempering of virulence in wild-type M. tuberculosis (this study) or M. marinum (Weerdenburg et al. 2012).

Perhaps the most striking finding of this study is that clinical strains of M. tuberculosis have high variations in the secretion of PE_PGRS proteins due to ppe38-71 mutations. This locus was previously shown to be hypervariable in M. tuberculosis (McEvoy, van Helden, et al. 2009), this means that there is also a high
variation in the production of PE_PGRS proteins between the different M. tuberculosis isolates. Furthermore, this variation seems to have an effect on the virulence of some of these strains, as strains of the Beijing lineage have previously been described to be hypervirulent (Reed et al. 2004; Tsenova et al. 2005; Via et al. 2013) and hypertransmissible (Hanekom et al. 2007; Aguilar et al. 2010) in humans and different animal models. The production of phenolic glycolipids (PGLs) has been implicated as a possible cause for the hypervirulence of Beijing strains (Reed et al. 2004; Tsenova et al. 2005). However, a later study (Sinsimer et al. 2008) showed that this mutation is responsible for hyperactivation of the immune responses against M. tuberculosis, but is not responsible for increased bacterial loads. In our study, a significantly higher amount of CFUs could be detected in the lungs of mice infected with strains lacking ppe38, indicating that the hypervirulence observed in strains with a defect in ppe38 is therefore likely different in mechanism from the PGL-related hypervirulence. The increase in lung CFUs could be fully complemented by reintroducing the ppe38-71 operon, which also resulted in a partial complementation of the increased mortality phenotype of SAWC_2135. Since the isolate SAWC_2135 is part of a large outbreak, caused by a single strain of M. tuberculosis, it is tempting to speculate that this strain is hypertransmissible due to its loss of PE_PGRS/PPE_MPTR secretion, but multiple factors might be responsible for this phenotype.

Although there was also a significant increase in lung CFUs in the ppe38-71 mutant of lab strain CDC1551, this difference was only significant at 60 days of infection. Perhaps these delayed differences are also responsible for the lack of detectable differences in mortality between wild-type CDC1551 and the mutant strain CDC1551-Δppe38-71. It should be emphasized that this experiment has only been performed once, with a methodological focus on analysing the CFU differences and that therefore the mortality of these strains should be assessed further before definitive conclusions can be drawn.

Surprisingly, Dong et al. have reported that a ppe38 mutant in M. marinum was attenuated in zebrafish embryos and adult fish (Dong et al. 2012). Whether this is due to species-specific differences, or is due to methodological differences between the infection models that were used, remains to be established. However, the findings that M. tuberculosis strains with a truncated ppe38 locus are highly transmitted in clinical cohorts (Hanekom et al. 2007; R. S. Lee et al. 2015) and show hypervirulence in a mouse model (Aguilar et al., 2010; this study), make the hypothesis that PPE38 is required for virulence of M. tuberculosis highly unlikely. Furthermore, the observed hypervirulence due to the mutation in the ppe38 locus of M. tuberculosis does correlate with previous results by our group, showing that an ESX-5 deficient strain of M. marinum is hypervirulent (Weerdenburg et al. 2012). This hypervirulence was also associated with a large increase of CFUs.

Finally, the observed effects may extend beyond the Beijing lineage. A recently published study (R. S. Lee et al. 2015) describes a highly dominant cluster of M. tuberculosis strains belonging to Lineage 4 (Euro-American lineage) in the Inuit
population in Canada that is efficiently transmitted. Intriguingly, although the authors suggest that strain-specific mutations are not responsible for the hypertransmissibility of these strains, all strains in this cohort have a deleted ppe38 locus. Similarly, genetic investigation of a patient cohort from San Francisco revealed a mutation called DS9 (M Kato-Maeda et al. 2001), which deletes the ppe38 locus as well as the adjacent locus of phospholipase C encoding genes, which are do not play a major role in mycobacterial virulence (Le Chevalier et al. 2015). The strains with this DS9 mutation had the highest number of patients with cavitary disease and had the highest percentage of tuberculin skin test positivity of direct contacts, suggesting hypervirulence and hypertransmissibility of the strains with this mutation (Table 1 in Kato-Maeda et al. 2001). It is of course tempting to speculate that the mutations of the ppe38-locus in these strains are also responsible for the evolutionary success of the strains in these cohorts.

How the PE_PGRS and/or PPE_MPTR proteins modulate the virulence of mycobacteria is currently not known, although different interactions with immune receptors have been reported (reviewed in Sampson, 2011). Our study provides some of the strongest evidence thus far that the PE_PGRS proteins indeed function as modulators of mycobacterial virulence. The ppe38 mutant strain will be an important tool to study the function of this sub-group of ESX-5 proteins in more detail, as there will probably be a high rate of redundancy among the highly similar PE_PGRS and PPE_MPTR proteins.

Taken together, these results show a major role for the modulation of virulence by PPE38 and the PE_PGRS proteins that is clinically relevant. These results warrant a more intensified research on the ESX-5 system and its substrates. The ESX-5 substrates have been relatively understudied, due to technical difficulties associated with working on these proteins, but we now show that they are crucial to understand mycobacterial virulence. Improved understanding of PPE38 and the PE_PGRS proteins will likely have an important impact on the field of mycobacterial pathogenesis.

EXPERIMENTAL PROCEDURES

**Strains and growth conditions**

All mycobacterial strains were grown on Middlebrook 7H10 plates (Difco) containing OADC supplement (BD bioscience) or liquid 7H9 medium containing ADC supplement (BD bioscience) and the appropriate antibiotics; 50µg/ml hygromycin (Sigma) and/or 25 µg/ml kanamycin (Roche). *M. marinum* strains were incubated at 30°C, while *M. tuberculosis* strains were incubated at 37°C. All mycobacterial strains and mutants are listed in supplemental table 1 (Supplemental Experimental Procedures). *Escherichia coli* strain DH5α was used for cloning procedures and was incubated on LB plates or liquid broth at 37°C.
### Table s1: Strain information

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Species</th>
<th>Original Strain</th>
<th>Relevant strain characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>M&lt;sup&gt;USA&lt;/sup&gt; ppe38::tn</td>
<td><em>M. marinum</em></td>
<td>M&lt;sup&gt;USA&lt;/sup&gt;</td>
<td>Clinical isolate</td>
</tr>
<tr>
<td></td>
<td><em>M. marinum</em></td>
<td>M&lt;sup&gt;USA&lt;/sup&gt;</td>
<td>Himar transposon in <em>mmar_3661</em> (position 4508719). Kan resistance cassette on transposon has same orientation as ORF <em>mmar_3661</em>.</td>
</tr>
<tr>
<td>LA10</td>
<td><em>M. marinum</em></td>
<td>M&lt;sup&gt;USA&lt;/sup&gt; pSMT3- mspA</td>
<td>Transposon mutant in <em>mmar_3661</em> (4510590). Orientation reversed to ORF <em>mmar_3661</em>.</td>
</tr>
<tr>
<td>M. tuberculosis WT</td>
<td><em>M. tuberculosis</em></td>
<td>CDC1551</td>
<td>Wild-type strain</td>
</tr>
<tr>
<td>Δppe38-71</td>
<td><em>M. tuberculosis</em></td>
<td>CDC1551</td>
<td><em>ppe38</em> locus deleted, see figure 3A</td>
</tr>
<tr>
<td>Δesx</td>
<td><em>M. tuberculosis</em></td>
<td>CDC1551</td>
<td><em>ppe38</em> locus partially deleted, See figure 3A</td>
</tr>
<tr>
<td>ΔBeijing</td>
<td><em>M. tuberculosis</em></td>
<td>CDC1551</td>
<td>See figure 3A</td>
</tr>
<tr>
<td>SAWC_2088</td>
<td><em>M. tuberculosis</em></td>
<td>Clinical isolate</td>
<td>Strain isolated in Cape Town cohort (Hanekom et al. 2007). Strain 1 in Aguilar et al 2010.</td>
</tr>
<tr>
<td>SAWC_2135</td>
<td><em>M. tuberculosis</em></td>
<td>Clinical isolate</td>
<td>Strain isolated in Cape Town cohort (Hanekom et al 2007).</td>
</tr>
<tr>
<td>SAWC_2701</td>
<td><em>M. tuberculosis</em></td>
<td>Clinical isolate</td>
<td>Strain isolated in Cape Town cohort (Hanekom et al 2007).</td>
</tr>
</tbody>
</table>

**Collection of clinical *M. tuberculosis* isolates**

In earlier studies (van der Spuy et al. 2009; Hanekom et al. 2007), sputum cultures were obtained from patients attending primary health care clinics in metropolitan Cape Town, South Africa. *M. tuberculosis* strains present in sputum culture were genotyped using IS6110 DNA fingerprinting (van Embden et al. 1993). DNA fingerprints were analysed with GelCompar software using the unweighted-pair group method, average linkages and Dice coefficients (Hermans et al. 1995). Strains with an IS6110 similarity index of ≥65% were grouped into strain lineages. Members of the Beijing genotype strain lineage were identified by the absence of region of difference RD105 (Tsolaki et al. 2005) and were further divided into sublineages as previously described (Tsolaki et al. 2005; Hanekom et al. 2007). The three selected strains were chosen because of their different genomic organizations of the *ppe38* genomic locus (McEvoy, van Helden, et al. 2009). Isolate SAWC_2088 belonged to the Beijing lineage 1 and therefore belongs to Beijing family F31; Isolate SAWC_2135 belongs to Beijing lineage 7, corresponding to family F27; Isolate SAWC_2701 belongs to Beijing lineage 2, corresponding to family F29. Isolate SAWC_2088 was earlier described to be lowly virulent in a mouse model by Aguilar et al. 2010 (strain 1 in Aguilar et al., 2010). Isolates SAWC_2135 and SAWC_2701 were not previously assessed for virulence.
Construction of plasmids

All plasmids are listed in supplemental table 2 (Supplemental Experimental Procedures). Primers used for the construction of plasmids can be found in supplemental table 3 (Supplemental Experimental Procedures). All plasmids containing *M. marinum* genes were constructed using *M. marinum* M\(^{55A}\) genomic DNA as template DNA. Plasmid pPPE38 was created by PCR amplification of *mmar_3661* with primers *ppe38-Fw* and *ppe38-Rv*. The PCR product was cloned into the pSMT3-LipY\(_{\text{ub}}\) plasmid replacing the *lipY* gene (Daleke et al. 2011), using restriction enzymes BamHI and HindIII. pPPE38-PN and pPN were constructed in the same way, using primers *ppe38-Fw* and *esxPN-Rv*, or *esxPN-Fw* and *esxPN-Rv*, respectively. The gene encoding PPE38-HA was created similarly by amplification with primers *ppe38-Fw* and *ppe38-Rv-HA* from pPPE38-PN template DNA. Since PCR amplification of *esxP_3 esxN_4* and the intergenic region between *mmar_3661* and *esxP_3* was repeatedly unsuccessful, this product (Supplementary procedures) was synthesized (GeneArt, Life technologies). The synthesized product was cloned after PPE38-HA to create pPPE38-HA-PN. Plasmid p3581-HA was constructed by cloning the PCR-product obtained by primers 3581-Fw and 3581-HA-Rv into pSMT3 using XbaI and BglII. pMV::ppe38-71 was

### Table s2: Primers

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5' → 3'</th>
<th>Enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ppe38-Fw</em></td>
<td>CGGGATCCAGGAGGGGTTGTGATGGTTTGGACTTTG</td>
<td>BamHI</td>
</tr>
<tr>
<td><em>Ppe38-Rv</em></td>
<td>TTAAGCTTTCGACACCGATCCGCGGCACCA</td>
<td>HindIII</td>
</tr>
<tr>
<td><em>Ppe38-Rv-HA</em></td>
<td>CCAAGCTTCACGCGTAGTCCGGCACGTCGTACGGGTA</td>
<td>HindIII</td>
</tr>
<tr>
<td><em>esxPN-Fw</em></td>
<td>AAAGGATCCGCGGGGGGTGATTGGTCGGG</td>
<td>BamHI</td>
</tr>
<tr>
<td><em>esxPN-Rv</em></td>
<td>AAGCTTATGGGCCACATCTGACTAGG</td>
<td>HindIII</td>
</tr>
<tr>
<td>3581-Fw</td>
<td>CCTCTAGAATGCCGCGTAAGCTTGGCTGG</td>
<td>XbaI</td>
</tr>
<tr>
<td>3581-HA-Rv</td>
<td>CCCAAGCTTCACGCGTAGTCCGGCACGTCGTACGGGTA</td>
<td>BglII</td>
</tr>
<tr>
<td>Mt2422-Fw</td>
<td>TTTGAATCTTTTTCGCTGGTGATGTG</td>
<td>EcoRI</td>
</tr>
<tr>
<td>Mt2419-Rv</td>
<td>TTAAGCTTCCGCGAGGTACATTTGGCAGCG</td>
<td>HindIII</td>
</tr>
<tr>
<td>PPE38 Del2 RR</td>
<td>TTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 Del2 RF</td>
<td>TTTTTTTTTTTTATTTTTGCTGGTGGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 Del2 LR</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 Del2 LF</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 (51bp) KO LR</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 (51bp) KO LF</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 (72bp) KO RF</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 (72bp) KO RR</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 Del LF</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 Del LR</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 Del RF</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
<tr>
<td>PPE38 Del RR</td>
<td>TTTTTTTTTTTTATTTTTTGGTGGTTGCTGGTCACTCTTG</td>
<td>Van91I</td>
</tr>
</tbody>
</table>
Table 3. Plasmid information

<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Plasmid backbone</th>
<th>Gene</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pPPE38</td>
<td>pSMT3, hsp60, hygromycin</td>
<td><em>Mmar</em> <em>3661</em></td>
<td>This study</td>
</tr>
<tr>
<td>pPPE38-PN</td>
<td>pSMT3, hsp60, hygromycin</td>
<td><em>Mmar</em> <em>3661-59</em></td>
<td>This study</td>
</tr>
<tr>
<td>pPN</td>
<td>pSMT3, hsp60, hygromycin</td>
<td><em>Mmar</em> <em>3660-59</em></td>
<td>This study</td>
</tr>
<tr>
<td>pPPE38-HA</td>
<td>pSMT3, hsp60, hygromycin</td>
<td><em>Mmar</em> <em>3661</em></td>
<td>This study</td>
</tr>
<tr>
<td>pPPE38-HA-PN</td>
<td>pSMT3, hsp60, hygromycin</td>
<td><em>Mmar</em> <em>3661-59</em></td>
<td>This study</td>
</tr>
<tr>
<td>p3581-HA</td>
<td>pSMT3, hsp60, hygromycin</td>
<td><em>Mmar</em> <em>3581</em></td>
<td>This study</td>
</tr>
<tr>
<td>p19kPro:: PE_PGRS45</td>
<td>p19kPro, 19kDa lipoprotein, hygromycin</td>
<td><em>rv2615c</em></td>
<td>Abdallah et al. 2009</td>
</tr>
<tr>
<td>p19kPro:: PE_PGRS33</td>
<td>p19kPro, 19kDa lipoprotein, hygromycin</td>
<td><em>rv1818c</em></td>
<td>Abdallah et al. 2009</td>
</tr>
<tr>
<td>pMV::ppe38-71</td>
<td>pMV, hsp60, kanamycin</td>
<td><em>mt2419-22</em></td>
<td>This study</td>
</tr>
</tbody>
</table>

constructed by amplification of CDC1551 genomic DNA by primers *mt2422-Fw* and *mt2419-Rv*. This PCR-product was cloned into the pMV361 vector (Stover et al. 1991) by restriction enzymes EcoRI and HindIII. The relevant sequence of all constructs was checked by nucleotide sequence analysis.

Construction of mutant strains

Transposon mutants in *M. marinum* were produced using the Mycomar phage system as described previously (Ates et al. 2015; Sassetti, Boyd, and Rubin 2003). Mutant LA10 was identified in a screen described in Ates et al., 2015 and is a transposon mutant in *mmar* _3661_ at bp 4510590. Targeted knock-outs of CDC1551 were created as described by Bardarov et al., 1997. The constructs for homologous recombination were created by amplification of CDC1551 genomic DNA. Primers PPE38 Del2 LF and PPE38 Del2 LR were used to amplify the sequence on the 3’ end of *mt2422*, while primers PPE38 Del2 RF and PPE38 Del2 RR were used to amplify the sequence on the 5’end of *mt2419*. Homologous recombination based on these constructs led to a complete deletion of *ppe38-71*, including the 36bp upstream of *mt2422* and the 57bp downstream of *mt2419*, resulting in strain Δppe38-71. Primers PPE38 (51bp) KO LF and PPE38 (51bp) KO LR were used to amplify the sequence spanning the region from 51bp of *mt2422* up to the 25bp on the 3’end of *mt2422*, while primers PPE38 (72bp) KO RF and PPE38 (72bp) KO RR were used to amplify the sequence spanning the 5’ end of *mt2419* up to 71bp of *mt2419*. Homologous recombination
based on these constructs led to truncations of both *mt2419* and *mt2422* and deletion of *mt2420-21*, resulting in strain ΔBeijing. Primers PPE38 Del LF and PPE38 Del LR were used to amplify the sequence from 168bp after the 3’ end of *mt2420* up to 26bp after the 3’ end after *mt2419*, while primers PPE38 Del RF and PPE38 Del RR were used to amplify the sequence spanning from 113bp of *mt2422* up to 46bp on the 3’ end of *mt2422*. Homologous recombination based on these constructs led to a complete deletion of *mt2420-21* including the 42bp upstream of *mt2421* and the 167bp downstream of *mt2420* resulting in strain Δesx.

**Secretion analysis and immunoblotting**

Secretion analysis of *M. marinum* was performed as described before (Daleke et al. 2011). Shortly, bacteria were cultured until mid-logarithmic phase and then washed three times in 7H9 medium without BSA. 0.35 OD/ml of bacteria were inoculated in 7H9 medium without BSA and cultured overnight. Subsequently, culture supernatant was isolated, filtered and the secreted proteins TCA precipitated. Also the bacterial cells were isolated and solubilized in denaturation buffer, lysed by sonication, and heated for 10 minutes at 95°C. Another aliquot of bacterial cells was treated for 30 min at room-temperature with 0.5% Genapol X-080 to extract surface localized proteins. For *M. tuberculosis* secretion analysis, cells were incubated in 7H9 medium without BSA for 40-48 hours, after which culture filtrate was isolated and secreted proteins were precipitated with TCA and pellet fractions were treated as above. SDS-Page followed by western blotting to nitrocellulose membranes were used to separate proteins. Staining of proteins was performed by the anti-PGRS antibody 7C4.1F7 (Abdallah et al. 2009), anti-GroEL2 antibody CS-44 (kind gift from J. Belisle, Colorado state University and the NIH, Bethesda, MD, USA), polyclonal anti-PPE41 (Abdallah et al. 2006), anti-EsxA antibody Hyb76-8 (Harboe et al. 1998) and an anti-hemagglutinin antibody (HA.11, Covance).

**LC-MS/MS analysis of *M. tuberculosis* culture filtrate**

*M. tuberculosis* strains SAWC_2088, SAWC_2135, SAWC_2701 and the SAWC_2135 and SAWC_2701 isolates complemented with pMV::ppe38-71 were grown as duplicates in liquid culture. Cultures were grown until all strains reached exponential growth with comparable optical densities. Subsequently, cells were washed and inoculated at 0.8 OD$_{600}$/ml in 100 ml 7H9 medium without BSA and grown for another 48 hours. Culture supernatant was collected and filtered. Culture filtrate proteins were concentrated to 1-2 ml, by centrifugation in Amicon® Ultra 3K centrifugal filter devices (Millipore, IRL), followed by TCA precipitation to reach a final volume of 100 μl. Protein samples were separated by SDS-PAGE, stained with Coomassie Brilliant Blue, subjected to in gel digestion, and analyzed by LC-MS/MS as previously described (Piersma, Warmoes, 2013). LC-MS/MS and protein identification and quantification were performed as described previously (Piersma et al. 2013). Briefly, Peptides were separated with a 25 cm x 75 μm ID fused silica C18 column (Dr Maisch GMBH, Ammerbuch-Entringen,
Peptides were trapped on a 10 mm × 100 μm ID trap column and separated at 300 nl/min in a 8–32% ACN in 0.5% HAc gradient in 90 min (120 min inject-to-inject). Eluting peptides were ionized at a potential of +2 kV into a Q Exactive mass spectrometer (Thermo Fisher, Bremen, Germany). Intact masses were measured at resolution 70.000 (at m/z 200) in the orbitrap. MS/MS spectra (top-10 precursors) were acquired at resolution 17.500 (at m/z 200).

Data analysis was performed by searching MS/MS spectra against the UniprotMycobacterium_tuberculosis_CDC1551.FASTA file (release October 2015; 4200 entries) using MaxQuant 1.5.2.8. (Cox and Mann 2008). Enzyme specificity was set to trypsin and up to two missed cleavages were allowed. Cysteine carboxamidomethylation was treated as fixed modification and methionine oxidation and N-terminal acetylation as variable modifications. Peptide precursor ions were searched with a maximum mass deviation of 4.5 ppm and fragment ions with a maximum mass deviation of 20 ppm. Peptide and protein identifications were filtered at an FDR of 1% using the decoy database strategy. Proteins were (label-free) quantified by spectral counting (Liu, Sadygov, and Yates 2004; Pham et al. 2010), i.e. the sum of all MS/MS spectra for each identified protein. For quantitative analysis across samples, spectral counts were normalized to the sum of the spectral counts per biological sample. Differential analysis of samples was performed using the beta-binominal test, which takes into account within- and between-sample variations, giving fold-change values and associated p-values for all identified proteins (Pham et al. 2010). Protein cluster analysis was performed using hierarchical clustering in R. The protein abundances were normalized to zero mean and unit variance for each individual protein. Subsequently, the Euclidean distance measure was used for protein clustering.

In vivo virulence assessment of M. tuberculosis strains in a BALB/c mouse model of pulmonary tuberculosis

Mice infection experiments were generally performed as previously described (Aguilar et al. 2010). Shortly, M. tuberculosis strains and mutants were grown until mid-logarithmic phase and 2.5 * 10⁵ bacteria were suspended in 100 μl PBS to create the infection inoculum. 6-8 week old BALB/c mice were anesthetized with sevoflurane and intratracheally inoculated. Forty-eight mice were infected per strain and were analyzed for survival. Six mice were sacrificed form each group at 1, 3, 7, 14, 21, 28, 60 and 120 days post infection and these mice were taken out of the analysis of survival curves at those time points. To determine bacillary loads, one lung per mouse was homogenized with a Polytron homogenizer (Kinematica) and was serially dissolved in isotonic saline and plated on 7H11 plates (Difco) with OADC supplement. Kaplan-Meier survival curves were created and analyzed by GraphPad Prism software version 6.0 for Windows. Mantel-Cox Log-Rank test was used to calculate significant differences between curves. Differences in CFUs between different strains were analyzed by two-way ANOVA followed by Tukey’s test for multiple comparisons.
Statistical calculations were performed in the GraphPad Prism software version 6.0 for Windows.

**ETHICS STATEMENT**
Animal studies were approved by the Institutional Ethics Committee of the National Institute of Medical Sciences and Nutrition “Salvador Zubiràn” in accordance to the guidelines of the Mexican national regulations on animal care and experimentation (NOM 062-ZOO-1999).

The collection of *M. tuberculosis* strains from human patients was approved by the Stellenbosch University Health Research Ethics Committee (approval reference number N10/04/126). Informed consent was not required as we retrospectively collected sputum cultures from the National Health Laboratory Service after routine processing. All patient identifiers were removed to ensure patient confidentiality.

**AUTHOR CONTRIBUTIONS**
Performed the experiments: LSA, RU, SRP, ADvdW, KvdK, MK, RH-P
Contributed to the manuscript: LSA, ENGH, RMW, RH-P, WB
Performed data analysis: LSA, AD, SRP, RMW, RH-P, WB
Contributed reagents and/or facilities: CRJ, RH-P, RMW

**ACKNOWLEDGMENTS**
We would like to thank Nico C Gey van Pittius, Ben Appelmelk, Joen Luijink and Astrid van der Sar for useful discussions and help with data interpretation.
We would like to thank Kim van der Kuij and Marion Sparrius for technical assistance.
We would like to thank Thang Phan for data analysis.
SUPPLEMENTAL INFORMATION

Supplementary procedures

*Synthesized DNA product ExsP_3-N_4*

CTACGCGTGAAAGCTGCGGGGGGTGATTGGTCGGGGTTTCGGGTTTGCGGTA
TTTGGTCCGGTCTGAATGTCTTTTTGGGAAAGGGGAATAGGCAGTGACTACTCG
GGTTATGACTGACCCGCATGCGATGCGGGACATGGCGGGTCGTTTTGAGATGC
ATGCGCAGACCGTGGATGAGGGCCGCAAGATGTTGGGCTCGTGCAGACA
ACATCGCAGCGCGGCTGAGCGCTGATGGCGCTCGGACTTCGCTGGACAT
CCATGGGTGATGAACACCGGCCCTCCGCAACATTGTGAACCATGCTGATG
GTGCCTGACCGGGTTTGTTCCGCGACCGCGAAACAGGACGAGCAAGAGCAG
GCTTCCGACGCAGGTTTGGGTAGCTAGAGGAGGTTGGGGGATGACGATCAA
TTACCATTCGCTGATGATGGCCATGAGGCGCTGATTCGGGCTCAGGCGG
CGGCGTTGAGGCTGACGTGAGCATCGGCGATGTGGTGTGGGCTGGCAG
GGATTTCGGGTGCGCGGGCTGGTGTCGCTGCCAGGAGTTCATCTCGCAG
TTGGGGCTAATTTCCAGGTGATCTACCCGAGGCAATGCCACCCAGGTCAAG
GGTGCAGGCTGCGGTAACAAACATGGCCAGCCACCGACTCCGCGGTGGGTTCG
AGCTGGGCATAGAGCTTATCGATACCG
**Table s4.** LC-MS/MS detected secreted proteins of *M. tuberculosis* isolates

<table>
<thead>
<tr>
<th>Protein name</th>
<th>MW kDa</th>
<th>% coverage</th>
<th>SAWC 2088</th>
<th>SAWC 2135</th>
<th>SAWC 2701</th>
<th>SAWC 2135-C</th>
<th>SAWC 2701-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>esxB</td>
<td>10.8</td>
<td>99</td>
<td>884.4</td>
<td>627.4</td>
<td>531.9</td>
<td>423.1</td>
<td>541.0</td>
</tr>
<tr>
<td>esxA</td>
<td>9.9</td>
<td>98.9</td>
<td>426.1</td>
<td>272.4</td>
<td>256.4</td>
<td>288.3</td>
<td>451.3</td>
</tr>
<tr>
<td>espB</td>
<td>47.6</td>
<td>55.7</td>
<td>144.6</td>
<td>39.4</td>
<td>92.1</td>
<td>58.6</td>
<td>84.5</td>
</tr>
<tr>
<td>mpt64</td>
<td>24.9</td>
<td>65.8</td>
<td>102.0</td>
<td>102.6</td>
<td>107.6</td>
<td>77.1</td>
<td>128.5</td>
</tr>
<tr>
<td>espA</td>
<td>39.9</td>
<td>31.4</td>
<td>99.4</td>
<td>133.5</td>
<td>133.7</td>
<td>142.0</td>
<td>141.9</td>
</tr>
<tr>
<td>esxK</td>
<td>11.0</td>
<td>95.9</td>
<td>84.6</td>
<td>100.4</td>
<td>125.7</td>
<td>78.8</td>
<td>143.0</td>
</tr>
<tr>
<td>espK</td>
<td>73.9</td>
<td>20.5</td>
<td>81.0</td>
<td>28.8</td>
<td>55.0</td>
<td>38.5</td>
<td>43.4</td>
</tr>
<tr>
<td>espE</td>
<td>42.1</td>
<td>19.4</td>
<td>61.1</td>
<td>17.5</td>
<td>34.4</td>
<td>25.4</td>
<td>32.8</td>
</tr>
<tr>
<td>espF</td>
<td>10.6</td>
<td>58.3</td>
<td>53.0</td>
<td>6.2</td>
<td>35.4</td>
<td>21.4</td>
<td>34.3</td>
</tr>
<tr>
<td>fbpC</td>
<td>36.8</td>
<td>38.2</td>
<td>48.8</td>
<td>28.7</td>
<td>52.7</td>
<td>35.9</td>
<td>61.1</td>
</tr>
<tr>
<td>mpt63</td>
<td>16.5</td>
<td>57.2</td>
<td>44.4</td>
<td>26.4</td>
<td>47.1</td>
<td>21.3</td>
<td>29.0</td>
</tr>
<tr>
<td>espC</td>
<td>10.8</td>
<td>28.2</td>
<td>39.5</td>
<td>17.5</td>
<td>31.7</td>
<td>28.8</td>
<td>31.0</td>
</tr>
<tr>
<td>esxN</td>
<td>9.9</td>
<td>71.3</td>
<td>39.4</td>
<td>40.0</td>
<td>63.2</td>
<td>43.8</td>
<td>64.5</td>
</tr>
<tr>
<td>PPE18</td>
<td>39.2</td>
<td>38.4</td>
<td>38.8</td>
<td>53.7</td>
<td>71.6</td>
<td>85.6</td>
<td>82.3</td>
</tr>
<tr>
<td>mpt53</td>
<td>18.4</td>
<td>39.9</td>
<td>30.8</td>
<td>22.0</td>
<td>20.6</td>
<td>16.2</td>
<td>32.0</td>
</tr>
<tr>
<td>espJ</td>
<td>27.4</td>
<td>29.3</td>
<td>28.5</td>
<td>3.0</td>
<td>18.8</td>
<td>12.3</td>
<td>19.7</td>
</tr>
<tr>
<td>fbpA</td>
<td>35.7</td>
<td>32.8</td>
<td>24.1</td>
<td>27.7</td>
<td>38.0</td>
<td>29.0</td>
<td>39.4</td>
</tr>
<tr>
<td>apa</td>
<td>32.7</td>
<td>32</td>
<td>21.6</td>
<td>38.1</td>
<td>44.7</td>
<td>35.9</td>
<td>43.7</td>
</tr>
<tr>
<td>PPE41</td>
<td>25.2</td>
<td>15.2</td>
<td>18.5</td>
<td>3.0</td>
<td>5.3</td>
<td>4.1</td>
<td>7.9</td>
</tr>
<tr>
<td>hhbA</td>
<td>21.5</td>
<td>29.6</td>
<td>17.9</td>
<td>13.5</td>
<td>7.0</td>
<td>10.9</td>
<td>1.2</td>
</tr>
<tr>
<td>mpt51</td>
<td>31.1</td>
<td>35.5</td>
<td>17.9</td>
<td>19.6</td>
<td>27.1</td>
<td>23.2</td>
<td>54.7</td>
</tr>
<tr>
<td>fbpB</td>
<td>34.6</td>
<td>24.9</td>
<td>16.7</td>
<td>26.0</td>
<td>25.1</td>
<td>23.2</td>
<td>30.5</td>
</tr>
<tr>
<td>PPE40</td>
<td>58.4</td>
<td>6</td>
<td>11.8</td>
<td>0.0</td>
<td>0.0</td>
<td>4.1</td>
<td>10.5</td>
</tr>
<tr>
<td>PPE52</td>
<td>39.1</td>
<td>15.2</td>
<td>9.3</td>
<td>0.0</td>
<td>0.0</td>
<td>6.1</td>
<td>9.3</td>
</tr>
<tr>
<td>PPE10</td>
<td>47.2</td>
<td>14.2</td>
<td>9.3</td>
<td>0.0</td>
<td>0.0</td>
<td>3.0</td>
<td>9.2</td>
</tr>
<tr>
<td>esxL</td>
<td>9.9</td>
<td>71.3</td>
<td>9.3</td>
<td>6.5</td>
<td>9.5</td>
<td>5.2</td>
<td>9.9</td>
</tr>
<tr>
<td>PE25</td>
<td>10.7</td>
<td>36.4</td>
<td>6.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.9</td>
</tr>
<tr>
<td>PE15</td>
<td>9.9</td>
<td>33.3</td>
<td>6.2</td>
<td>1.0</td>
<td>2.6</td>
<td>6.9</td>
<td>4.0</td>
</tr>
<tr>
<td>PPE68</td>
<td>37.3</td>
<td>7.6</td>
<td>6.2</td>
<td>1.1</td>
<td>3.5</td>
<td>0.9</td>
<td>2.7</td>
</tr>
<tr>
<td>PPE53</td>
<td>57.7</td>
<td>3.8</td>
<td>5.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
<td>4.0</td>
</tr>
<tr>
<td>PPE20</td>
<td>55.4</td>
<td>5.6</td>
<td>5.0</td>
<td>2.1</td>
<td>0.0</td>
<td>3.0</td>
<td>1.2</td>
</tr>
<tr>
<td>PPE19</td>
<td>39.7</td>
<td>19.8</td>
<td>4.3</td>
<td>3.2</td>
<td>8.6</td>
<td>3.9</td>
<td>8.6</td>
</tr>
<tr>
<td>PPE32</td>
<td>39.9</td>
<td>8.8</td>
<td>4.3</td>
<td>2.1</td>
<td>2.6</td>
<td>3.9</td>
<td>4.0</td>
</tr>
<tr>
<td>PE35</td>
<td>9.8</td>
<td>24.5</td>
<td>3.1</td>
<td>0.0</td>
<td>3.5</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>espL</td>
<td>12.2</td>
<td>20</td>
<td>3.1</td>
<td>3.0</td>
<td>1.8</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>PPE16</td>
<td>60.2</td>
<td>3.7</td>
<td>2.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>3.4</td>
</tr>
<tr>
<td>PE4</td>
<td>53.1</td>
<td>2.6</td>
<td>1.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>MT3449</td>
<td>127</td>
<td>0.7</td>
<td>1.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>PE_PGRS38</td>
<td>44.5</td>
<td>5.5</td>
<td>1.2</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
<td>4.0</td>
</tr>
<tr>
<td>PE_PGRS24</td>
<td>49.8</td>
<td>2.7</td>
<td>1.2</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
<td>2.7</td>
</tr>
<tr>
<td>PE_PGRS35</td>
<td>53.7</td>
<td>3.2</td>
<td>1.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.9</td>
</tr>
<tr>
<td>EsxH</td>
<td>10.4</td>
<td>19.8</td>
<td>1.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.8</td>
</tr>
<tr>
<td>EsxG</td>
<td>9.8</td>
<td>52.6</td>
<td>1.2</td>
<td>3.2</td>
<td>1.8</td>
<td>5.2</td>
<td>5.2</td>
</tr>
<tr>
<td>PPE36</td>
<td>27.5</td>
<td>21</td>
<td>0.0</td>
<td>22.8</td>
<td>0.0</td>
<td>21.6</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Table s4. (continued)

<table>
<thead>
<tr>
<th>Protein name H37Rv</th>
<th>MW kDa</th>
<th>% coverage</th>
<th>SAWC 2088</th>
<th>SAWC 2135</th>
<th>SAWC 2701</th>
<th>SAWC 2135-C</th>
<th>SAWC 2701-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPE51</td>
<td>38.0</td>
<td>12.6</td>
<td>0.0</td>
<td>14.3</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>PPE65</td>
<td>40.7</td>
<td>4.8</td>
<td>0.0</td>
<td>3.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>PPE26</td>
<td>40.6</td>
<td>11.4</td>
<td>0.0</td>
<td>4.1</td>
<td>0.9</td>
<td>3.0</td>
<td>1.9</td>
</tr>
<tr>
<td>EspD</td>
<td>19.8</td>
<td>6.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>PPE38</td>
<td>37.4</td>
<td>9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>4.1</td>
<td>15.3</td>
</tr>
<tr>
<td>PE_PGRS39</td>
<td>38.7</td>
<td>10.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>4.1</td>
<td>7.3</td>
</tr>
<tr>
<td>PE31</td>
<td>9.8</td>
<td>17.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>PE_PGRS29</td>
<td>32.9</td>
<td>9.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9</td>
<td>2.5</td>
</tr>
<tr>
<td>PPE33</td>
<td>47.2</td>
<td>2.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
<td>1.3</td>
</tr>
<tr>
<td>PE_PGRS59</td>
<td>37.0</td>
<td>5.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.9</td>
</tr>
<tr>
<td>PE_PGRS42</td>
<td>57.2</td>
<td>1.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>PE12</td>
<td>31.0</td>
<td>7.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>PE5</td>
<td>9.6</td>
<td>27.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.5</td>
</tr>
<tr>
<td>PPE64</td>
<td>61.6</td>
<td>4.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.2</td>
</tr>
<tr>
<td>PE22</td>
<td>10.4</td>
<td>12.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.1</td>
<td>0.0</td>
</tr>
</tbody>
</table>
EXPLOITING ESX-5: OPTIMIZING LIPY AS A CARRIER FOR HETEROLOGOUS SECRETION IN MYCOBACTERIA

Louis S. Ates¹, Maroeska Burggraaf¹, Kim van der Kuij¹, Coen Kuijl¹, Edith Houben² & Wilbert Bitter¹²
ABSTRACT
Heterologous secretion in mycobacteria can be a valuable tool in vaccine development and fundamental research, but thus far attempts to achieve it have been limited in success and/or reproducibility. In this study, we initially examined whether heterologous protein domains could be secreted when coupled to a diverse set of mycobacterial secreted proteins that are probably transported via different pathways. However, no secretion for any of the chimeric proteins could be confirmed. Therefore, a more detailed approach was taken to establish heterologous secretion by the ESX-5 secretion system using the efficiently secreted lipase LipY as a carrier. Truncations in the linker domain that connect the N-terminal PE domain with the lipase domain of LipY impaired secretion and lipase activity. Therefore, it was concluded that the complete linker domain as well as the PE domain, which contains recognition motifs for secretion, should be used when using LipY as a secretion carrier. This knowledge was used to create a chimeric construct of the first 205 amino acids of LipY containing the PE and linker domain fused to a synthetic protein fragment containing the dominant epitopes of ovalbumin (OVA). This chimeric protein (LipY-OVA-HA) was secreted to the cell surface and culture filtrate of M. marinum in an ESX-5 dependent manner, albeit with low efficiency. To identify sequences that limited transport of this heterologous construct, random mutagenesis was used, resulting in the identification of multiple clones with enhanced LipY-OVA-HA secretion efficiency. Two of these clones exhibited highly improved secretion efficiency due to mutations in the stop codon of the chimeric protein, resulting in an extended C-terminus. These data show that LipY can be used as a secretion carrier for heterologous secretion and that random mutagenesis is a viable approach to optimize secretion of chimaeric proteins.

1Department of medical microbiology and infection prevention, VUMC medical centre, Amsterdam the Netherlands.
2Section Molecular Microbiology, Amsterdam Institute of Molecules, Medicine & Systems, Vrije Universiteit, Amsterdam, the Netherlands.
INTRODUCTION

Species of the genus *Mycobacterium* are responsible for many diseases in humans and animals. *Mycobacterium tuberculosis* is still considered the most deadly bacterial pathogen worldwide, killing approximately 1.5 million people each year (World Health Organization 2013). Although there is no fully protective vaccine against tuberculosis, there is a live vaccine that offers a partial protection against the most severe forms of tuberculosis (Abubakar et al. 2013). This strain, *Mycobacterium bovis* Bacillus Calmette-Guérin (BCG), is an attenuated strain of *M. bovis* that was created by serial culturing in vitro. Although BCG is only able to confer partial protection against tuberculosis, it is still administered in many countries, especially in regions with high prevalence. With 3 billion people vaccinated in the last century, BCG is the most widely used vaccine worldwide (Bloom and Fine 1994).

Isolated cell wall components of BCG are already highly immunogenic and induce high amounts of pro-inflammatory cytokines upon administration. However, the strongest immune responses occur when live BCG bacteria are used for vaccination (Youmans and Youmans 1969; Trujillo-Vargas et al. 2005). BCG is attenuated, but fully able to infect host cells and persist there, leading to prolonged activation of both cellular and humoral immune responses. However, since BCG is not able to disseminate, it is generally recognized as safe for immuno-competent individuals.

The immunogenic capacities of BCG are also used in treatment of non-muscle invasive superficial bladder cancer. After resection of tumor tissue, multiple instillations of intra-vesical BCG are administered, which cause inflammatory reactions in the bladder (Babjuk et al. 2011). This intra-vesicular BCG helps to prevent outgrowth of remaining tumor cells with similar or higher success rates than traditional chemotherapy and is the standard treatment for these cases in many countries (Gontero et al. 2010; Babjuk et al. 2011). Although BCG has many clinically important capacities that are exploited in the treatment or prevention of different diseases, many of these applications are still sub-optimal and could be improved. One way to improve the efficiency of BCG as a vaccine is to engineer the bacterium to produce and secrete heterologous proteins or protein epitopes. For instance, several studies have attempted to improve BCG by expressing human cytokines to skew immune responses in a particular direction (Bastos et al. 2009). People also tried to improve the immunogenic properties of BCG by introducing antigenic epitopes from HIV or *Plasmodium* (reviewed in Bastos et al. 2009). Different studies have provided evidence that different localization of proteins can lead to different immune responses (Stover et al. 1993; Bastos et al. 2009; Hayward et al. 1999). For instance, secreted or cell-wall associated antigens are thought to induce a stronger antibody response than intracellularly expressed proteins (Hayward et al. 1999; Stover et al. 1993). Therefore, the cellular localization of the final heterologously expressed product is an important parameter for applications. However, mycobacterial protein secretion is markedly different from that of typical Gram-positive and Gram-negative bacteria and the establishment of successful heterologous secretion in mycobacteria has therefore
been difficult. Efficient secretion of heterologous proteins in BCG probably requires more detailed understanding of mycobacterial protein secretion systems in general.

Mycobacteria possess a diderm cell-envelope consisting of an inner membrane and a highly specific outer membrane, predominantly consisting of long chain (C_{60-90}) mycolic acids (Zuber et al. 2008; Sani et al. 2010; Patrick J Brennan and Crick 2007; Hoffmann et al. 2008). In order to secrete proteins over this unusual cell envelope, mycobacteria have developed specialized secretion systems. These secretion systems have been classified as type VII secretion (T7S) systems and up to five of these systems are present in mycobacteria (reviewed in Abdallah et al. 2007). These systems have been called ESX-1 to ESX-5 (Bitter et al. 2009). Together these ESX systems are responsible for the secretion of dozens of substrates necessary for the growth and virulence of the mycobacteria. These secreted substrates belong to different protein families such as the Esx proteins and the PE and PPE proteins, the latter two named after the conserved proline and glutamic acid residues in their N-terminus (M. J. Brennan and Delogu 2002; Gey van Pittius et al. 2006). In addition to the T7S systems, proteins containing the canonical N-terminal signal sequence can be found in culture supernatants, suggesting that also a transport mechanism of substrates of the classical Sec and Tat systems over the outer membrane exists. Finally, mycobacteria also possess a second copy of the motor protein of the Sec system, i.e. SecA2, that could be involved in the secretion of extracellular and cell surface proteins that lack a signal sequence (Ligon, Hayden, and Braunstein 2012; van der Woude et al. 2014).

In this study, we attempted to secrete a human cytokine or a tumor associated protein by fusing these protein sequences to different mycobacterial cell surface proteins. This approach was unsuccessful, requiring us to investigate heterologous secretion in a more detailed manner. We decided to utilize a T7S substrate, secreted via ESX-5, to target heterologous proteins for secretion in the model organism M. marinum. This substrate, LipY, is a lipase that has been shown to be efficiently secreted in different species of Mycobacteria (Mishra et al. 2008; Daleke et al. 2011). LipY of M. tuberculosis consists of three domains, an N-terminal PE domain, a linker domain and a C-terminal lipase domain. While the PE domain contains the general T7S signal (Daleke, Ummels, et al. 2012) and is targeted for secretion by the ESX-5 system, the linker domain is cleaved by an unknown protease upon secretion. The C-terminal part of the linker domain together with the lipase domain remain attached to the cell surface (Daleke et al. 2011). In this study, we investigated the role of the linker domain of LipY in the secretion process, by making truncations in this region. We subsequently used the obtained knowledge to establish secretion of a LipY chimera and further improved its secretion by exposing the construct to random mutagenesis.
RESULTS

Heterologous secretion by fusion of GCSF or MART-1 to mycobacterial secreted proteins

We first tried to secrete the human cytokine granulocyte colony-stimulating factor (GCSF) and the melanoma antigen recognized by T-cells (MART-1) by different mycobacterial species (Kawakami et al. 1994). This was attempted by cloning the genes encoding these proteins after different genes of mycobacterial secreted proteins. Additionally, an optimized version of MART-1 was constructed, which was called MART_{opt}. This construct consisted of the MART-1 gene codon-optimized for expression in *Mycobacterium bovis* BCG and had 4 cysteine residues changed to alanine residues to avoid disulfide bond formation. Furthermore, a stretch of four hydrophobic amino acids (LLLI) that are part of the hydrophobic transmembrane region of MART-1 were deleted to avoid protein aggregation or membrane incorporation. Selected secretion carriers were mpt64 (*Rv1980c*) that contains an N-terminal signal sequence, the lipoprotein LpqH (*Rv3763c*), the ESX-5 substrate LipY (*Rv3903c*) and HbHA (*Rv0475c*) that does not have any clear secretion signals of *M. tuberculosis*. For the lipoprotein LpqH only the described signal sequence was used (Langermann et al. 1994; Stover et al. 1993; Hayward et al. 1999) and for LipY only an N-terminal fragment representing amino acids 1-158 was used, since cleavage of LipY upon secretion happens between residues 149 and 150 (Daleke et al. 2011). Since the requirements for transport across the outer membrane of signal sequence containing proteins is unknown and no defined secretion signal is known HbHA, the complete coding sequence of mpt64 and hbha were used as a secretion carrier. The chimaeric genes were cloned into pSMT3:eGFP and expressed under control of a constitutively active hsp60 promoter. Plasmids were transformed to *Mycobacterium smegmatis*, *Mycobacterium marinum* E11 and *M. bovis* BCG (Copenhagen strain). Using an anti-MART-1 antibody A103 we could not detect any expression of any of the chimeric proteins for all the species examined (data not shown), except for the constructs LpqH-MART-1 or LpqH-MART_{opt}. *M. smegmatis* expressing LpqH-MART-1 produced a protein product with the expected apparent molecular weight of 24 kDa and a 15 kDa product, probably representing a degradation product (Figure 1A). Also an LpqH fusion with the optimized MART domain could be detected, with clearly less of the 15kDa degradation product. Since LpqH-MART-1 and LpqH-MART_{opt} were expressed at similar quantities, codon-optimization, removal of a stretch hydrophobic amino acids and cysteines of MART-1 did not seem to lead to higher levels of protein production, but did seem to create a more stable chimaeric protein (Figure 1A). Next, we tested the localization of these chimeric products by Genapol X-080 extraction of surface proteins, proteinase K treatment of intact bacteria or by examining the culture filtrate. Unfortunately, no secretion or surface localization of LpqH-MART-1 or LpqH-MART_{opt} could be detected using these methods in *M. smegmatis* or *M. marinum*. In *M. bovis* BCG the 15 kDa product appeared to be partly susceptible for proteinase-K treatment, indicating that this product might be surface localized.
Figure 1. Codon-optimization of LpqH-MART-1 leads to higher protein stability, but not higher expression or secretion. The secretion sequence of *M. tuberculosis* LpqH was fused to full-length MART-1, or the optimized MART-1 (MART<sub>opt</sub>). A) Expression of LpqH-MART-1 and LpqH-MART<sub>opt</sub> in *M. smegmatis* led to expression of a full-length chimaeric protein product (white arrow) and a degradation product (black arrow). Lower amounts of the degradation product were observed in the MART<sub>opt</sub> construct, indicating increased protein stability. Genapol treated cells (GP) had similar levels of protein compared to whole-cell lysates (P) indicating that the chimaeric protein was not surface localized in *M. smegmatis*. B) Expression and secretion analysis of LpqH-MART-1 and LpqH-MART<sub>opt</sub> in *M. bovis* BCG. The degradation products (white arrow) of MART-1 and MART<sub>opt</sub> were more susceptible to proteinase K treatment (K), but were not extracted by Genapol-X-080 (GS).

(Figure 1B). However, this product was not Genapol-extractable (Figure 1B) and surface labeling by ELISA or spotblot assays was unsuccessful, due to background staining of the A103 antibody (data not shown). We therefore concluded that we were not able to prove that these chimeric proteins were secreted or surface localized.

**Investigation of the linker domain of LipY.**

Previous study by our group (Daleke et al. 2011) has revealed that the PE domain of LipY from *M. tuberculosis* is necessary for targeting this protein to the ESX-5 secretion machinery. Upon secretion, the linker domain is cleaved after residue 149. The processed form of LipY then remains attached to the cell-surface in the capsular layer. The linker region of LipY is specific for this protein and has no homologues in other ESX-5 substrates or any other proteins. However, the function of this domain is not known. Since fusion of GCSF or MART-1 to the first 158 amino-acids of LipY did not result in expression or secretion of these proteins we investigated the role of this linker domain in secretion and define the minimal domain required. Because processing at position 149 was shown to be important for secretion (Daleke et al. 2011), the N-terminal and C-terminal cleavage site-flanking sequences were truncated resulting in LipYΔ100-145 and LipYΔ158-205, respectively.
Figure 2. Truncations in the linker domain of LipY impair secretion and lipase activity. A) Schematic representation of HA-tagged LipY (Rv3097c) consisting of the PE (Dark Grey), linker (White) and lipase (Light grey) domains. The linker domain is cleaved at amino acid positions 148-150 (Black arrow) upon secretion. Truncations of indicated amino-acids were made in pSMT3::LipY-HA and were expressed in M. marinum which was used to assess secretion (B) and lipase activity (C). B) Secretion analysis of LipY-HA truncation constructs. The indicated truncations were expressed and surface localization was analyzed by Genapol X-080 extraction. Full length LipY-HA and M. marinum wild-type (empty) were used as positive and negative control respectively. Black arrows indicate the full-length protein products. Low levels of surface localized LipY (GS) were detected in the ∆158-180 and ∆158-205 constructs, while the ∆181-205 surface localization was similar to the full-length construct. In ∆100-145 no full-length LipY-HA was detected in the intracellular fraction (GP) indicating that the PE domain is always cleaved of, but no surface localization (GS) was detected. C) Lipase activity of M. marinum expressing the indicated truncations of LipY-HA. Lipase activity was measured after 30 min. incubation of M. marinum cells with 4-Nitrophenyl Stearate substrate followed by absorbance measurement of the supernatant at 405nm. Three independent measurements (black dots) were taken, error bars indicate the standard deviation. These data are representative of three individual experiments.
(Figure 2A). Additionally, two smaller truncations were made in the C-terminal domain of the linker, based on sequence homology or divergence of various LipY molecules from different mycobacterial pathogens. This created LipYΔ158-180 (divergent region) and LipYΔ181-205 (homologous region), respectively (Figure 2A). Secretion of the full-length and truncated LipY-HA constructs was analyzed by Genapol X-080 extraction of surface-localized proteins followed by immunoblotting. As seen before, the full-length construct was efficiently processed and Genapol X-080 extractable (Figure 2B). The truncated protein LipYΔ100-145 was detected at a molecular weight of 25 kDa, indicating that this truncated protein is processed at the right position. Surprisingly however, the processed form of LipYΔ100-145 was not detected in the surface-enriched protein fraction (Genapol supernatant), but could only be observed in the cell pellet fraction. Less efficient processing, but more surface-localized LipY could be observed for LipYΔ158-205. The Genapol-extractable material contained the processed and secreted form of around 23 kDa, as well as the ~39 kDa product that is also observed for LipY-HA. Additionally, some full-length surface-localized LipY could be observed. If we now look at the behavior of the two smaller deletions in this second part of the linker, we observe that there is a pronounced secretion defect in LipYΔ158-180, whereas LipYΔ181-205 shows a secretion pattern very similar to that of the full-length construct (Figure 2B). Therefore, the first part of this domain seems to be most important for secretion.

To investigate whether secreted LipY was properly folded and therefore functional as a lipase, the activity of the lipase domain was assessed as described previously (Daleke et al. 2011). Surprisingly, the data of the lipolytic assay did not fully correspond to what was observed in immunoblotting experiments (Figure 2C). The strongest lipolytic activity was observed in the full-length LipY-HA construct (1.86 +/- 0.006). Strikingly, cells with the LipYΔ100-145 constructs showed moderate lipase activity, whereas we could previously not detect this protein in the surface-enriched protein fraction. The large deletion of the second linker domain, i.e. LipYΔ181-205, showed reduced but significant lipase activity (0.53 +/- 0.04). The two other deletions in the second linker domain showed a more dramatic effect, since both LipYΔ158-205 (0.21 +/- 0.008) and LipYΔ158-180 (0.179 +/- 0.01) showed no activity. In fact, lipase activity in these cells was in fact even below that of wild-type M. marinum and closer to the ESX-5 deficient strain (Daleke et al. 2011). Although LipY is the most active secreted lipase (Deb et al. 2006), Mycobacteria produce many different lipases. Because wild-type M. marinum cells show higher lipase activity than the esx-5 mutant, we previously suggested that also other lipases are secreted via this pathway. Now we see that two LipY truncations have a reduced lipase activity, this could mean that these constructs in fact disturb secretion through ESX-5. To investigate this, protein samples were analyzed for the expression and secretion of PE_PGRS proteins (Supplemental Figure S1A). Indeed, cells expressing LipYΔ158-205 and LipYΔ158-180 exhibited impaired secretion of PE-PGRS proteins to the cell surface (Genapol supernatant) or culture filtrate, while the other constructs did not show any
Figure S1. PE_PGRS secretion profiles of strains expressing different LipY constructs. A) Immunoblot of PE_PGRS proteins secreted by M. marinum expressing the indicated LipY-HA truncations. Expression and secretion of PE_PGRS proteins is markedly lower in strains containing ∆158-205 & ∆158-180 indicating that expression of these constructs might interfere with ESX-5 secretion. B-D) Immunoblots of PE_PGRS expression and secretion by strains expressing indicated mutated LipY-OVA constructs. Genapol treated cells (B), Genapol supernatant (C) and culture filtrate (D) fractions do not show marked differences in ESX-5 dependent PE_PGRS secretion.

These data indicate that the 158-180 region of the linker sequence of LipY is important in ESX-5 secretion and that truncation of this domain blocks ESX-5 secretion. Because all truncations in the linker domains of LipY resulted in lower secretion levels or reduced lipase activity, the following heterologous secretion experiments were performed with the complete PE and linker domains of LipY.

LipY as a carrier for heterologous secretion of OvaL

In order to provide a proof of concept for heterologous secretion via ESX-5, a chimaeric protein was made consisting of the PE and linker domain of M. tuberculosis LipY fused to ovalbumin fragment from Gallus gallus. This synthetic ovalbumin fragment, called OvaL, contains the major immunogenic epitopes of ovalbumin, but is smaller in size (99 amino acids) than the previously used heterologous proteins and lacks cysteine residues or hydrophobic regions that could interfere with secretion. The resulting heterologous protein, called LipY-OVA also contains a hemagglutinin tag at the C-terminus to facilitate detection (Figure 4A, Figure 4B).

LipY-OVA was expressed in M. marinum wild-type (WT) strain E11 and in the ESX-5 secretion-deficient mutant 7C1 (Abdallah et al. 2009; Daleke, Woude, et
Surface localization and secretion of LipY-OVA was analyzed by Genapol X-080 extraction and precipitation of culture supernatant, respectively, followed by immunoblotting using an anti-HA antibody. The expected full-length chimaeric protein, with an apparent molecular weight of ~30 kDa, could be detected in the cell pellet fractions of both WT and 7C1 strains (Figure 3). Also two smaller forms of LipY-OVA could be detected, with an apparent molecular weight of ~16 kDa. These bands could represent processed chimaeric protein. As expected, these putative processed forms were only present at the cell-surface or supernatant of WT cells, confirming that LipY-OVA is secreted in an ESX-5 dependent manner. However, secretion was inefficient.

Optimizing secretion of LipY-OVA by random mutagenesis

Since LipY-OVA was expressed in high amounts, but was secreted only in relatively small amounts, we performed random mutagenesis to increase the secretion efficiency. An error-prone DNA-polymerase (Biles and Connolly 2004) was used to amplify LipY-OVA and subsequently cloned in the pSMT3expression vector. A library of approximately 1,000 clones was electroporated in M. marinum E11 and screened for improved secretion via a double filter assay (Abdallah et al. 2006; van der Woude et al. 2012). Nine colonies that showed improved secretion (Figure 4C) were selected for further analysis. The secretion improved LipY-OVA constructs contained between one to five non-synonymous single nucleotide polymorphisms (SNPs) per construct (Figure 4C). Striking was a 2 bp deletion or 1 bp insertion in the 3’ end of the construct of mutants KL1, KL3, KL5 & KL7, which caused a similar frameshift mutation.

Figure 3. LipY-OVA is secreted in an ESX-5 dependent manner. Immunoblots of secretion analysis of wild-type M. marinum, or a transposon mutant in the ESX-5 specific chaperone espG, expressing pSMT3::LipY-OVA-HA. Low amounts of full-length and processed LipY-OVA-HA could be detected in the cell-surface enriched (GS) or culture filtrate (S) fractions of wild-type M. marinum, but not in the ESX-5 secretion mutant espG::tn. Staining for GroEL2 was used as a loading and lysis control.
Figure 4. Random-mutagenesis of LipY-OVA-HA identified several constructs with increased secretion. Error-prone PCR was performed on Lip-OVA-HA, after which mutated PCR products were ligated into an empty pSMT3 vector and transformed to *M. marinum*. ~1000 colonies were screened for HA secretion by double filter analysis, which resulted in the identification of 9 supersecretor mutants that were further analyzed. A) Schematic overview of the LipY-OVA-HA chimeric protein. B) The protein sequence of the chimeric protein LipY-OVA-HA, with the respective domains indicated on the right. Highlighted in grey are amino acids that were mutated in at least one of the identified supersecretor constructs, indicating that no ‘hotspots’ for mutations with increased secretion could be identified. B) Double filters stained with an anti-HA antibody (top row) and the corresponding colonies on the top filter (bottom row) of the non-mutated LipY-OVA-HA (‘WT’) or the nine identified supersecretor strains (KL1-KL14). Indicated below the strains are the specific amino acid substitutions in respectively the LipY (above black line) and the OVA-HA (Below black line) regions of the chimaeric proteins. *: HA+ indicates an insertion of 1 bp or a deletion of 2 bps resulting in an alternative stop codon increasing the length of the protein by 19 amino acids.

in all these strains leading to an extended C-terminus of 18 or 19 amino-acids (Figure 4B, indicated with HA+).

To analyze the LipY-OVA supersecretion clones in more detail, we performed secretion analysis. As seen previously (Figure 3), LipY-OVA was partially detected in the Genapol extractable material, suggesting that this protein is surface localized
Corresponding to the frameshift mutations, the mutant strains KL1, KL3, KL5 and KL7 showed a protein product with a somewhat higher apparent molecular weight as compared to the original LipY-OVA construct (Figure 5). An increase in surface localization compared to the non-mutated LipY-OVA could be observed for strains KL3, KL11 and KL14 (Figure 5C), whereas strains KL1, KL3 and KL10 showed a markedly improved secretion in the culture supernatant (Figure 5D). Of these clones, KL10 exhibited supersecretion of a ~35 kDa protein band instead of the expected processed form of 16kDa. PE_PGRS protein secretion was used as a loading control and was similar in all samples (Supplemental Figure S1 B-D).

To test whether mutations in the LipY sequence or in the OVA-HA sequence were responsible for the phenotype of supersecretion, swap constructs were created. To create these swap constructs, the non-mutated LipY sequence was ligated to mutated OVA-HA sequences (OVA-1, OVA-3 & OVA-10) or vice versa (LipY-1, LipY-3 & LipY-10). The supersecretor phenotype was evident for *M. marinum* transformed with plasmids OVA-1 and OVA-3, but not for the other constructs, indicating that the mutations in the OVA-HA part of the protein are responsible for the supersecretion phenotype of these strains (Figure 6A, 6B). These results indicate that the extended C-terminus of the KL1 and KL3 constructs positively influence secretion of LipY-OVA-HA. For genomic *M. marinum* mutations responsible for the phenotype of supersecretion, mutated plasmids were isolated from strains KL1, KL3 or KL10. These isolated plasmids were re-transformed to wild-type *M. marinum* and were checked for supersecretion by double-filter analysis. The supersecretion phenotype associated with these plasmids was confirmed for the strains containing mutated plasmids pKL1 and pKL3, but not pKL10 (Figure 6A, B). Together these results show that, when secretion of heterologous proteins is sub-optimal, random mutagenesis combined with a double-filter screening of mutants is a powerful tool to identify mutants with improved secretion. This led to the finding that a C-terminal extension on the heterologous protein product has a positive effect on its secretion.

**DISCUSSION**

In this study, we set out to achieve heterologous secretion of the tumor-associated protein MART-1 or the human cytokine GCSF by mycobacteria. Fusion of these heterologous proteins to sequences of secreted mycobacterial proteins, some of which have been utilized and published before for heterologous export (Stover et al. 1991; Bastos et al. 2009), did not lead to detectable secretion in *M. smegmatis*, *M. marinum* or *M. bovis* BCG. This could be due to the nature of the selected heterologous proteins. MART-1 is a transmembrane protein with one classical hydrophobic transmembrane domain. Secretion of MART-1 via the well-established HBP-autotransporter heterologous secretion platform (Jong et al. 2012; Jong et al. 2014) was also unsuccessful (Jong & Luirink, personal communication). However, we also tested an optimized MART-1 with reduced hydrophobicity and also this form was not secreted. The other substrate, GCSF, is a soluble and largely...
Figure 5. Secretion analysis of LipY-OVA-HA supersecretion mutants. The identified *M. marinum* supersecretion mutants expressing mutated LipY-OVA-HA constructs were grown in liquid culture. Secretion analysis was performed by analyzing whole cell lysates (A), Genapol-X-080 treated cells (B), the surface enriched Genapol-X-080 extractable proteins (GS) and culture filtrate (D) by immunoblotting using an anti-HA antibody (A-D), or an anti-PGRS antibody as a control (Supplemental Figure 1 B-D). A protein product with increased molecular weight was detected in strains KL1, KL3, KL5 and KL7 corresponding with the 3' insertions or deletions leading to an alternative stop codon. Improved secretion to the culture filtrate was detected for strains KL1, KL3 and KL10 (D).

hydrophilic protein. This protein however does have two disulfide bonds and erroneous di-sulfide bridge formation could be responsible for misfolding and/or secretion defects. The only chimeric protein that resulted in successful expression was the fusion between the LpqH signal sequence and the MART-1 or MART_{opt} heterologous proteins.
Figure 6. Supersecretion phenotype of strains KL1 and KL3 is due to mutations in OVA-HA and not to the LipY secretion sequence. Colonies (A) and double filter analysis (B) of M. marinum expressing non-mutated LipY-OVA-HA or the plasmids isolated from colonies KL1 (pKL1), KL3 (pKL3) or KL10 (pKL10). Swaps of either the mutated LipY or the OVA-HA and their respective wild-type genes were made, resulting in plasmids with non-mutated LipY, but mutated OVA-HA (OVA-1, OVA-3 & OVA-10), or with mutated LipY, but non-mutated OVA-HA (LipY-1, LipY-3 & LipY-10). Supersecretion was clearly detected in strains expressing pKL1, pKL3, OVA-1 and OVA-3, indicating that the mutations in the OVA-HA part of this chimeric protein are responsible for the supersecretion phenotype.

Unfortunately, we could not show any definite prove of secretion for these proteins, due to methodological difficulties. The LpqH signal sequence is presumed to localize proteins to the membrane of M. bovis BCG, but this is often based on fractionation studies, which detect LpqH fusions in the insoluble fraction (e.g. Hayward et al. 1999; Stover et al. 1993). This fraction contains membrane proteins, but misfolded and therefore aggregated proteins are also expected to end up in this fraction and are not by definition membrane localized. Similarly, the increased proteinase K susceptibility of the LpqH-MART-1 and LpqH-MARTopt degradation products is not enough evidence to confirm that these chimaeric proteins are surface localized.

In the second part of this study, we studied the effect of LipY deletions on secretion. All four truncation mutants showed a reduced secretion compared to the full-length construct, indicating that the complete PE and linker domain should be included when LipY is used as a carrier for heterologous secretion. Interestingly, this analysis did result in the identification of two constructs that showed a broad ESX-5 secretion defect. One explanation for this observed phenotype is that these truncated substrates are recognized by the ESX-5 system, but are then trapped in secretion system. Heterologous secretion via ESX-5 was subsequently achieved by fusing the N-terminal 205 amino acids of LipY to a synthetic ovalbumin fragment. The chimaeric protein LipY-OVA was expressed in high amounts and secreted in an ESX-5 dependent manner. However, secretion of the chimaeric protein was inefficient.
Improved secretion efficiency could be achieved by random mutagenesis combined with a double-filter assay. This approach could be the way forward to increase secretion of heterologous constructs. However, it is very important to ensure that the epitopes of interest remain intact, when using this approach. Strikingly, the mutants with the strongest and most reproducible secretion phenotype both contained frame-shift mutations in the sequence coding for the HA-tag, which caused an alternative stop codon in these constructs. The role for this extended C-terminus in the detection or secretion of these constructs remains to be further investigated. Perhaps the charged nature of the HA tag interferes with secretion and is neutralized by the extended C-terminus of these mutants.

Development of a reliable platform for heterologous secretion is an important and perhaps underestimated tool in mycobacterial research. The immune-stimulating potential of BCG, as well as the experience in producing and administering this bacterium in a safe way make it a highly interesting vaccine-development tool. Strong humoral immune responses are achieved by heterologous protein expression in BCG (Medeiros et al. 2005; Bastos et al. 2009; Stover et al. 1993). However, cellular immune responses against proteins expressed in BCG are often inefficient. One possible explanation is that secreted, surface-exposed or cell-wall associated antigens might induce different immune responses as compared to cytosolically expressed proteins. This has been tested in BCG by fusion of proteins to different carriers for protein localization (Bastos et al. 2009; Stover et al. 1993; Hayward et al. 1999; Langermann et al. 1994; Matsumoto et al. 1998). However, in these studies it is impossible to discern the effect of protein targeting from immuno-stimulating capacities of the carrier sequence. Evidence that secretion is important in triggering a strong cellular immune response comes from studies of other intracellular pathogens. The strongest cellular (CD8+ T-cell) immune responses against intracellular bacterial pathogens such as *Listeria monocytogenes* or *Salmonella* sp. are found against secreted proteins (reviewed in Pamer 1998). Additionally, vaccination studies using *Salmonella typhimurium* as a vaccine vehicle have shown that cellular immune responses against heterologous proteins require secretion of these proteins (Hess et al. 1996). In analogy, heterologous secretion of antigens is expected to be an important aspect of developing BCG as a vehicle for vaccination against other diseases.

Although cell-wall targeting is usually achieved by fusion to the 19 kDa antigen and extracellular secretion is achieved by fusion to the secretion signal of Ag85b (reviewed in (Bastos et al. 2009)), no reliable secretion signal is known for surface localization in BCG. The LpqH (19 kDa lipoprotein) is often cited as a secretion carrier for surface localization. However, this is often shown by localization in the insoluble protein fraction, which is not necessarily membrane localization and also contains aggregated proteins. Even if membrane localization occurs, this would not necessarily imply surface localization, since it could also be inner membrane or periplasmic localization.
Therefore, the use of the better defined surface localized protein LipY as a carrier for heterologous proteins could be of added value. Another promising application for heterologous secretion in Mycobacteria is in the treatment of non-muscle invasive bladder cancer (reviewed in Begnini et al. 2015). Ongoing research in our group focuses on improvement of BCG for the treatment of bladder cancer by aiming to secrete human proteins that specifically drive tumor cells into apoptosis. Finally, heterologous secretion can also be applied to the research on Mycobacterial virulence mechanisms. For instance, the ESX-1 system in pathogenic mycobacteria is deleted in BCG due to loss of the RD1 region in serial culture (Pym et al. 2002). BCG strains that do secrete the ESX-1 substrate EsxA (ESAT-6) after reconstitution with the esx-1 locus, are known to confer stronger protective immunity against tuberculosis in animals (Pym et al. 2003). However, these strains are also more virulent and therefore not recognized as safe for human use. BCG secreting specific immunogenic ESX-1 substrates via the ESX-5 system could be an improvement for current BCG vaccination without impairing the safety of vaccination. This could also be a way to study the role of individual ESX-1 substrates in virulence in more detail, since substrates of the ESX-1 system are interdependent on secretion (Fortune et al. 2005).

This study illustrates some of the difficulties associated with heterologous secretion in mycobacteria. Although many beneficial effects of expressing and secreting heterologous proteins in BCG have been shown in individual studies, a reliable platform for heterologous secretion in mycobacteria is still lacking. Although we have not been able to provide this reliable platform in this study either, these results could be a step forward to achieve such a secretion platform.

**MATERIALS AND METHODS**

**Strains and culture conditions**

*M. marinum* E11 (van der Sar et al. 2004) was used as a wild-type strain of *M. marinum* in this study. Additionally the ESX-5 deficient mutant strain 7C1 was used, which contains a transposon insertion in the 5’ region of *mmar_2767* (Abdallah et al. 2009). *M. smegmatis* mc²155 and *M. bovis* BCG strain Copenhagen were used for expression of heterologous proteins. Plate cultures were performed on 7H10 agar plates supplemented with OADC (BD bioscience). Liquid cultures were performed in liquid Middlebrook 7H9 medium (Difco) supplemented with 10% albumin-dextrose-catalase supplement (BD bioscience) and 0.05% Tween-80. Appropriate antibiotics were added as selection markers (50μg/ml hygromycin and/or 25μg/ml kanamycin). All cultures were grown in Erlenmeyer flasks while shaking at 150 rpm at 30°C. *Escherichia coli* DH5α was used for cloning experiments and was grown in Luria-Bertani broth or plates at 37°C.
Cloning and constructs

GCSF was amplified from plasmid pORF-hGCSFb (Purchased from Invivogen, CA, USA) with primers GCSF_F and GCSF_R. MART-1 was amplified from plasmid LS7, containing MART-1 cDNA (kindly supplied by Dr. Erik Hooijberg) using primers MART-1_F and MART-1_R. Optimized MART-1 was synthesized by Geneart (Life technologies, Paisley, United Kingdom). GCSF, MART-1 or MART-1_{opt} were cloned into pSMT3::eGFP linearised by restriction enzymes EcoRV and HindIII. The different secretion domains of LipY, Mtp64, LpqH or HBHA were amplified with the respective primer pairs shown in Table 1 using H37Rv genomic DNA as a template. Chimaeric genes were created by ligating the respective signal sequence amplification products in pSMT3::MART-1, pSMT3::MART-1_{opt} or pSMT3::GCSF using restriction enzymes EcoRV and NheI.

pSMT3::lip_{Yub}-HA was previously described (Daleke et al. 2011) and used as the template for amplification of LipY truncation constructs. In order to create

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence 5’ → 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCSF_F</td>
<td>AAAGATATCGCTAGCTCCCTGCCCACAGGTTT</td>
</tr>
<tr>
<td>GCSF_R</td>
<td>GGGAGACCTTCAAGGCTTTTCTCAGTCT</td>
</tr>
<tr>
<td>HetLipY_F</td>
<td>AAAGATATCGTTTTGCGTCGGAGAGTCTT</td>
</tr>
<tr>
<td>HetLipY_R</td>
<td>AAAGCTAGCAAGTGATTGTGGGATGGT</td>
</tr>
<tr>
<td>Mtp64_F</td>
<td>GAGCTCTGTCAGGCGATCGTC</td>
</tr>
<tr>
<td>Mtp64_R</td>
<td>AAAGCTAGCGTACTGCGGAGAGTCTT</td>
</tr>
<tr>
<td>HbHA_F</td>
<td>AAAGATATCGAAAAATAGTTAATGAAAGGA</td>
</tr>
<tr>
<td>HbHA_R</td>
<td>AAAGCTAGCTTCTCTGGTGACCTCTTGG</td>
</tr>
<tr>
<td>19kDa_F</td>
<td>AAAGATATCGCTACTGCGGAGAGTCTT</td>
</tr>
<tr>
<td>19kDa_R</td>
<td>AAAGCTAGCAGGCGTACGCTT</td>
</tr>
<tr>
<td>MART-1_F</td>
<td>AAAGATATCGCTACTGCGGAGAGTCTT</td>
</tr>
<tr>
<td>MART-1_{Goed}_R</td>
<td>CCCAGCTTTCTCaAGGTGAAATAGTTGCTT</td>
</tr>
<tr>
<td>MDTBlip_{Y_F}</td>
<td>ACTGCTAGCTAGGCGATCGTC</td>
</tr>
<tr>
<td>LA RP570 (HA_R)</td>
<td>ATGGATATCCCTAAAGCATAATCAGGAAATCATA</td>
</tr>
<tr>
<td>∆100-145 F</td>
<td>TGGCCAAACGCTCTTGCGCTGAGCAGGCCAGGATGCTT</td>
</tr>
<tr>
<td>∆100-145 R</td>
<td>GGTGAAGATCTTGGCGCTGAGCAGGCCAGGATGCTT</td>
</tr>
<tr>
<td>∆158-205 F</td>
<td>CCGGATTGCACTGCGTCTTGCGTCTGCGTGCAGGCAGG</td>
</tr>
<tr>
<td>∆158-205 R</td>
<td>ACCGTCTTGCCGACGCGAAGCGATGCGAATCGCC</td>
</tr>
<tr>
<td>∆158-180 F</td>
<td>CCGGATTGCACTGCGTCTTGCGTCTGCGTGCAGGCAGG</td>
</tr>
<tr>
<td>∆158-180 R</td>
<td>CTGGCAATGAGCAGGCTAAAGCGGAAGCGATGCGAATCGCC</td>
</tr>
<tr>
<td>∆181-205 F</td>
<td>ACTTTGCAATCCCAAACAATTCGCTGAGCAGGCCAGGATGCTT</td>
</tr>
<tr>
<td>∆181-205 R</td>
<td>ACCGTCTTGCCGACGCGAAGCGATGCGAATCGCC</td>
</tr>
<tr>
<td>LipY Fw</td>
<td>TTTGCTAGCTTGCGTCTTGCGTCTGCGTGCAGGCAGG</td>
</tr>
<tr>
<td>LipY Rv</td>
<td>AAAGATATCGCTACTGCGGAGAGTCTT</td>
</tr>
<tr>
<td>OVA Fw</td>
<td>TTTGCTAGCTTGCGTCTTGCGTCTGCGTGCAGGCAGG</td>
</tr>
<tr>
<td>OVA-HA Rv</td>
<td>AAAGATATCGCTACTGCGGAGAGTCTT</td>
</tr>
</tbody>
</table>

Table 1. Primers used in this study
the truncated constructs, independent PCR reactions were performed with a LipY forward primer (MDTB\textunderscore LipY\_F (Daleke et al. 2011)) and a reverse primer containing the 5’-truncation flanking sequence, as well as a 3’-tail with a complementary sequence of the 3’ truncation-flanking sequence (Δ100-145 R, Δ158-205 R, Δ158-180 R and Δ181-205 R respectively). Similarly, a reverse primer specific for sequence of the C-terminal HA-tag (LA RP570 (HA\_R)) was combined with a forward primer with the reverse complement sequence spanning the specific truncation (Δ100-145 F, Δ158-205 F, Δ158-180 F and Δ181-205 F respectively). PCR products of these individual reactions were mixed in a 1:1 molar ratio and diluted 1,000 fold. A second PCR with primers MDTBLipY\_F and LA RP570 (HA\_R) created the truncated LipY constructs. These products were cloned into the vector pSMT3 restricted with BamHI and EcoRV, which replaced the full-length construct. This created vectors pSMT3::lipYΔ100-145, pSMT3::lipYΔ158-205, pSMT3::lipYΔ158-180 and pSMT3::lipYΔ181-205, which were transformed to E. coli Dh5α and thereafter electroporated in M. marinum E11.

pSMT3::lipY\textunderscore OVA was constructed by amplifying the sequence corresponding to the first 205 amino acids of LipY with primers LipYss Fw and LipYss Rv from the template pSMT3::lipY\textunderscore tub\_HA. In parallel, OvaL was amplified with an added C-terminal HA-tag from plasmid pEM3-HbpDL-Glyk-OvaL (kind gift from Maria Daleke) using primers OVA Fw and OVA-HA Rv. Both fragments were cut with BamHI and ligated together. This ligated product was amplified with primers LipYss fw and OVA-HA Rv and ligated into the pSMT3 vector using NheI and EcoRV.

Secretion analysis and immunoblot
Mycobacterial liquid cultures in exponential growth were washed and inoculated in 7H9 medium without bovine serum albumin (BSA) at an optical density of 0.35 OD/ml and were incubated overnight. Cells were harvested by centrifugation at 5,000 rpm and an equivalent of 1.5 OD units (~3*10^8 bacterial cells) were suspended in solubilization/denaturation buffer. Subsequently, these samples were homogenized by mild sonication and incubated at 95°C. Another aliquot of 1.5 OD units was suspended in 100 μl of 0.5% Genapol-X-080 solution and incubated 30 minutes under head-over-head rotation at room-temperature. Extracted material and cells were harvested and were suspended in solubilization/denaturation buffer. Culture supernatant was filtered with a 0.2μm filter and was concentrated by TCA precipitation. All protein samples were separated with SDS-PAGE and were transferred to nitrocellulose membranes by western blot. Proteins were visualized by staining with polyclonal rabbit anti-G-CSF serum (AbCAM, Cambridge, UK) or monoclonal antibodies against MART-1 (anti-Human Melan A, cloneA103, Monosan, Uden, the Netherlands) influenza hemagglutinin (HA.11 Covance), GroEL2 (kind gift from J. Belisle), or mouse PE\_PGRS (7C4.1F7) (Abdallah et al. 2006)
Lipolytic Assay

Lipolytic activity of surface exposed LipY constructs was performed as described before (Mishra et al. 2008; Daleke et al. 2011). Shortly, bacteria were grown in dark conditions until cultures reached an optical density of 0.8-1.2. Cells were harvested by centrifugation and three aliquots of $2 \times 10^8$ bacteria were resuspended in 100mM Tris-HCl buffer (pH 8.0) containing 0.5mM p-Nitrophenyl stearate (Sigma). The suspension was vortexed and incubated at 37°C during 15 minutes. Supernatant of the reaction was collected and absorbance at 405nm was measured in 96-wells plates.

Error Prone PCR and double filter secretion screen.

A random mutagenesis library of LipY-OVA was created by performing a PCR reaction using the primers LipYss Fw and OVA-HA Rv and an error prone DNA-polymerase derived from Pyrococcus furiosus (Biles and Connolly 2004). After 30 cycles of amplification, the PCR product was isolated and cloned into a pSMT3 vector using the restriction enzymes Nhel and EcoRV. The library of ligation products were transformed to *E. coli* Dh5α and this library was plated on LB agar plates. Approximately 1,000 colonies were harvested from the plates and plasmids were isolated from this library. The mutated constructs were electroporated to *M. marinum* E11 and the cells were plated on nitrocellulose filters placed on 7H10 plates. After 11 days of growth, the colony-containing filters were placed on fresh plates that also contained clean nitrocellulose filters. After overnight incubation the second filter without the colonies were analyzed using a monoclonal antibody against influenza hemagglutinin (HA.11 Covance dilution 1:500). Stained filters were compared to the control filters and supersecretion colonies were selected for further analysis by western blot. Plasmids from investigated mutants were isolated by isolating mycobacterial DNA and transforming this to *E. coli* DH5α. Plasmid inserts were sequenced by Macrogen Europe to detect single-nucleotide polymorphisms.

ACKNOWLEDGEMENTS

We thank Maria Daleke (Abera Bioscience, Amsterdam, the Netherlands) for design of OvaL. We thank Vincent van Winden and Roy Ummels for technical assistance. We thank Wouter Jong, Erik Hooijberg, Tanja de Gruijl and Joen Luirink for valuable discussions.
7

SUMMARIZING DISCUSSION
In this thesis, we have used transposon-based genetic screens to gain further insight in the function and mechanism of ESX-5 secretion. ESX-5 is one of the type VII secretion (T7S) systems in mycobacteria and one of the major secretion pathways in slow-growing mycobacteria. The latter can be concluded from two findings: first, the number of ESX-5 substrates identified in other studies (Abdallah et al. 2009) and in our work (Chapter 3) far exceeds the number of substrates of all other T7S systems, and second, structural components of the ESX-5 system are detected in higher amounts than components of other T7S systems, at least in M. marinum (Chapter 3; van der Woude et al. 2012). Despite its apparent dominance, the importance and function of ESX-5 substrates have been less studied thus far. A possible explanation for this is that the major substrates of ESX-5, i.e. the PE and PPE proteins, are notoriously difficult to examine. In this study we set out to start to fill this gap in our knowledge. One of the major findings was that the ESX-5 system is essential for the in vitro growth of mycobacteria because of its role in outer membrane permeability. Furthermore, we have identified two important PPE substrates of ESX-5 with major functions: PPE38 and PPE10 facilitate the secretion of other PE/PPE substrates and shape the cell envelope respectively (especially the capsule layer).

**ESX-5 ESSENTIALITY AND MEMBRANE PERMEABILITY**

Previous research indicated that ESX-5 is essential for slow-growing mycobacteria such as M. marinum and M. tuberculosis (Abdallah et al. 2006; Di Luca et al. 2012; van der Woude et al. 2012). In contrast, fast-growing mycobacteria do not have an ESX-5 secretion system and therefore seem able to circumvent the essential functions of ESX-5 by other mechanisms. Even more remarkable was the finding that ESX-5 essentiality seems to be strain-dependent in M. tuberculosis (E. N. G. Houben et al. 2012; Di Luca et al. 2012). We describe in chapter 3 that expression of the major outer membrane porin MspA of the fast-growing species Mycobacterium smegmatis rescued the essentiality of ESX-5 in the slow-growing M. marinum. This poses an interesting dichotomy between fast-growing and slow-growing mycobacteria: all slow-growing mycobacteria possess an orthologue of the ESX-5 secretion system (Gey van Pittius et al. 2006; Gey Van Pittius et al. 2001), while fast-growing mycobacteria possess general hydrophilic MspA-like outer membrane pores (Wolschendorf, Mahfoud, and Niederweis 2007; M Niederweis et al. 1999; Stephan et al. 2004). The individual branches of this dichotomy have been described before, but our findings suggest that there may be a causal link between these findings. Increase in membrane permeability is known to attenuate mycobacteria in macrophage infections (Kirksey et al. 2011; Sharbati-Tehrani et al. 2005; Yu et al. 2012). We can therefore speculate that the loss of general hydrophilic pores, such as MspA, in slow-growing mycobacteria was an evolutionary step, which was necessary for pathogenic mycobacteria to adapt to an intracellular life-style. Possibly, the ESX-5 system and (a subset of) its substrates evolved in these slow-growing mycobacteria to create more specific outer membrane channels that are more selective in their substrates.
and that can be more tightly regulated. This would allow the mycobacteria to better withstand the harsh environments encountered during infection, whereby they remain able to obtain nutrients when conditions are more favorable. This trade-off between a permeable outer membrane and therefore efficient nutrient uptake on the one hand and impermeability and resistance to host attack on the other, has been elegantly postulated in a previous study on outer membrane lipids of *M. tuberculosis* (Yang et al. 2014). Trehalose di-mycolates (TDMs) are an important part of the outer membrane of *M. tuberculosis*, and contribute to the impermeability of the cell-envelope. When these TDMs are hydrolyzed by mycobacterial cutinases, nutrient uptake is increased, but so is susceptibility to the attack of activated macrophages (Yang et al. 2014). This effect of membrane lipids on membrane permeability, appears comparable to the role of outer membrane channels, which are required for nutrient uptake, and depend on ESX-5 for their insertion in the membrane. This raises an important question: which ESX-5 substrates are responsible for membrane permeability and/or nutrient uptake? A question that we unfortunately have not been able to answer, probably due to redundancy as there are dozens of ESX-5 substrates. However, that T7S substrates can form porins in the mycobacterial outer membrane is illustrated by the recently described CpnT protein (Danilchanka et al. 2014; Danilchanka et al. 2015). The actual T7S system that is involved in CpnT secretion is thus far unknown, but this porin has all characteristics of a T7S substrate (Daleke et al. 2012). Like ESX-5, CpnT and its orthologues are only present in slow-growing mycobacteria. We were not able to detect CpnT in our membrane fractions or in the cell-surface localized proteome (Chapter 3). Possibly, this was due to low amounts of CpnT being produced in *vitro*, but it could also be that outer membrane proteins are lost during the enrichment that we used in our experiments. There are more indications that T7S could fulfil a role in nutrient uptake. Recent structural analysis of the ESX-1 associated protein EspB has shown that this protein multimerizes into a barrel-shaped heptamer with a central pore (Solomonson et al. 2015; Korotkova et al. 2015). Furthermore, on one side of the barrel there is a hydrophobic region with potential lipid binding domains. The monomeric structure of EspB is highly similar to that of PE/PPE dimers, suggesting that PE/PPE protein pairs may form a similar quaternary. This barrel-like structure appears large enough to transport nutrients through its pore. Furthermore, certain PPE proteins, such as the ESX-3 associated PPE4 and conserved proteins PPE1 and PPE2, have highly hydrophobic domains in their C-terminal extensions, which are classified as putative transmembrane domains, and could possibly assist in channel formation. Whether PPE4 is involved in the uptake of iron and zinc, the main known function of ESX-3 (Siegrist et al. 2009; Serafini et al. 2009), is an interesting hypothesis. If correct, this could help to improve the understanding of PPE proteins as nutrient channels in a more defined setting than ESX-5, which has dozens of substrates.

More hints on the specificity of certain ESX-5 substrates for different nutrients could come from the study of *Mycobacterium bovis* BCG strains. After the initial
development of BCG in the 1920’s in France, BCG has been distributed and maintained in different laboratories worldwide. Culturing of these strains on different nutrient media and over long periods of time has resulted in many genomic mutations and differences in expression profiles between different strains (Brosch et al. 2007). Recently published data (Abdallah et al. 2015) shows that BCG Tice, a strain of BCG that is currently mainly used for treatment of bladder cancer, contains a duplication of the complete esx-5 locus that leads to increased expression of the ESX-5 conserved components and of a few substrates (Abdallah et al. 2015). Likewise, BCG Connaught was shown to over-express the PE-PPE couple PE31-PPE18 (Abdallah et al. 2015). Future work by our group will focus on exploring whether these mutations and changes in expression lead to improved growth on specific nutrients. The presence of such nutrients in growth media during the many years of subculture in the laboratory might have led to these genomic changes.

The link between membrane permeability and the ESX-5 system might also be explained direct effects of ESX-5 substrates on lipid profiles of mycobacteria. This could possibly occur through interactions with cutinases or other lipolytic proteins. Since the many different ESX-5 substrates perform diverse biological roles, both these hypotheses might be partially correct and ESX-5 might influence membrane permeability in multiple ways. Unfortunately, we have not been able to study the lipid profiles of ESX-5 mutants in detail during my PhD research, but initial analyses did show differences in lipid profiles that, as of yet, remain uncharacterized (Ates & van de Weerd, unpublished results). No differences were detected in TDM or PDIM lipids however, which are mycobacterial lipids that do have a role in membrane permeability (Ates et al., 2015; Ates & van de Weerd unpublished results). Further research in this area is certainly warranted, but was unfortunately not possible in the timeframe of my research.

That certain ESX-5 substrates are important in nutrient uptake does not mean that this is the only biological role of all ESX-5 substrates. Chapter 4 and chapter 5 of this thesis describe that ESX-5 substrates have different functions in the biology of mycobacteria. However, methodological difficulties associated with analysis of PE and PPE proteins and genes, have thus far hampered research on these ESX-5 substrates. Many PE/PPE encoding genes are not only highly homologous, but also have a high GC-content (above 65%) and contain repetitive domains. Therefore, these genes are often excluded from genome and RNA sequencing analyses (E.g. McMurry et al., 2005). Furthermore, PE/PPE proteins are difficult to detect by methods such as mass spectrometry, due to a paucity of trypsin cleavage sites and high homology (Schubert et al. 2013). The specific and important biological functions for two PPE proteins described in this thesis show that a more individual investigation of PE/PPE proteins is necessary if we want to understand the biological role of these proteins. This is a logical conclusion when looking at substrates of other secretion systems from different organisms. For instance, Sec-secretion substrates vary greatly in function and can only be grouped by their mechanism of secretion. This is also
supported by the study by Daleke et al. which shows that the ESX-5 substrate LipY can be transported by different N-terminal secretion motifs in different species of mycobacteria (Daleke et al. 2011). In other words, the lipase LipY is secreted by a PE-domain, a PPE-domain or by a classical signal sequence in different species. These different routes of secretion do not alter the function of this lipase and therefore inversely, it should not be assumed that PE and PPE proteins share biological functions because they share a similar route of secretion. Perhaps the conserved PE/PPE domain is, at least for the longer substrates, only required for secretion and not for the ultimate function of these proteins.

The improved understanding of the secretion of LipY (Daleke et al. 2011; Daleke, Ummels, et al. 2012) motivated us to explore this ESX-5 substrate as a carrier for heterologous secretion in mycobacteria, as described in chapter 6. Achieving heterologous secretion in mycobacteria is an arduous task that thus far has not been analyzed systematically. Fusions of heterologous substrates to predicted signal sequences or full-length secreted mycobacterial proteins often fail and fall into obscurity (Rosenberger, Brülle, & Sander, 2013; Ates et al., unpublished results; Chapter 6). However, sometimes they do (seem to) lead to secretion (Bastos et al. 2009; De Maio et al. 2014). However, reports on successful heterologous secretion are generally anecdotal and miss data on cellular localization or are poorly controlled. In the study described in chapter 6 our initial attempts to induce heterologous secretion failed. Therefore we attempted to approach heterologous secretion by mycobacteria in a more structured manner. We selected the N-terminus of the ESX-5 substrate LipY_{\text{tub}} for this approach, since it secretes the lipase domain of LipY_{\text{marinum}} in M. marinum (Daleke et al. 2011). Targeted truncations in the linker domain showed that, additional to the PE-domain, the complete linker domain is required for optimal secretion of LipY. Therefore the PE-domain and the linker domain were fused to an artificial protein sequence containing the main epitopes of Gallus gallus ovalbumin (Ova). This chimerical protein was secreted in an ESX-5 dependent manner, albeit inefficiently. Mutagenesis of the chimerical protein combined with a screening approach identified mutants with greatly improved secretion. As the mutations with a positive effect were shown to occur in the C-terminus of the chimeric protein, it seems unlikely that the N-terminal domains of LipY of these mutants would improve secretion when fused to other heterologous sequences. The technique that we describe can be used to improve the secretion of any heterologous protein of interest. Many challenges still remain before reliable platforms for heterologous secretion can be used on a large scale to create new vaccine candidates or provide answers to fundamental research questions. Recently improved knowledge on the fundamental workings of T7S warrants new efforts to endeavor and improve heterologous secretion by mycobacteria, but this needs to be undertaken in the context of fundamental research on protein secretion in mycobacteria before moving to in vivo approaches.
PPE10

In Chapter 4 we have shown that the ESX-5 substrate PPE10 is involved in capsular integrity of pathogenic mycobacteria. Mutants in esx-5 or ppe10 shed capsular components more easily in the presence of detergent, which resulted in reduced virulence in the early stages of infection due to reduced phagosomal rupture by capsular proteins. This is another major phenotype that was observed in ESX-5 mutants of both *M. marinum* and *M. tuberculosis* and could be completely attributed to PPE10 in *M. marinum*. Unfortunately also here a molecular mechanism for this phenomenon is not yet identified. It is however tempting to speculate on the possible function of PPE10, although we still lack experimental data to support these possible hypotheses. One of the possibilities that would fit our data is that PPE10 plays a structural role in the mycobacterial capsule. Although previous data make it clear that PPE10 is not the most abundant protein in the capsule (Sani et al. 2010), perhaps PPE10 interacts with other capsular components and keeps them in place. This could explain why most capsular components seem to be present but less efficiently retained in the PPE10 mutant. An alternative hypothesis is that PPE10 has a role in the biogenesis or export of an unidentified capsular or membrane component (which could be a protein, lipid or other molecule), which performs the actual capsule stabilizing function. Immunoprecipitation experiments with tagged PPE10 or an anti-PPE10 antibody may help to identify the interacting partners of this protein, thereby helping to elucidate the molecular function of PPE10. BLAST searches of PPE10 did not reveal any possible molecular functions for the C-terminus of this protein. The middle part of the protein contains penta-peptide repeat domains that have not been associated with any known protein function, but which might be important for the structure of this part of the protein. The protein MYCSM_05776 from *M. smegmatis* JS623 contains a domain with low homology to the C-terminus of PPE10 (~40% identity), which is putatively secreted by a Sec secretion signal sequence. This could be a good start to investigate which *M. smegmatis* proteins are involved in capsule integrity in fast-growing mycobacteria.

The significance of the mycobacterial capsule in human tuberculosis is still uncertain. It is even still under debate whether the mycobacterial capsule is a structure that is actively created by the mycobacterium or whether it is composed of cellular remnants that stick to the cells. Our data indicate that PPE10 is secreted to stabilize the capsule. Combined with our findings that integrity of the capsule is important in virulence of *M. marinum*, this suggests that the mycobacterial capsule should be considered as a virulence factor and not as an artifact of mycobacterial growth. Capsular components have been implicated in immune-modulation and this could be a significant biological role for the mycobacterial capsule (Geijtenbeek et al. 2003; Driessen et al. 2009; Geurtsen et al. 2009). Whether the mycobacterial capsule also plays a role in biological processes such as antibiotic resistance or resistance to host-attack remains to be established. Our lab is currently developing PPE10 mutants in *M. tuberculosis* and *M. bovis* BCG to help answer these questions. It is unclear whether
the capsule will play similar roles in different species of mycobacteria. For instance, surface-localized ESX-1 substrates are important for virulence of *M. marinum*, but this has not been shown for *M. tuberculosis*. This could be due to differences in biogenesis of the capsule, since we see that most capsular components are only detected on the surface of *M. tuberculosis* after 5 days of growth without detergent (Sani et al. 2010; Chapter 4; van der Wel & Bitter unpublished results). Growing *M. tuberculosis* without detergent leads to bacterial clumping that prevents *in vivo* or cell-infection experiments to be performed in a reproducible way. The finding that the capsule of *M. tuberculosis* seems to regenerate more slowly than that of *M. marinum* could explain why *M. tuberculosis* seems to need more time to rupture the phagosome (Abdallah et al. 2011; van der Wel et al. 2007), or why many substrates that are surface localized in *M. marinum* are secreted and detected in the culture filtrate of *M. tuberculosis* (Houben et al., 2012; Chapter 3; Chapter 5).

An unexpected finding described in chapter 4 was that the *ppe10* mutant strain of *M. marinum* showed a complete single-cell phenotype in liquid culture, even when grown without detergent. Current efforts in our lab are aimed at exploiting this phenotype to make *in vivo* and cell-infection experiments more reliable by eliminating the variation due to mycobacterial clumping. High-throughput analysis of zebrafish embryos by automated injection of early stage (2-8 cells) embryos is a promising technique that is being developed for *in vivo* drug screens against mycobacterial infection. Mycobacterial clumping introduces variation in these experiments, which is highly problematic for large screens. Initial results in our lab show that these variations can be circumvented by using the *ppe10::tn* mutant (Deventer, van der Sar, Bitter et al. unpublished results). Whether this is a useful approach for drug screens must still be confirmed, but these preliminary results illustrate the methodological problems still attached to mycobacterial clumping and how the results in chapter 4 might help to find new solutions for these problems.

**PPE38**

In chapter 5 we show that loss or disruption of a single genetic locus, *ppe38-71*, leads to a complete loss of PE_PGRS secretion. The very specific function of this locus suggests that PPE38 could be the designated PPE partner for the PE_PGRS proteins, which have no known secretion partner. There is only limited information on the structure of PE and PPE proteins (Strong et al. 2006; Ekiert and Cox 2014; Korotkova et al. 2014), but based on the available data it has been hypothesized that PE and PPE proteins form heterodimers in the bacterial cytosol. Secretion of these heterodimers is dependent on a chaperone-like protein known as EspG (Daleke, Woude, et al. 2012; Korotkova et al. 2014). PPE38 has been shown to interact with EspG₅ (Daleke, Woude, et al. 2012), the designated chaperone of the ESX-5 system. Structural analyses of the ESX-5 substrates PE25 and PPE41 in complex with EspG₅ have shown that the EspG proteins interact with PPE proteins, but not directly with PE proteins (Korotkova et al. 2014; Ekiert and Cox 2014).
Furthermore, interaction of PPE’s with EspG could be required for system specificity (Phan, Houben, Bitter unpublished results). Therefore, PE_PGRS proteins probably do need a PPE partner protein. Current efforts in our laboratory are now aimed at elucidating whether there indeed is an interaction between PPE38 and PE_PGRS proteins, or whether the observed effects must be explained by another mechanism. The finding that PPE38 is essential for the secretion of PE_PGRS and PPE_MPTR proteins, but not for the secretion of other ESX-5 substrates, suggests that the role of PPE38 is specific for certain sub-groups of ESX-5 substrates. Indeed, a recently published paper (Shah et al. 2015) shows that another genomic locus containing both a PPE and Esx proteins(pe8-ppe15-esx1-esxJ) is essential for the secretion of another sub-group of ESX-5 substrates. The authors propose that ppe-loci containing esx-genes might all have similar functions for subgroups of ESX-5 substrates. This hypothesis is supported by our results, since the ppe38 locus also contains two esx genes. LC-MS/MS analysis of the ppe38::tn mutant in M. marinum showed that this strain was impaired in the secretion of all detected ESX-5 substrates (Ates, Ummels, Piersma, unpublished results), while in M. tuberculosis only two specific subsets of substrates seemed affected. In M. marinum ESX-5 substrates that are not secreted are not detected intracellularly anymore. This is not seen in M. tuberculosis (e.g. Abdallah et al., 2009; Chapter 3; Chapter 5), where intracellular protein expression remains mostly unaffected when secretion is impaired. Loss of intracellular expression of non-secreted substrates is not due to reduced transcription of genes or regulation of mRNA since the mRNA levels are even more abundant in the ESX-5 mutant of M. marinum compared to the wild-type strain (Abdallah, Ummels, Bitter unpublished results). In other words: M. marinum seems to have a strong interdependency of secretion of ESX-5 substrates similar to what was previously reported for ESX-1 (Fortune et al. 2005), indicating that while M. marinum is a great model to detect secretion deficiencies, results should be verified in M. tuberculosis before it can be stated whether these are general or specific secretion defects.

Although many questions remain about the molecular mechanism behind the effects of ppe38 mutations on ESX-5 secretion, the fact that these mutations occur in hypertransmissible clinical isolates is a major finding that will have significant impact on the understanding of ESX-5 secretion and its role in virulence of M. tuberculosis. The hypervirulence of M. tuberculosis strains of the Beijing lineage has been a topic of intense research and speculation over the past decade (Abebe and Bjune 2006; Luo et al. 2015; Hanekom et al. 2011). Several mutations have been detected in strains of the Beijing lineage, all of which have been implicated in this hypervirulence (Reed et al. 2004; Reed et al. 2007; Alonso et al. 2011). These mutations were mostly involved in regulation or production of virulence-associated lipids. However, the observed associations with these mutations and hypervirulence are usually not exclusive (Aguilar et al. 2010; Ribeiro et al. 2014; Midori Kato-Maeda et al. 2012; Sinsimer et al. 2008). The hypervirulence caused by mutation of the ppe38-locus may help to further understand the population dynamics of specific strains of M. tuberculosis.
Our results add to previous evidence that loss of PE_PGRS secretion leads to hypervirulence. This was first shown by our group in adult zebrafish, where the ESX-5 deficient strain of \textit{M. marinum} (7C1/espG5::tn), proved to be hypervirulent (Weerdenburg et al. 2012). Future research should focus on elucidating the mechanism behind this hypervirulence, since as yet, we do not know whether this is a direct result of the loss of PE_PGRS or PPE_MPTR secretion. It has been postulated that PE_PGRS proteins interact with and regulate the immune system (Banu et al. 2002; Abdallah et al. 2008). Alternatively, the lack of secreted ESX-5 substrates may lead to alterations in the cell wall of the mycobacteria, which somehow leads to hypervirulence. It also remains unknown how widespread \textit{mt2419-22} mutations are among the Beijing lineage. This could be a phenomenon linked to this South African clinical setting, but might also be present in all hypervirulent Beijing strains. Analysis of published \textit{M. tuberculosis} genomes for these mutations is difficult, because these are often created by assembling short reads to the H37Rv sequence (Cole et al. 1998; Kapopoulou, Lew, and Cole 2011). H37Rv is unfortunately erroneously annotated as having a RvD7 mutation, that is a deletion of \textit{mt2420-22} (McEvoy, van Helden, et al. 2009). Interestingly, a recently published study (R. S. Lee et al. 2015) describes the successful spread of a single strain of \textit{M. tuberculosis} belonging to the Euro-American lineage. All isolates in this cohort show a mutation in the \textit{ppe38}-locus. Therefore, these ppe38 mutations are not exclusive for the Beijing lineage of \textit{M. tuberculosis}, which is further verified by strain T92 that belongs to the East-African-Indian lineage and also has a mutation deleting the \textit{ppe38}-locus (McEvoy, van Helden, et al. 2009). This shows that ppe38-mutations are more widespread and occur in different \textit{M. tuberculosis} lineages. This suggests convergent evolution of these mutations and that they might be (partially) responsible for the successful spread of these isolates. Current and future efforts by our group in collaboration with Stellenbosch University will focus on assessing how widespread the ppe38-locus mutations are and how this is correlated to variations in PE_PGRS secretion and virulence. The findings described in chapter 5 show that there are large differences in PE_PGRS secretion between different clinical strains of \textit{M. tuberculosis}. This is also reflected in laboratory strains of \textit{M. tuberculosis}: CDC1551 secretes high amounts of PE_PGRS proteins, while the most widely used strain, H37Rv, secretes only low amounts of PE_PGRS proteins (Houben et al., 2012; Chapter 5). Apart from differences in secretion levels of all PE_PGRS proteins, preliminary analyses also show different patterns of expression and secretion of particular PE_PGRS proteins in different clinical isolates (Chapter 5; Bastiaansen & Gey van Pittius unpublished results; Houben & Young unpublished results). Further analysis of these differences in larger numbers of clinical isolates is needed to prove and understand the association between PE_PGRS expression or secretion profiles and virulence. An important current topic of study is to find ways to predict whether patients infected with \textit{M. tuberculosis} will develop active tuberculosis or not by assessing patient or bacterial biomarkers. The finding that strains deficient in PE_PGRS secretion cause more aggressive and fulminant
disease might significantly aid in predicting disease progression based on bacterial factors and could therefore help in the development of predictive diagnostics of disease progression.

Our results possibly provide the tools to separate the effects of PPE_MPTR and PE_PGRS proteins from the effect of other ESX-5 substrates that are not affected by PPE38. This is already leading to increased understanding of the different subgroups of ESX-5 substrates and their individual roles in the biology of mycobacteria. E.g., although the ESX-5 conserved membrane components are essential for in vitro growth of pathogenic mycobacteria, this is not the case for PPE38, indicating that the PE_PGRS, or PPE_MPTR proteins are not likely candidates to be the essential substrates of ESX-5.

Taken together, the results presented in chapter 5 highlight the importance of the PPE and PE_PGRS substrates in the virulence of M. tuberculosis and should be an incentive to intensify research on these proteins, which have so far been relatively neglected in mycobacterial research.

CONCLUDING REMARKS

The results described in this thesis focus on important new biological roles for substrates of the ESX-5 system, such as membrane permeability, nutrient uptake, capsular integrity and modulation of virulence. That we have pinpointed individual ESX-5 substrates to be responsible for some of these phenotypes (Chapters 4 & 5) is a major breakthrough in our understanding of the function of T7S. These results already seem to have direct implications for our understanding of mycobacterial virulence and could therefore help with progress in this field. Our results raise a whole array of new questions that we cannot answer at this time. We can only guess what and if these results will ever mean anything for the 9 million people that develop TB each year. However, I am convinced that new solutions to combat tuberculosis, be it through improved vaccines, predictive diagnostics or new antibiotics, will require a deeper understanding of the biological role of T7S and its substrates than we currently have. I hope that the results described in this thesis will contribute to that understanding.
DANKWOORD/
ACKNOWLEDGEMENTS
Het leven van een promovendus gaat niet over rozen, maar eigenlijk is het een
droombaan, waarin je word gestimuleerd om dingen te proberen, dingen te leren,
kennis over te brengen en tegelijkertijd aan de rand van de menselijke kennis nieuw
territorium te verkennen. Dat neemt niet weg dat het meeste van wat je doet gewoon
mislukt. Ik begon aan mijn promotie als een zelfverzekerde (misschien zelfs wel lichtelijk
arrogante) ex-student die altijd zonder moeite door zijn studie was gefietst en daarom
dacht dat hij het wel even zou gaan maken… Ik kan me nog goed herinneren hoe
ik een jaar later huilend op de fiets zat en niet wist of het ooit nog wat zou worden.

Als er één ding is wat me duidelijk is geworden, is het dat er niets belangrijker
is in het onderzoek dan goede mensen om je heen te hebben om je te helpen en
leuke mensen om je heen te hebben om je te ondersteunen en vol te laten houden.
Het geweldige aan mijn promotietijd was dat de goede mensen om me heen ook
nog leuke mensen bleken te zijn en dat het daarom een prachtige tijd is geworden. Ik
besef me maar al te goed dat zo’n goede sfeer op een afdeling niet overal te vinden
is en ben er dan ook superdankbaar voor.

En dan moet je iedereen bedanken, waar begin je? Misschien moet ik gewoon nog
een keer schrijven dat jullie een superfijne groep collega’s waren en dat de volgorde
hier dan ook niet zoveel betekend.

Maar als ik dan toch ergens moet beginnen dan begin ik met Wilbert te bedanken.
Bedankt dat je mij de kans hebt gegeven om aan de slag te gaan op de afdeling.
Je bent een geweldige promotor. Ik heb er ongelooflijk veel respect voor hoe je
overall in alle projecten steeds de rode draad weet te vinden, zelfs als de data
er oorspronkelijk als een blauw bolletje wol uitzien. Zonder jou briljante ideeën
(zoals MspA en het combineren van PPE10 en de hemolyse) was mijn onderzoek
niet half zo mooi geworden en je wist natuurlijk al veel eerder welke kant ik beter
op kon gaan, ook al zag ik het zelf nog niet. Ik ben blij dat ik aan het eind van mijn
promotie ook zelf een paar ideeën kon toevoegen, maar als ik in mijn carrière jou
niveau als wetenschapper, begeleider en (tegen wil en dank) manager maar enigszins
can benaderen kan ik tevreden zijn.

Maar ik had natuurlijk een overvloed aan begeleiders tijdens mijn promotie.
Christina, ook aan jou bedankt voor de kans die je me hebt gegeven door me aan
nemen. Gedurende mijn promotie was je altijd op de achtergrond aanwezig als
beschermende baas. Ik wist dat als er iets mis zou gaan op de afdeling dat ik dat met
jou kon bespreken, gelukkig is dat nooit nodig geweest. Maar je hebt me wel scherp
gehouden met je vragen vanuit een net ander oogpunt tijdens werkbesprekingen
en AiO gesprekken. En Ben, nu ben je met pensioen, maar gelukkig kwam je nog
steeds elke week langs. Op die momenten kunnen we nog steeds heerlijk botsen als
jij ideeën hebt voor extra controles die het experiment nog mooier maken en ik mijn
projecten alleen maar zo snel mogelijk af wil hebben. Maar altijd als er iets was op
persoonlijk vlak, kon ik het met jóu bespreken en we hebben altijd een klik gehad als
het ging om hoe we in de wetenschap en in het leven staan denk ik.
Ook veel dank aan Jeroen die belangrijk was in de beginfase van mijn project - wie had gedacht dat ik nog een paper over het kapsel zou gaan schrijven? En Coen, die een vergelijkbare rol vervulde in de laatste fase van mijn project – altijd als ik een vraag aan één van jullie had, hadden jullie een antwoord. Geen gelul, gewoon een duidelijk antwoord. Coen, je hebt in een enorm korte tijd een paar ideeën gehad die onze publicatie heel mooier hebben gemaakt en ik vind het jammer dat onze samenwerking wat dat betreft zo kort was. Ik wens je veel succes met het uitbouwen van je eigen groep op de MMI! Ook Astrid en Joen bedankt voor de goede invalshoeken tijdens werkbesprekingen die zo geholpen hebben om de juiste weg in mijn projecten te vinden. Edith, na mijn moeizame omzwervingen in de eerste twee jaar van mijn promotie, mocht ik meehelpen aan het project van jou en Roy. Roy had al zoveel werk gedaan aan de essentialiteit van ESX-5 en ik kwam daar een beetje halverwege bij aanwaien. Hoewel het aan het begin misschien een makkelijk opstapje leek te zijn, bleek het een waar epos te worden om de reviewers te overtuigen van ons verhaal. Hoewel ik een snelle publicatie natuurlijk fijner had gevonden, heb ik nu wel meer het gevoel gekregen dat ik mijn plaats op het paper, naast jou en Roy daarmee ook verdient heb. Het was heel fijn om met je samen te werken en vooral bij het schrijven en herschrijven (ook van ons review) was het heel fijn om altijd zo snel en goede feedback te krijgen, daar heb ik heel veel van geleerd. En dan moet ik natuurlijk ook Roy bedanken, wat een held ben je toch. Ik vond het superfijn om met jou op het lab te zitten en ik ben je veel verschuldigd. Zeker aan het einde van mijn projecten (die nog steeds lopen) ben jij van onschattbare waarde geweest door je werk op het ML-III lab en je miraculeuze cloneringskunsten. Het blijft toch leuk dat Felix geboren is rond dezelfde tijd dat ik ben begonnen aan mijn promotie en dat hij nu een lopende en pratende dreumes is, laten we dat maar als metafoor voor mijn promotieproject zien.

Ook de andere analisten hebben mij natuurlijk heel veel geholpen. Marion, Janneke, Gunny, Theo en daar reken ik dan ook onze Wim nog bij, zonder jullie is de MMI natuurlijk niet de MMI. De analisten vormen de constante factor van een afdeling en dat doen jullie als de besten. Er heerst natuurlijk veel onduidelijkheid en stress over de aanstaande verhuizing naar het nieuwe gebouw. Hoewel ik blij ben dat ik me daar zelf niet mee bezig hoeft te houden, ben ik overtuigd dat het binnen een paar weken allemaal weer als van ouds is. De sfeer van de MMI hangt van jullie af en niet van het gebouw waar jullie inzitten.

Dan zijn er natuurlijk veel AiO en Post-Doc collega’s waar ik veel aan heb gehad. In mijn beginstadium, was ik qua onderwerp een beetje een vreemde eend in de bijt. Aan het begin waren daar de myco-angels (wat een verschrikkelijke naam, en waar zal Wilbert mee komen voor de jongens…), Maria, Esther, Aniek en Evelien. Jullie waren al een hele hechte club en waren duidelijk verder in jullie projecten dan ik, maar jullie maakten echter altijd tijd om mij te helpen bij mijn experimenten en het opzetten daarvan. Ik vind het vooral ook leuk dat ik nu jaren nadat Aniek onze afdeling heeft verlaten, twee van haar projecten weer extra leven heb in kunnen blazen en dat hier
waarschijnlijk prachtige publicaties uitkomen. En ook Abdallah en Jovanka waren aan het begin belangrijk in het begrijpen van het werk hier. Eigenlijk jammer dat we nooit samen echt hebben kunnen sparren over de projecten waar ik nu aan werk. Wie had gedacht dat ons werk zo dicht bij elkaar zou komen?

Maar ook de collega’s die het stokje weer over nemen waren van onschatbare waarde. Robert, valt daar natuurlijk niet helemaal onder. Wij hebben het grootste gedeelte van onze promoties samen op de kamer doorgebracht en nu zijn we ook allebei toe aan de volgende stap, ik hoop dat het tij zich snel voor je keert en je snel aan een nieuwe uitdaging kan beginnen. Het was zo fijn toen Maikel, Lisanne, Kin Ki, Vincent en Susanna op de MMI kwamen. Het begon toen echt stil te worden en opeens was er weer leven in de brouwerij! Borrels tot in de diepe uurtjes en weer mensen om het enthousiasme en de frustratie mee te delen. Vincent en Susanna, bedankt dat jullie mijn paranimfen willen zijn!

En ook de meer recent begonnen mensen: Maroeska, jij was al een bekende en ik ben heel blij dat je bij ons op de afdeling bent begonnen. Misschien kan je toch nog wat maken van een aantal van de losse eindjes die ik op mijn beurt weer achterlaat. Und auch Alex, Ich wünsche dir viel Spaß und Erfolg auf diese Abteilung.

And then all the colleagues of the MolMic: What insanity that these two departments are not one single department. Thanks so much for all the fun times we had! Karlijn and Sadeeq I could always share some of the PhD pain with one of you guys, but you both have great new jobs now, so it all worked out for all of us. And Trang, Catalin, Melanie, Marjolein, Dat, Maurice and Markus, I wish you all the best in your projects; I’m certain you guys are going to discover some great things in the near future – good luck! Peter, Mattijs, Bart, Corinne, Diane, Greg: passing by molmic was always fun or useful, usually both, thanks to your input on my questions or on the borrels.

Ook bedankt aan de drie stagiaires die ik heb begeleid en die mij hebben geholpen. Parveen, Marit en Kim, ik hoop dat jullie er iets van hebben opgestoken om bij ons op de afdeling rond te lopen. Ik vond het leuk om te zien hoe jullie beter werden in het wetenschappelijke gedeelte, maar alsook bij Marit en Kim in jullie zelfverzekerdeheid en presentatievaardigheden. Ik hoop dat jullie nog een mooie carrière voor je hebben en wens jullie daar veel succes mee.

I would also like to thank all the international people that helped me with my publications and the path leading up to them. Rob Warren, Anzaan Dippenaar and Nico Gey van Pittius from Stellenbosch University; it was a pleasure working with you. I’m so glad that the projects that I started with you will probably be continued by new PhD students from South Africa. And Rogelio, every time we wanted to carefully ask you if you could measure an extra parameter you were already two steps ahead of us! I hope the work we did will continue as a long collaboration between these universities. Arnab Pain, and the colleagues at KAUST, and Rainer Kalscheuer and Marina Alber in Düsseldorf: thank you for the invaluable help on my first publication! Maar er waren ook veel samenwerkingen binnen Nederland. Nicole en Daisy, dankjewel voor de prachtige EM-plaatjes die onze ideeën zo mooi illustreren. Sander en Connie, zonder
jullie massa-spectrometrie zouden mijn publicaties nooit het volledige beeld kunnen geven. En ook René Musters en Jeroen Kole, onze samenwerking was maar kort, maar leverde prachtige plaatjes op. And Juan Garcia Vallejo, thanks for letting Coen and me use your amazingly cool toy, I could think of so many great experiments that we could perform together. En last, but not least, Tanja, Jurjen en Anita van het CCA, heel erg bedankt voor de begeleiding bij mijn pogingen om mijn MART-1 BCG te testen. Er bleken gewoon te veel haken en ogen aan te zitten, maar aan jullie hulp heeft het niet gelegen en ik heb er veel van geleerd.

I would also like to thank Roland Brosch, for giving me the opportunity and trust to continue my career in the mycobacterial field. I’m grateful and excited for the time to come at the Institut Pasteur.

En dan zijn er natuurlijk zo veel mensen buiten de wetenschap waar ik heel veel aan heb gehad tijdens mijn promotieperiode. Een van de belangrijkste uitlaatkleppen voor de frustratie die er soms bij het onderzoek komt kijken was uiteraard het frisbeeën. Ik vond het superfijn om te kunnen trainen en spelen met alle teams waar ik de afgelopen 5 jaar bij betrokken was. De ervaringen met het 2012 Nederlands mixed team voor het WK in Japan waren geweldig, net als het coachen van het U17 jongens team. Maar natuurlijk het meeste van allemaal, al het afzien en genieten met de gasten van de Cakes, onbetwist (door ons) het leukste team van Nederland. Het was een beetje belachelijk om naast mijn promotie ook nog voorzitter te worden van onze club Crunch, maar dankzij mijn bestuursgenootjes was het toch niet zo’n slecht idee en hebben we een aantal mooie dingen kunnen waarmaken! Nu is ook dat stokje weer doorgegeven, veel succes daarmee.

Maar het allerbelangrijkste is natuurlijk het thuisme. Goede vrienden waar je dingen mee kan delen, het waren er gelukkig zo veel dat ik eigenlijk niet wil beginnen ze allemaal te noemen. Maar ik kom er in ieder geval niet mee weg om niet ons vaste groepje te bedanken voor de avonden met praten, drinken en spellen waar de stoom afblazen kon worden - Arne, Jeroen en Robin. Maar ook al die andere vrienden waarvan ik altijd het gevoel had ze te weinig te zien, maar waarvan je weet dat het meteen weer super is als je elkaar wel ziet. Maaieke, Tristan, Eva en al die anderen.

En hoe verder je komt in het dankwoord hoe dichter bij mezelf ik terecht kom. Tjeerd, ik vind het superfijn om jou als kleine broer te hebben. En hoewel onze band altijd goed is geweest, zijn we de laatste jaren ook echt gegroeid tot gelijkwaardige vrienden. En ik vind weinig leuken dan elke verdediging van een tegenstander samen met jou aan flarden te spelen. Ik vind het superleuk om te zien dat je langzamerhand ook je drive en doelen in het leven begint te vinden ook al moet nog even uitkristallisieren waar dat op uit gaat komen. De tijd dat ik trots ben als je iets bereikt gaat voorbij en verandert in een vertrouwen dat je goede dingen doet en gaat doen.

Mieke, we zeggen nooit mama tegen je, maar wat is het toch fijn om een mama te hebben. Langskomen voelt altijd als thuiskomen. Dat is natuurlijk ook logisch, want dat is het ook, maar wat ik bedoel is dat het rust geeft en dat is heel fijn. Ook de
hulp met alle lastige dingen die er nog liepen en lopen in de afgelopen tijden was natuurlijk onontbeerlijk. Alle problemen rond het huis, de financiën en dergelijke. Af en toe kon ik heel even als een klein kindje mijn kop in het zand steken en aan jou vragen om me te helpen.

En als laatste natuurlijk Julia. Ik hou van je. Dat is alles.
LITERATURE


the Role of IS6110 in a Highly Transmissible Mycobacterium Tuberculosis Beijing Strain, GC1237.”


Host Cell Mitochondria by the Mycobacterium Tuberculosis PE_PGRS33 Protein."


Mycobacterium Tuberculosis in the Inuit." *Proceedings of the National Academy of Sciences*, October, 201507071.


