7

SUMMARIZING DISCUSSION

Over the last years, it has become clear that pathogenic mycobacteria are highly dependent on type VII protein secretion (T7S) systems to successfully infect their host. In addition to the well-studied ESX-1 system, the ESX-5 system is thought to play an important role in virulence because of its exclusive presence in the slow growing, mostly pathogenic mycobacterial species. Work from our lab revealed that ESX-5 is responsible for secretion and cell-surface localization of numerous PE and PPE family proteins (described in **Chapter 2**). By performing zebrafish infection studies we could indeed establish a role for ESX-5 in mycobacterial virulence. The work described in this thesis provides valuable insight in the functional effects of ESX-5 and in addition, extends the knowledge on ESX-1. Furthermore, it enables us to paint a unique picture of mycobacterial virulence factors that are important for infection in different hosts. The characterization of these virulence mechanisms is of great importance for the development of treatment options for mycobacterial diseases, including TB. The knowledge acquired in this thesis can therefore contribute to finding new solutions for this devastating disease.

DISRUPTION OF ESX-5

In this work, we have made use of an *M. marinum* $espG_5$- transposon mutant to functionally study ESX-5. The results from our TraDIS experiments, described in **Chapter 6**, indicate that ESX-5 is essential for growth of the *M. marinum*. We found that mutants with transposon insertions in the genes coding for the systemic ESX-5 components $EccB_5$, $EccC_5$, $EccD_5$, $EccE_5$, $MycP_5$ and $EspG_5$ cannot or hardly be identified in the transposon mutant pool and are therefore likely to be essential. Previous efforts from our lab to knock-out these components, have indeed remained unsuccessful. How then could we have used an $espG_5$- transposon mutant? This specific mutant, isolated in a functional screen [1], was found to contain a transposon insertion at the second TA site from the 5'-end of the gene and in fact, the TraDIS data confirms that the first two TA sites of $espG_5$ can contain such an insertion. Possibly, protein translation starts at an alternative start codon in these transposon mutants, resulting in a minor modification of the EspG protein which only partially disturbs its function. The ESX-5 membrane complex is still intact in the *M. marinum* $espG_5$- transposon mutant (unpublished data, E. Houben). Analysis of its secretome showed that PE_PGRS secretion is strongly but not completely reduced to zero, suggesting that ESX-5 can function at a residual level assuming that these proteins are exclusively transported by ESX-5 (**Chapter 2**). Therefore, the *M. marinum* $espG_5$- transposon mutant allowed us to work with an ESX-5-defective strain, while ESX-5 in itself is essential. The essential nature of ESX-5 has recently also been described for the *M. tuberculosis* laboratory strain H37Rv, in which deletion of the $eccB_5$-$eccC_5$ operon was shown to be impossible [2]. Interestingly, transposon insertions in $eccC_5$ and $eccD_5$ were not lethal for the CDC1551 strain of *M. tuberculosis*. However, despite their strain differences, disruption or deletion of single core components does lead to severe growth defects for both *M. tuberculosis* CDC1551 and H37Rv strains, demonstrating that ESX-5 is of great importance for *M. tuberculosis* [2,3]. Thus far, it is not clear why ESX-5 plays such

an important role in mycobacteria. Its restriction to the slow-growing and generally more pathogenic species suggests that this acquired system is involved in virulence. However, its essentiality irrespective of host context points towards a more basal function in mycobacterial survival. As ESX-5 is a system that transports substrates across the mycomembrane, it is not unthinkable that it also functions as (nutrient) import system. Indeed, recent work from our lab has shown that the introduction of a pore in the outer membrane enabled the construction of *M. marinum esx-5* knock-out strains, suggesting that the presence of a membrane channel lies at the base of ESX-5 essentiality (Houben and Bitter, manuscript in preparation). Possibly, the introduction of ESX-5 in pathogenic mycobacteria has led to more flexibility in the biogenesis of outer membrane pores, which resulted in gene loss of classical porins such as MspA through evolutionary events, making ESX-5 essential for these species. An alternative explanation for ESX-5 essentiality could be the toxic accumulation of substrates when their export is blocked. As no less than one tenth of the pathogenic mycobacterium's coding capacity is dedicated to PE and PPE proteins, their inappropriate location within the bacterium or incorrect folding state could have a major effect on bacterial viability. In the ESX-5-defective *M. marinum espG*$_5$- transposon mutant, expression of PE_PGRS proteins is dramatically decreased. Other, more severely affected ESX-5 mutants, may not be able to regulate substrate expression that well, leading to lethal intracellular protein accumulation. A third possibility that could explain the necessity of ESX-5, is the requirement for secretion of a single or multiple substrates. The TraDIS data shows that indeed nine PE_PGRS proteins and two PPE proteins, such as ME11_4805 (Mmar_4955), ME11_5183 (Mmar_5362) and ME11_2743 (Mmar_2822), do not contain transposon insertions (**Chapter 6**), suggesting that ESX-5 could be essential due to the inability to secrete or transport these specific proteins to the cell wall. If these PE/PPE proteins would form porins in the outer membrane, as hypothesized in **Chapter 4**, this theory would fit well with our data showing that ESX-5 knock-out strains can only be generated in the presence of an extra membrane pore (Houben and Bitter, manuscript in preparation). Altogether, given the amount of substrates of ESX-5, the conservation of the secretion system across pathogenic species and its essential nature, the ESX-5-dependent PE/PPE proteins are probably very important for mycobacterial survival and virulence, as also discussed in the introductory chapter of this thesis.

ESX-5 AND DORMANCY

The research described in this thesis has led to the surprising discovery that *M. marinum* bacteria that are unable to use their ESX-5 protein secretion machinery, display increased virulence in a natural host (described in **Chapter 3**). Because we found in **Chapter 2** that the secretion of a substantial number of PE and PPE proteins is impaired in these bacteria, our results strongly suggest that these proteins are involved in the establishment of a chronic infection and prevent the induction of acute disease. ESX-5 deficiency in *M. marinum* was previously found to result in increased pro-inflammatory cytokine production by human macrophages [4]. In zebrafish embryos and phagocytic cells, it leads to attenuation of

bacterial growth (**Chapter 3 and 6**). Despite the attenuating effect of ESX-5 abrogation on *M. marinum* infection in these innate model systems, the hypervirulent phenotype observed in adult zebrafish suggests that ESX-5 specifically acts upon conditions that are only present in these fully developed animals. One of these conditions is the presence of an adaptive immune system, which starts to develop in zebrafish embryos after the first week of development. However, we could rule out a role for the adaptive immune system by using T-cell deficient adult zebrafish (**Chapter 3**). Another hypothesis was that ESX-5 would be involved in dormancy. Dormancy has been the key to success for *M. tuberculosis*. Because the bacterium reaches a low metabolic state with limited or no active cell division, it is able to prevent eradication by the human immune system and survive within hypoxic granulomas for decades until a chance occurs to spread to a new host (as described in the introductory chapter). Dormancy has been suggested to be an evolutionary old response in mycobacteria, that initially developed to deal with changing oxygen levels in the soil [5]. One of the important regulators of dormancy, the DosR regulon, is conserved across mycobacterial species, including those that cause non-tuberculous diseases (NTM) [6]. Antibiotic-susceptibility *in vivo* is low for these bacteria, indeed suggesting a dormant-like response. In addition, hypoxia-induced dormancy with high similarity to *M. tuberculosis* dormancy has been described for the non-pathogenic environmental species *M. smegmatis* [7]. Although these findings indicate that dormancy is not specific for slow-growing, pathogenic mycobacterial species, these bacteria have undergone several evolutionary adaptations to an intracellular lifestyle within their host. The dietary switch to lipids, which are abundantly present in host granulomas, could be one of these specific adaptations leading to bacterial survival during dormancy *in vivo*. As we observed in **Chapter 4**, ESX-5 seems to be involved in the uptake of host fatty acids, partly by the action of the cell wall localized PPE protein LipY. As a functional lipase domain was found to be dispensable for fatty acid uptake, LipY may exert this effect by binding these lipids and guiding them towards an import system, of which it may be part itself as well. It is possible that the strong attenuation of an *M. tuberculosis* mutant deleted for all ESX-5-associated *pe* and *ppe* genes may be partly contributable to defects in fatty acid uptake, leading to altered energy provision during infection [8]. Remarkably, this mutant was reported to be still capable of secreting PE_PGRS proteins [9]. In contrast, we observed that abrogation of ESX-5-dependent protein secretion (including PE_PGRS proteins) in *M. marinum* leads to induction of acute disease and possibly immediate progression towards active cell division in adult zebrafish (**Chapter 3**). If ESX-5-mediated uptake of fatty acids would be a prerequisite for dormancy, this system may be a good target for anti-mycobacterial treatment. By inhibiting dormancy, bacteria would probably remain susceptible to antibiotics. This way, duration of treatment could be shortened and latent infection prevented. However, direct progression to active disease requires timely and adequate treatment of patients, which is difficult in most TB endemic regions where recourses are scarce and healthcare is poor. In addition, these patients are not protected from re-infection, which is likely to occur in these regions as the majority of the population is latently infected with *M. tuberculosis*. In order to eradicate disease, preventing

progression of latent infection to active disease would be a better approach. To this end, strategies aimed at the induction of immunity to antigens that are specifically expressed during dormancy are now being explored [10]. As the prevention of infection itself is still far away at this point, improvement of current pre-exposure vaccines in order to prevent reactivation seems a more feasible approach. The immunogenic ESX-5-secreted PE and PPE proteins, which are expressed during chronic infection, could be good candidates to incorporate in such vaccines [11]. Recently, several of these proteins were described to be dominantly recognized by memory T cells of latently infected individuals that successfully contain infection [12]. Genes encoding these PE and PPE proteins were mainly localized on antigenic islands that also contained T7s-associated genes. These antigenic islands were proposed as main determinants of immunodominance [12]. However, it has to be noted that, as described in the introductory chapter of this thesis, individual *M. tuberculosis* strains display a high level of genetic and gene expression diversity for the PE and PPE proteins. Therefore, new targets to incorporate in vaccines should be chosen carefully.

FUNCTIONAL GENOMICS: GENE EXPRESSION PROFILING OF MYCOBACTERIA

In this work, we have made use of high-throughput RNA sequencing to determine gene expression profiles of *M. marinum* strains. The rapidly developing field of next-generation sequencing offers a world of possibilities, as massive amounts of information can be generated in relatively short time. Clear research questions and clever experimental design are essential to filter this information in order to be of value. By focusing on ESX-1-mediated protein secretion, we were able to identify a novel substrate and a transcriptional regulator of this system amongst a set of down-regulated genes in an *M. marinum* secretion mutant (**Chapter 5**). Gene expression profiles can provide important information on the bacterial factors required at different stages of infection. The pronounced up-regulation of the *espA* operon we observed in ESX-1-deficient *M. marinum* during the first hours of macrophage infection, suggests that the proteins encoded by this operon are part of an ESX-1-mediated early virulence mechanism that may precede the action of other ESX-1-substrates. The time of action and function of individual substrates of this highly important virulence system, are thus far unknown. Gene expression profiling at different stages of infection may be a good method to learn more about these dynamics. Over the last years, several studies have documented such experiments with wild-type strains of *M. tuberculosis*. Using micro-arrays, the mycobacterial transcriptome has been determined under several conditions, including infection *in vitro* and *in vivo* [13]. These studies have contributed to the identification of genes required for bacterial pathogenesis [14]. Recent developments in deep sequencing techniques and the decreasing costs for these initially expensive experiments, have now led to the use of RNA sequencing (RNA-Seq) for transcriptional studies on several organisms, including bacterial pathogens [15]. The major advantages of RNA-Seq over the old microarray techniques include the increased mapping and quantification accuracy

and sensitivity, the generation of actual transcript sequences, the high resolution and high reproducibility. In this work, we have employed this relatively new technique to determine transcriptional differences between *M. marinum* and *M. tuberculosis* wild-type and ESX-1-defective strains. Our results revealed a large degree of overlap in the transcriptional response of these two species, suggesting that ESX-1-associated gene transcription is regulated similarly. This first report describing the successful use of RNA-seq for mycobacteria has provided promising results. More RNA-Seq data for *M. tuberculosis* can be expected to become available in the near future. As the sequencing field will continue to develop, it is important to standardize terminology and data analysis methods in order to enable comparison of the increasing pool of data that is being generated. The TB Database Project (www.TBDB.org) has created a tool by which all published expression data can be easily explored, providing a useful resource of already existing information. Together, gene expression data can be instrumental in the characterization of mycobacterial virulence mechanisms and may lead to the identification of new drug targets [16]. However, gene expression data by itself is not enough to understand virulence. As post-transcriptional regulation, protein expression level alterations and the constitutive expression of virulence factors are not taken into account, additional research methods are necessary to draw a more complete picture of virulence-associated processes in mycobacteria.

GENOME SEQUENCING OF MYCOBACTERIA

The introduction of genome sequencing techniques started with the Sanger sequencing method in 1977 and has developed rapidly since [17]. Technological improvements enabled sequencing of whole genomes, which resulted in 1995 in the first sequenced genome of a free-living organism, *Haemophilus influenza* [18]. Three years later, the genome sequence of *M. tuberculosis* H37Rv was released [19]. Subsequent sequence polymorphism-based genome comparison studies revealed an unexpectedly high degree of genetic variation between different *M. tuberculosis* strains [20,21], which may affect outcome of infection and epidemiologic success. Many other studies aimed to determine sequence variation in specific genomic regions of *M. tuberculosis* and various other mycobacterial species followed. With the introduction of next-generation sequencing methods in 2005, which enabled high-throughput sequencing by running multiple reactions in parallel, sequencing costs decreased dramatically leading to a massive increase in its application. Complete genome sequences of several different clinical isolates of *M. tuberculosis* are now available (TBDB.org). The genome sequence of the *M. marinum* reference strain M has been determined in 2008 [22]. This pathogen, which causes TB-like disease in ectothermic animals, is able to infect a broad range of hosts. A genome comparison study of several *M. marinum* isolates derived from different sources showed that polymorphisms vary according to their host [23]. A strong differentiation between fish and human derived isolates was found, as was also shown in a previous study from our lab, which additionally showed major differences in virulence [24]. The observed diverse range of genotypes points

to strain-specific virulence mechanisms and ecological adaptation of *M. marinum*. Adaptation to a diverse range of hosts involves the development of specific virulence mechanisms. In Chapter 6, we used next-generation sequencing to determine host-specific virulence factors of a fish-derived *M. marinum* strain. To this end, we first determined the whole genome sequence of this *M. marinum* isolate (E11 or Mma11). In contrast to the 'human' isolate M, the E11 strain was isolated directly from sea bass and causes a more chronic infection in fish species, resembling human TB [24]. Its genome sequence can therefore provide valuable information on factors that determine pathogenicity. Genome sequence analysis revealed the presence of a large plasmid, which is absent from the M strain. Preliminary results from our lab indicate that this (conjugal) plasmid is involved in horizontal gene transfer and is present in several *M. marinum* isolates. Interestingly, it contains genes that are homologous to those that encode the type VII secretion systems (Bitter, manuscript in preparation). At this point, the function of the plasmid remains to be determined. In the study described in Chapter 6, we found that the *M. marinum* E11 strain is, similar to the M strain, able to infect a broad range of hosts. In order to determine the contribution of all *M. marinum* genes to virulence in phagocytic cells derived from each of these hosts, we made use of a transposon mutant library. With transposon-directed insertion site sequencing (TraDIS) and subsequent read mapping onto the E11 sequence, we were able to establish which *M. marinum* genes are essential, beneficial or disadvantageous for infection of five separate host organisms. Although similar transposon-based studies in mycobacteria have been performed before, our study is the first to compare infection requirements in different hosts. Such a unique study can only be performed with a broad-host pathogen like *M. marinum* and provides valuable insight in the common processes that mycobacteria require for infection in general. Importantly, it also shows which specific factors the pathogen needs to adapt to the human host. Based on the TraDIS data, we have already identified the putative transcriptional regulator CpsA as a novel *M. marinum* virulence factor that is important for infection of mammalian and fish phagocytes and zebrafish embryos. As this protein is conserved in both *M. tuberculosis* and *M. leprae*, it may be essential for virulence of these species as well. CpsA function and other interesting virulence-associated genes that the TraDIS data analysis revealed will be studied in the near future. Extra attention should be paid to mammalian cell-specific virulence mechanisms, as this may provide more information on the evolutionary processes that occurred in pathogens that have adapted to a human host, such as *M. tuberculosis*. Ultimately, the wealth of information obtained in this study my lead to the identification of potential vaccine candidate genes or drug targets for TB. In our *M. marinum* TraDIS study, we have shown that this technique works well for mycobacteria and gives highly reproducible results. In the future, TraDIS can be employed to answer several other research questions. For example, by using this technique on mycobacteria containing a mutation that restrict these bacteria to the phagosome, the genes specifically required for survival in this cellular compartment can be determined. Furthermore, it can be used to determine growth requirements

in the different environments that are encountered *in vivo*, such as varying oxygen levels or lipid content. TraDIS combined with a long-term infection of a natural host would especially be informative to determine the factors involved in pathogenicity and the establishment of a chronic infection. This could however be quite a challenge, considering the infection inocula required to achieve complete genomic coverage.

CONCLUDING REMARKS

This thesis describes several lines of work aimed at the characterization of virulence mechanisms of pathogenic mycobacteria. Using *M. marinum* as a model organism, we were able to study in depth the type VII protein secretion system ESX-5 and its contribution to virulence. Given the conservation of ESX-5 across pathogenic mycobacteria, our research results obtained with *M. marinum* may also be applicable to *M. tuberculosis*. Indeed, the process of ESX-5-mediated protein secretion is similar in both species. However, care should be taken when extrapolating results, as the available ESX-5 mutants of *M. tuberculosis* have varying phenotypes *in vivo* and the number, sequences and expression patterns of PE/PPE proteins differ substantially between the two species. Further research on the function of ESX-5 substrates is necessary to determine their importance for pathogenic mycobacteria. In this thesis, we have established a role for ESX-5 of *M. marinum* in virulence *in vivo* and lipid uptake *in vitro*, providing important information on the function of this system. Furthermore, we have shed light on the transcriptional pathways involved in ESX-1-mediated protein secretion. In a general experimental approach aimed at the simultaneous identification of multiple virulence factors, we could demonstrate the significance of ESX-5 and ESX-1. In addition, this study revealed many other virulence factors important for infection of specific or multiple hosts, providing valuable information on bacterial adaptation to the host. Altogether, these studies give more insight in the virulence mechanisms employed by pathogenic mycobacteria, which may aid in the development of improved vaccines and treatment options for TB.

REFERENCES

1. Abdallah AM *et al*. PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. Mol Microbiol (2009); 73(3): 329-40.
2. Di Luca M *et al*. The ESX-5 associated *eccB5-eccC5* locus is essential for *Mycobacterium tuberculosis* viability . PLoS One (2012); 7(12): e52059.
3. Houben EN *et al*. Composition of the type VII secretion system membrane complex. Mol Microbiol (2012); 86(2): 472-84.
4. Abdallah AM *et al*. The ESX-5 secretion system of *Mycobacterium marinum* modulates the macrophage response. J Immunol (2008); 181(10): 7166-75.
5. Boon C, Dick T. How *Mycobacterium tuberculosis* goes to sleep: the dormancy survival regulator DosR a decade later. Future Microbiol (2012); 7(4): 513-8.
6. Gerasimova A *et al*. Comparative genomics of the dormancy regulons in mycobacteria. J Bacteriol (2011); 193(14): 3446-52.
7. Dick T *et al*. Oxygen depletion induced dormancy in *Mycobacterium smegmatis*. FEMS Microbiol Lett (1998); 163(2): 159-64.
8. Sayes F *et al*. Strong immunogenicity and cross-reactivity of *Mycobacterium*

tuberculosis ESX-5 type VII secretion: encoded PE-PPE proteins predicts vaccine potential. Cell Host Microbe (2012); 11(4): 352-63.

9. Bottai D *et al*. Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Mol Microbiol (2012); 83(6): 1195-209.
10. Ottenhoff TH, Kaufmann SH. Vaccines against tuberculosis: where are we and where do we need to go? PLoS Pathog (2012); 8(5): e1002607.
11. Sampson SL. Mycobacterial PE/PPE proteins at the host-pathogen interface. Clin Dev Immunol (2011); 2011: 497203.
12. Lindestam Arlehamn CS *et al*. Memory T cells in latent *Mycobacterium tuberculosis* infection are directed against three antigenic islands and largely contained in a CXCR3+CCR6+ Th1 subset. PLoS Pathog (2013); 9(1): e1003130.
13. Haller R *et al*. The transcriptome of *Mycobacterium tuberculosis*. Appl Microbiol Biotechnol (2010); 86(1): 1-9.
14. Ward SK *et al*. Transcriptional profiling of *Mycobacterium tuberculosis* during infection: lessons learned. Front Microbiol (2010); 1: 121.
15. Wang Z *et al*. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet (2009); 10(1): 57-63.
16. Mukhopadhyay S *et al*. Pathogenesis in tuberculosis: transcriptomic approaches to unraveling virulence mechanisms and finding new drug targets. FEMS Microbiol Rev (2012); 36(2): 463-85.
17. Sanger F *et al*. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A (1977); 74(12): 5463-7.
18. Fleischmann RD *et al*. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science (1995); 269(5223): 496-512.
19. Cole ST *et al*. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature (1998); 393(6685): 537-44.
20. Fleischmann RD *et al*. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J Bacteriol (2002); 184(19): 5479-90.
21. Tsolaki AG *et al*. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc Natl Acad Sci U S A (2004); 101(14): 4865-70.
22. Stinear TP *et al*. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. Genome Res (2008); 18(5): 729-41.
23. Broutin V *et al*. Genetic diversity and population structure of *Mycobacterium marinum*: new insights into host and environmental specificities. J Clin Microbiol (2012); 50(11): 3627-34.
24. van der Sar AM *et al*. *Mycobacterium marinum* strains can be divided into two distinct types based on genetic diversity and virulence. Infect Immun (2004); 72(11): 6306-12.