Modeling metabolism of *Mycobacterium tuberculosis*

Rienk A. Rienksma

Modeling metabolism of *Mycobacterium tuberculosis*

Rienk A. Rienksma

Thesis committee

Promotors
Prof. Dr V.A.P. Martins Dos Santos
Professor of Systems & Synthetic Biology
Wageningen University & Research

Co-promotors
Dr P.J. Schaap
Associate Professor, Systems & Synthetic Biology Group
Wageningen University & Research

Dr M. Suárez-Diez
Assistent Professor, Systems & Synthetic Biology Group
Wageningen University & Research

Other members
Prof. Dr H.F.J. Savelkoul, Wageningen University & Research
Prof. Dr F.J. Bruggeman, VU Amsterdam
Dr D.E. Martens, Wageningen University & Research
Dr E.M. van Schothorst, Wageningen University & Research

Modeling metabolism of *Mycobacterium tuberculosis*

Rienk A. Rienksma

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 22 May 2019
at 1.30 p.m. in the Aula.

Contents

Chapter 1

Introduction

Mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis (TB), a disease that has been plaguing mankind for millennia, with estimates of its occurrence dating back 9,000 years. Currently, an astonishing one quarter of the world's population has latent tuberculosis. In these cases, Mtb is inactive and the disease cannot be transmitted [1].

As Mtb is inhaled via the lungs, it is phagocytosed by alveolar macrophages and resides in the phagosome. Knowledge about the environment Mtb encounters inside this phagosome has increased over the years as it is known to be hypoxic and nutrient poor [2], nitrosative, oxidative, and conductive to increased iron uptake by Mtb [3] and that Mtb cocatabolizes multiple amino acids, C1, C2, and C3 substrates within this phagosome, wherein the major carbon source is likely acetate or acetyl-CoA derived from host lipids [4].

Mtb is renowned for its waxy cell wall and its ability to remain 'dormant' within the host macrophage (latent tuberculosis). In such a dormant state Mtb does not grow, but is able to survive. This quiescent non-growing part of an Mtb population is particularly problematic, as these bacteria are phenotypically resistant, or 'drug tolerant'. When using an effective medicine, such as isoniazid, 99% of Mtb is killed during the first two days of administration. The drug tolerant part of the population however remains, resulting in current treatments requiring at least 6 months to reach acceptable cure. The situation can become direr with multiple drug resistant TB (MDR-TB), which is thought to have arisen by poor compliance to the drug administration regimen for normal TB. With this type of TB, Mtb does not respond to isoniazid and rifampicin, and expensive and toxic second line drugs are required. The total treatment period for MDR-TB is at least 2 years [1]. Extensively drug resistant TB (XDR-TB) is even worse as it is only cured in less than one-third of the cases, while a little over half of the cases of MDR-TB are cured.

These prolonged treatment periods and the rise of resistant Mtb requires the development of novel therapeutic interventions. Mtb metabolism provides opportunities for such development, as it encompasses over 1000 potential drug targets in the form of metabolic enzymes (Rienksma, 2018). If we are able to develop a model that can recreate the metabolic state of Mtb, especially for dormant or VBNC (viable but non-culturable) cells, we have a chance to pinpoint enzymes essential for survival in such a state and develop novel therapeutic interventions that target these enzymes.

Metabolic networks and metabolic models

Metabolism can be viewed as an interconnected network of nodes and edges, wherein the nodes represent metabolites and the edges represent reactions. These reactions can be transport reactions from one compartment to another, wherein

the compartments can be viewed as different enclosed spaces within or outside of the microorganism, such as the cytoplasm and the extracellular space. These reactions can also be conversions of one metabolite into another. In order to pinpoint essential enzymes, such a genome-scale metabolic network can be cast in a mathematical form resulting in a genome-scale constraint-based metabolic model (GSM). GSMs enable the prediction of metabolic states. A metabolic state is defined as the whole of fluxes or conversion rates (in mmol/h) throughout metabolism, including transport reactions. The earliest constraint-based model of Mtb metabolism was not 'genome-scale' but focused on the mycolic acid pathway, a specific kind of lipid produced by mycobacteria [5]. In time, various GSMs have been created, most based on other earlier versions (Figure 1.1).
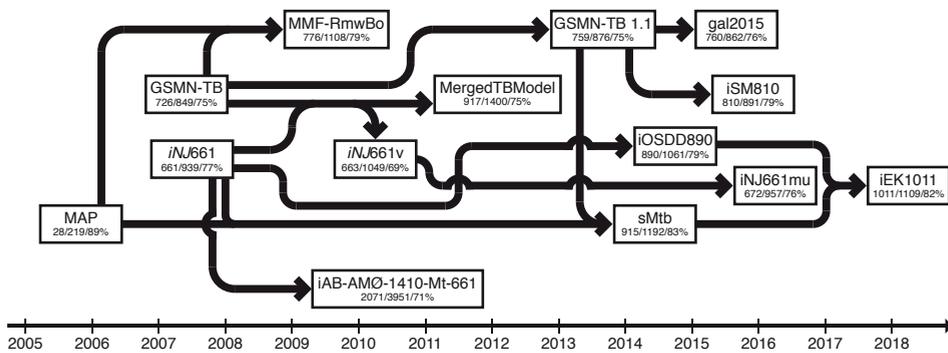

Figure 1.1 - Timeline of GSM of Mtb metabolism
Arrows indicate dependencies of one model on another. Models are given by rectangles, wherein the names of the models are given in the top line. The numbers represent from left to right: The number of genes, the number of reactions, and the percentage of reactions that are not orphan.

The metabolic state is captured as a vector of fluxes in vector **v**. The CB metabolic model basically consists of a series of steady state reaction differential equations describing each conversion of one metabolite to another and each transport reaction of a metabolite. Transport reactions can be uptake or secretion processes, or, in eukaryotes, transport over various cellular compartments. This series of steady state equations is founded on an underlying metabolic network wherein all metabolites are interconnected by conversion and transport reactions. All these steady state equations are captured in stoichiometric matrix, **S**, wherein the columns represent the various reactions and the rows represent the various metabolites associated with these reactions. A negative value in the stoichiometric matrix indicates consumption of the respective metabolite and positive value indicates formation. Metabolites that are not associated with a given reaction are indicated with a zero at the respective position in the stoichiometric matrix. The most straightforward approach for calculating a metabolic state is to construct

a Linear Programming Problem, LP-Problem, consisting of an objective and constraints. An LP-Problem based on a metabolic network is often referred to as a flux balance analysis problem (Text Box).

FBA Problems

An LP-Problem based on a metabolic network is often referred to as a flux balance analysis problem and has the following canonical form:

$$w = \max\{\mathbf{c} \cdot \mathbf{v}\}$$

subject to:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b}$$
$$\mathbf{l} \leq \mathbf{v} \leq \mathbf{u}$$

Wherein w indicates the optimal objective function value, dependent on the maximization (minimization is the alternative) of a vector of reaction weights, $\mathbf{c}$, and a vector of reaction fluxes, $\mathbf{v}$. The objective function value, w, as well as the associated vector of reaction fluxes, $\mathbf{v}$, are limited by constraints. These constraints are given below the 'subject to'-part. In this case, the first constraint specifies that the system is in steady state; hence vector $\mathbf{b}$ contains only zeros. The other constraint specifies that the values that individual fluxes can obtain are limited to a range defined by an upper bound vector, $\mathbf{u}$, and a lower bound vector, $\mathbf{l}$.

Solving a flux balance analysis problem as the one given above and thus simulating metabolism using a CB metabolic model is referred to as flux balance analysis (FBA). When this problem has been solved, the vector of reaction fluxes could assume different values while still obtaining the optimal objective function value, w. The range of these values is generally known as the solution space and in the case of CB metabolic models a set of fluxes within this range is known as a metabolic state. This solution space can be very large if the number of constraints is generally low and the difference between values in the $\mathbf{l}$ vector and their corresponding values in the $\mathbf{u}$ vector is large. As fluxes within microorganisms are very difficult to measure, uptake and secretion fluxes are mostly used to constrain GSMs and bring (a few) of the values in the $\mathbf{l}$ vector and $\mathbf{u}$ vector closer together. When biomass concentration and nutrient concentrations are measured over time in a controlled environment, these (specific, i.e. per weight of biomass) uptake and secretion fluxes can be calculated.

Experimental data to create genome-scale models
Obviously, an important part of a GSM is the $\mathbf{S}$-matrix, which is based on a metabolic network. Such a metabolic network is based on enzymes and their function. Therefore, data on substrates and products of metabolic enzymes is a prerequisite to create such a network. This data mainly comes from two sources: biochemical enzymology experiments to provide evidence on the functioning of individual enzymes and bioinformatics based on the Mtb genome, to provide predictions of functioning of individual enzymes.

Another valuable source of data is the biomass composition of Mtb, acquired by measuring the fraction of as many specific compounds as possible that make up Mtb biomass. This biomass composition can subsequently be cast into a biomass

reaction wherein a plurality of these compounds is consumed. Subsequent maximization of this biomass reaction gives a prediction of the maximal specific growth rate attainable under the given constraints. Conversion of a metabolic network in a GSM can be visually represented (Figure 1.2). The sum of the molecular weights of each metabolite multiplied by its respective number in the stoichiometric matrix should equal zero for each column in the S-matrix. In other words, there should be no net accumulation or consumption of any metabolite.
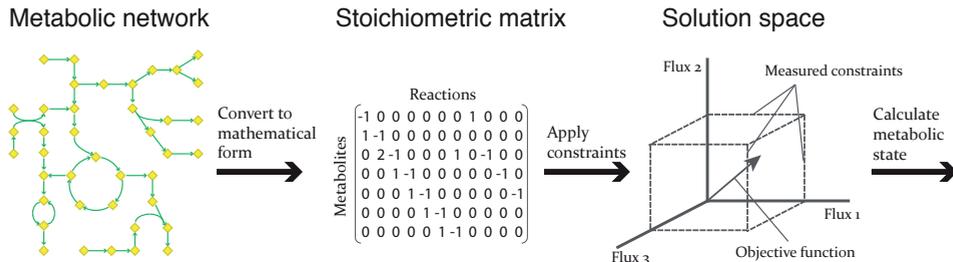

Figure 1.2 - Converting a metabolic network into a GSM [6]
Each reaction (green arrows) converting one metabolite (yellow diamond) into another within a metabolic network (left panel) is written down in a stoichiometric matrix (middle panel). In this **S**-matrix metabolites that are consumed are denoted by a negative number, while metabolites that are produced are denoted by a positive number. By adding (measured) constraints to the model and optimizing for an objective, a metabolic state can be calculated (right panel).

Experimental data to shrink the solution space

By constraining a GSM, using the **l** and **u** vectors, such that only nutrients available to a given microorganism can be taken up, a metabolic state of the organism growing on those nutrients can be predicted. To sufficiently shrink the solution space, it is useful to know a couple of uptake/secretion rates and have exact knowledge on the types of nutrients available. In general, it is preferred that the uptake rates of the main (or more preferably all) carbon, nitrogen, and oxygen sources are known. Insufficient knowledge on the uptake/secretion rates and the nutrients available quickly leads to a very large solution space and an inability to predict the metabolic state.

Genome-scale data

Genome-scale data, such as gene expression data and protein expression data form valuable resources to shrink the solution space. A general assumption is that these two types of expression data are somehow related to flux. Protein expression data is more closely related to flux as compared to gene expression data and thus represents the more favorable type of data.

Metabolomics provides a source of data that is perhaps the most closely related to flux. However, an inherent problem of (untargeted) metabolomics is that it provides no information about flux differences [7]. Metabolic flux analysis using

stable isotopes, such as ^{13}C, do provide information on fluxes and have been used for Mtb [4, 8, 9]. This information is however indirect, as it must be inferred via a metabolic network.

Perhaps the most reliable and direct information on fluxes is provided by measuring nutrient concentrations (e.g. with HPLC) and biomass in a medium over time and use these to calculate the uptake rates.

Methods to integrate genome-scale data and GSMs

A wealth of methods have been created to integrate genome-scale data, particularly gene expression data, such as GIMMI [10], iMAT [11], MADE [12], E-Flux [13], PROM [14], E-Fmin [15], uFBA [16], and TRFBA [17], just to name a few. These methods all try to "overlay" the respective data and the GSM to infer fluxes from the data.

Predictions by genome-scale models

If sufficient data on these rates and nutrients is available, it allows the modeler to make predictions on essential genes under those conditions. In a GSM each gene is associated to (part of) an enzyme product, which is in turn associated to at least one reaction, i. If the enzyme product is essential for the reaction to carry flux, its flux boundaries are set to zero ($l_i = u_i = 0$), and the biomass reaction is maximized, resulting in the optimal objective function value, w. With this value, w, a prediction can be made as to whether the respective gene is essential. In general, if the ratio of the optimal objective function value obtained with the aforementioned constraint, $w_{knockout}$, to the optimal objective function value obtained without such constraint, $w_{wildtype}$, falls below a certain limit, the gene is deemed essential [18].

The specific growth rate, i.e. the flux through the biomass reaction, is often predicted using GSMs. These predictions can serve to validate a GSM or discrepancies between predictions and data can lead to the discovery of novel metabolites and/ or enzymes [19].

Uptake and secretion rates can be measured in steady state conditions and subsequently be used to constrain the metabolic state of the organism. In general, the number of constraints increases with the number of measured uptake and secretion rates. These constraints subsequently reduce the size of the solution space, allowing a more accurate metabolic state prediction [19]. Although it is also possible to predict uptake rates using GSMs [20], this is not common procedure.

Outline of this thesis

The objectives of this thesis are:

- Provide a genome-scale constraint-based model (GSM) of Mtb metabolism that is up to date and fully mass balanced.
- Use this GSM to simulate the metabolic state within the host.
- Point out vulnerable metabolic pathways of Mtb in an in-host environment.

In 2007 two large GSMs of Mtb had been published: GSMN-TB (genome scale metabolic network tuberculosis) and iNJ661 (*in silico* Neema Jamshidi 661 genes) [21, 22]. The almost simultaneous publication of both models resulted in two independently created, and quite different models. In **chapter 2**, these two models are combined together with a model of the mycolic acid synthesis pathway, MAP [5], to create a consensus GSM, sMtb (*in silico* Mycobacterium tuberculosis). Model sMtb is tested for its ability to predict an *in vitro* metabolic state and its ability to correctly identify essential genes. This performance is compared with iNJ661 and GSMN-TB.

Despite this increased knowledge, a precise picture of the nutrients available and their role in uptake and secretion rates is still unknown. Therefore, an experimental model representing a real infection scenario, wherein *Mycobacterium bovis* BCG infects human macrophage-like THP-1 cells is studied using RNA-sequencing to get a complete genome-wide picture, is described in **chapter 3**. The thus obtained dual RNA-seq data can be used to constrain GSMs by assuming a linear relationship between such data and metabolic fluxes catalyzed by the encoded enzymes. This results in condition-specific models, as gene expression data used to constrain the metabolic fluxes within the GSM are associated with a given condition [23, 24]. A slightly different approach is given **chapter 4** where condition-specific objective functions are formulated to predict the metabolic state of Mtb inside the host.

By assuming that anti-TB drugs are not always fully effective, the effect of known drugs on the metabolic state of Mtb is assessed. Using model sMtb, some pathways are predicted to become more important during administration of these drugs to a patient in need thereof, while other pathways become less important. To obtain a more detailed understanding of the metabolic processes that occur during infection, the metabolism of both Mtb and its host, the human macrophage, need to be taken into account. Therefore, in **chapter 5**, the flux rerouting process upon drug application is visualized using a combined GSM of Mtb and its host, and vulnerable pathways are highlighted. Finally, a general discussion and outlook is given in **chapter 6**.

References

1. WHO, *Global Tuberculosis Report 2017*. 2017, France: World Health Organisation.
2. James Philip, E., Y. Grinberg Oleg, G. Michaels, and M. Swartz Harold, *Intraphagosomal oxygen in stimulated macrophages.* Journal of Cellular Physiology, 2005. **163**(2): p. 241-247.
3. Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan, and G.K. Schoolnik, *Transcriptional Adaptation of* Mycobacterium tuberculosis *within Macrophages: Insights into the Phagosomal Environment.* J Exp Med, 2003. **198**(5): p. 693-704.
4. Beste, Dany J.V., K. Nöh, S. Niedenführ, Tom A. Mendum, Nathaniel D. Hawkins, Jane L. Ward, Michael H. Beale, W. Wiechert, and J. McFadden, *¹³C-Flux Spectral Analysis of Host-Pathogen Metabolism Reveals a Mixed Diet for Intracellular* Mycobacterium tuberculosis. Chemistry & Biology, 2013. **20**(8): p. 1012-1021.
5. Raman, K., P. Rajagopalan, and N. Chandra, *Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs.* PLoS Comput Biol, 2005. **1**(5): p. e46.
6. Rienksma, R.A., M. Suarez-Diez, L. Spina, P.J. Schaap, and V.A.P. Martins dos Santos, *Systems-level modeling of mycobacterial metabolism for the identification of new (multi-) drug targets.* Seminars in Immunology, 2014. **26**(6): p. 610-622.
7. Zamboni, N., A. Saghatelian, and G.J. Patti, *Defining the Metabolome: Size, Flux, and Regulation.* Molecular cell, 2015. **58**(4): p. 699-706.
8. Beste, D.J., B. Bonde, N. Hawkins, J.L. Ward, M.H. Beale, S. Noack, K. Noh, N.J. Kruger, R.G. Ratcliffe, and J. McFadden, *¹³C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in mycobacteria which requires isocitrate lyase and carbon dioxide fixation.* PLoS Pathog, 2011. **7**(7): p. e1002091.
9. Basu, P., N. Sandhu, A. Bhatt, A. Singh, R. Balhana, I. Gobe, N.A. Crowhurst, T.A. Mendum, L. Gao, J.L. Ward, M.H. Beale, J. McFadden, and D.J.V. Beste, *The anaplerotic node is essential for the intracellular survival of* Mycobacterium tuberculosis. Journal of Biological Chemistry, 2018. **293**(15): p. 5695-5704.
10. Becker, S.A. and B.O. Palsson, *Context-Specific Metabolic Networks Are Consistent with Experiments.* PLoS Comput Biol, 2008. **4**(5): p. e1000082.
11. Shlomi, T., M.N. Cabili, M.J. Herrgard, B.O. Palsson, and E. Ruppin, *Network-based prediction of human tissue-specific metabolism.* Nat Biotech, 2008. **26**(9): p. 1003-1010.
12. Jensen, P.A. and J.A. Papin, *Functional integration of a metabolic network model and expression data without arbitrary thresholding.* Bioinformatics, 2011. **27**(4): p. 541-547.
13. Colijn, C., A. Brandes, J. Zucker, D.S. Lun, B. Weiner, M.R. Farhat, T.Y. Cheng, D.B. Moody, M. Murray, and J.E. Galagan, *Interpreting expression data with metabolic flux models: predicting* Mycobacterium tuberculosis *mycolic acid production.* PLoS Comput Biol, 2009. **5**(8): p. e1000489.

14. Chandrasekaran, S. and N.D. Price, *Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in* Escherichia coli *and* Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences, 2010. **107**(41): p. 17845-17850.

15. Song, H.-S., J. Reifman, and A. Wallqvist, *Prediction of Metabolic Flux Distribution from Gene Expression Data Based on the Flux Minimization Principle.* PLOS ONE, 2014. **9**(11): p. e112524.

16. Bordbar, A., J.T. Yurkovich, G. Paglia, O. Rolfsson, Ó.E. Sigurjónsson, and B.O. Palsson, *Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics.* Scientific Reports, 2017. **7**: p. 46249.

17. Motamedian, E., M. Mohammadi, S.A. Shojaosadati, and M. Heydari, *TRFBA: an algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data.* Bioinformatics, 2017. **33**(7): p. 1057-1063.

18. Bordbar, A., J.M. Monk, Z.A. King, and B.O. Palsson, *Constraint-based models predict metabolic and associated cellular functions.* Nat Rev Genet, 2014. **15**(2): p. 107-20.

19. O'Brien, E.J., J.M. Monk, and B.O. Palsson, *Using Genome-Scale Models to Predict Biological Capabilities.* Cell, 2015. **161**(5): p. 971-987.

20. Edwards, J.S., R.U. Ibarra, and B.O. Palsson, In silico *predictions of* Escherichia coli *metabolic capabilities are consistent with experimental data.* Nature Biotechnology, 2001. **19**: p. 125.

21. Jamshidi, N. and B. Palsson, *Investigating the metabolic capabilities of* Mycobacterium tuberculosis *H37Rv using the* in silico *strain iNJ661 and proposing alternative drug targets.* BMC Syst Biol, 2007. **1**: p. 26.

22. Beste, D., T. Hooper, G. Stewart, B. Bonde, C. Avignone-Rossa, M. Bushell, P. Wheeler, S. Klamt, A. Kierzek, and J. McFadden, *GSMN-TB: a web-based genome-scale network model of* Mycobacterium tuberculosis *metabolism.* Genome Biol, 2007. **8**: p. R89.

23. Blazier, A.S. and J.A. Papin, *Integration of expression data in genome-scale metabolic network reconstructions.* Front Physiol, 2012. **3**: p. 299.

24. Machado, D. and M. Herrgard, *Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism.* PLoS Comput Biol, 2014. **10**(4): p. e1003580.
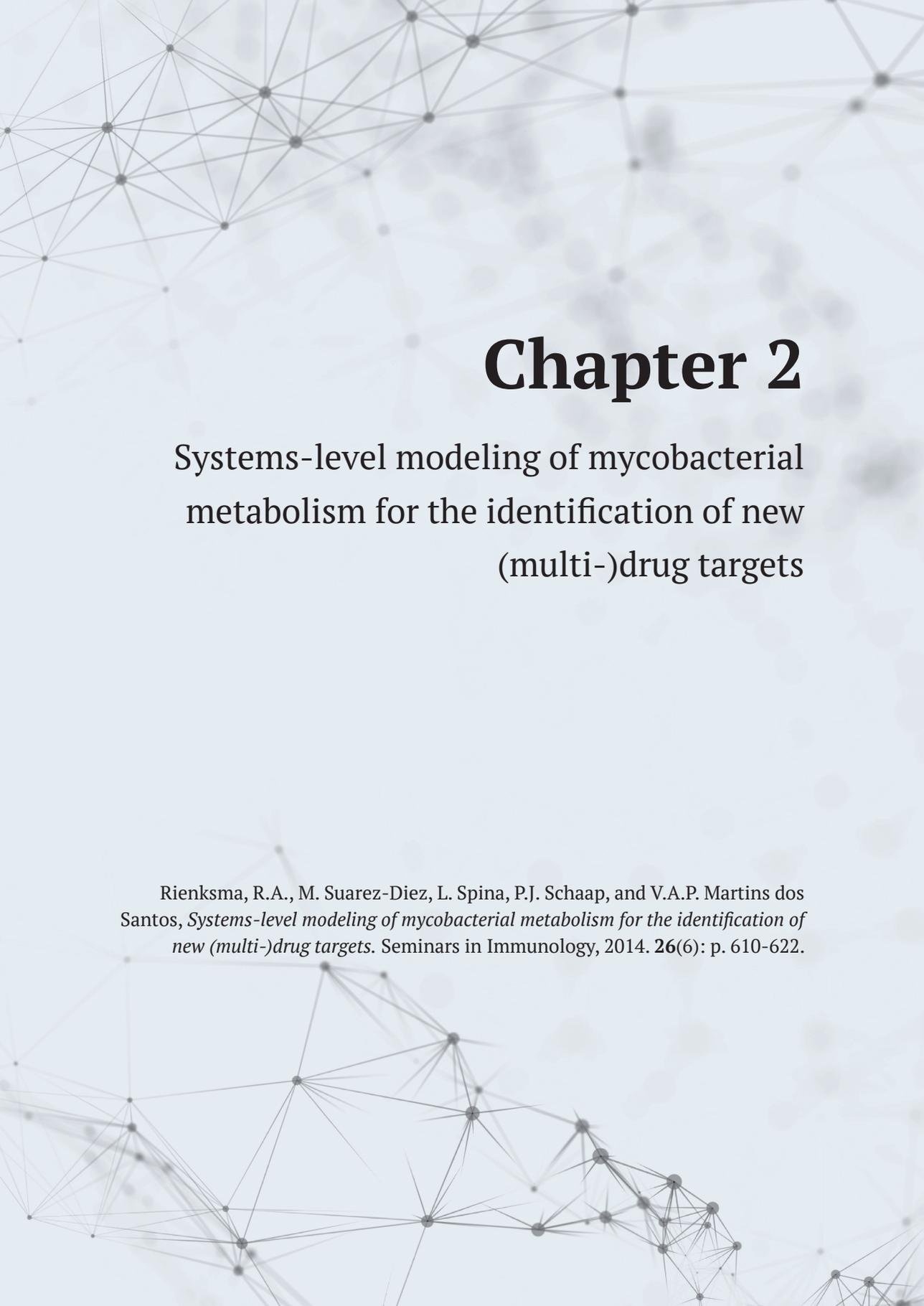
Chapter 2

Systems-level modeling of mycobacterial metabolism for the identification of new (multi-)drug targets

Abstract

Systems-level metabolic network reconstructions and the derived constraint-based (CB) mathematical models are efficient tools to explore bacterial metabolism. Approximately one-fourth of the *Mycobacterium tuberculosis* (Mtb) genome contains genes that encode proteins directly involved in its metabolism. These represent potential drug targets can be systematically probed with CB models through the prediction of genes essential (or the combination thereof) for the pathogen to grow. However, gene essentiality depends on the growth conditions and, so far, no *in vitro* model precisely mimics the host at the different stages of mycobacterial infection, limiting model predictions. These limitations can be circumvented by combining expression data from *in vivo* samples with a validated CB model, creating an accurate description of pathogen metabolism in the host. To this end, we present here a thoroughly curated and extended genome-scale CB metabolic model of Mtb quantitatively validated using ^{13}C measurements. We describe some of the efforts made in integrating CB models and high-throughput data to generate condition specific models, and we will discuss challenges ahead. This knowledge and the framework herein presented will enable to identify potential new drug targets, and will foster the development of optimal therapeutic strategies.

Keywords: *Mycobacterium tuberculosis*, metabolic model, constraint-based metabolic model, gene essentiality, metabolic state, systems biology.

The rise of multi-resistant *Mycobacterium tuberculosis* and the need for new intervention strategies

2

Mycobacterium tuberculosis (Mtb) is the etiological agent of tuberculosis (TB) and has re-emerged as a serious threat for human health. In 2012, TB claimed the lives of 1.3 million people [1]. The rapid appearance of multi, extensively and totally drug-resistant strains, emphasizes the adaptability of Mtb and has raised concerns of its impact to human health. Furthermore, due to the diverse genetic predisposition of the infected subjects, uncertainties on long-term adverse effects and other safety concerns regarding the rise of drug resistant strains, the development of new, effective and affordable TB drugs has been slow [2]. New (combined) therapeutic strategies are urgently required to combat these drug-resistant strains [3].

In vitro studies have revealed sets of genes that are essential for growth and survival under laboratory growth conditions [4, 5]. Due to the differences between the *in vivo* and the *in vitro* environments this does not automatically imply that these sets of genes are suitable drug targets. Besides, given all cellular components from different types of networks, genes (and their products) that may be not essential on their own can be indispensable in combinations not immediately obvious. A vital improvement would be the expansion of these studies to *in vivo* or *ex vivo* models, such as animal models, which would as faithfully as possible mimic the onset and progression of the infection, as well as the strategies against it [6]. An alternative and complementary method to identify suitable drug targets is to use mathematical descriptions of the metabolism of Mtb under *in vivo* conditions, circumventing experimental difficulties that arise with *in vivo* and *ex vivo* studies.

Approximately one-fourth of the annotated mycobacterial gene pool encodes structural proteins known to be involved in its metabolism presenting a wealth of enzymes and metabolites as potential drug targets. Stoichiometric genome-scale models of metabolism are essential to identify possible metabolic drug targets, as they provide a holistic view on metabolism. Drug targets in the form of enzymes encoded by their specific genes, have been identified by gene essentiality predictions based on modeling the *in vivo* environment [7]. Recent insights have clarified the picture of available metabolites to Mtb inside the host and shed new light on *in vivo* gene essentiality predictions [8-11].

Predictions on gene essentiality can be done using constraint-based (CB) metabolic models by simulating the effect of total loss of an enzyme function in a metabolic network. This black and white scenario where a drug is able to completely shut down an enzymatic reaction is not fully realistic. In most cases, drug effects are subtler, leading to only a partial loss of function [12]. Furthermore, and owing to the network structures in which they are embedded, genes may code for proteins that are not essential *per se*, but which do become so if equally non-essential

proteins to which they are connected become dysfunctional or absent. A reliable metabolic network topology, knowledge of the available metabolites in the host, *in vivo* growth and survival requirements and strategies, and reliable and quantitative predictions of metabolic activity are important and thus far overlooked.

A stoichiometric genome-scale CB metabolic model that is experimentally validated, not only qualitatively for the correct network topology, but also quantitatively for predicting fluxes, provides many opportunities to further identify metabolic bottlenecks and weak spots. Instead of using only qualitative, topology based, methods such models can be explored for new drug targets and novel synergistic drug combinations using more realistic quantitative approaches. For example, in addition to simulating the effect of a knockout of given genes or combinations thereof, the effect of a partial loss of function induced by a drug can also be simulated. Simulating the effect of decreasing the function of enzymes that can be targeted with known drugs can highlight alternative metabolic escape routes that become more relevant under these conditions paving the way to the development of more efficient therapeutic strategies.

Here we present a new genome-scale CB model of Mtb metabolism, sMtb (*in silico Mycobacterium tuberculosis*), which builds upon three previously published models and which is experimentally validated in great detail. Our model also includes recently discovered or annotated reactions and pathways, has undergone extensive manual curation and outperforms its predecessors in terms of both qualitative and quantitative predictions. We discuss the applications of this model for the identification of possible drug targets, to the unraveling of potentially unknown interconnections and for the development of future intervention strategies.

Mathematical models of metabolism

There are different types of metabolic models, all of them based on networks of metabolites that are interconnected through enzymatic, spontaneous, or transport reactions. These metabolic networks are reconstructed from literature and annotated genome data.

CB metabolic models are stoichiometric, mass, charge and energy-balanced scaffolds that describe steady-state kinetics, whereas dynamic metabolic models are explicitly time-dependent and enable to determine the changes in the concentration of metabolites over time. Thus, dynamic metabolic models enable more accurate descriptions of metabolism, but require many detailed kinetic parameters, such as rate-constants of every enzyme. Such kinetic parameters are often unknown and obtaining them experimentally is often difficult or impossible. Therefore, for a genome-scale dynamic model, many of these parameters are

unavailable and many of them would have to be fitted to the model, which would diminish its predictive power. In addition, simulations with these models are computationally costly, making dynamic models thus far unsuitable to describe metabolism on a genome-scale.
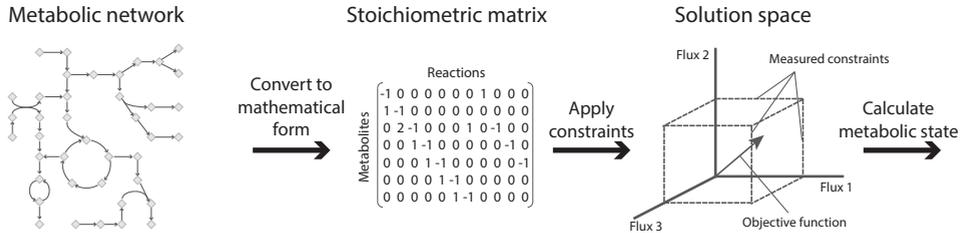

Figure 2.1 - Constraint-based model creation and functioning.
A scaffold metabolic network is constructed from an annotated genome and completed after a rigorous survey of organism specific databases and literature. This metabolic network represents all the different possibilities for metabolites to travel through the network (metabolic states). After this network has been constructed, a stoichiometric matrix is created that encompasses the stoichiometry of all metabolic reactions under steady state conditions. Constraints on uptake and/or secretion rates are subsequently set, and the optimization of one or multiple objectives leads to the prediction of a metabolic state.

Genome-scale CB metabolic modeling provides a holistic view on metabolism and transport. A metabolic network forms the foundation of a CB metabolic a model (Figure 2.1). The stoichiometry of each reaction is written in stoichiometric matrix where negative numbers represent the consumption of metabolites and positive numbers represent the formation of metabolites. This stoichiometric matrix ensures that the system is in steady-state, as for every reaction no metabolite can accumulate. Through the application of constraints, hence they are called 'constraint-based', the number of possible metabolic states can be lowered, to best predict the actual metabolic state of an organism under given genetic and environmental conditions [13]. Applying too many constraints can result in an infeasible model where no possible metabolic state can be found. CB metabolic models can be used to predict genes [14] and metabolites that are essential to synthesize precursors for growth [15]. A major advantage of genome-scale CB metabolic models as compared to dynamic models is that few parameters are required to describe the entire known metabolism of an organism. On the other hand, CB metabolic models are not easily adapted to describe the dynamics of the system, since they contain a stoichiometric matrix and are thus designed to operate in steady-state conditions where uptake and secretion fluxes are constant and there is no net accumulation of metabolic intermediates, which is only valid if the time scales under consideration are different enough. These metabolic models are based on optimization principles and need one or more optimization objectives to function. Optimization objectives in CB metabolic models can be

multiple and describe what the organism 'aims' for. Examples of frequently used metabolic objectives are: maximizing the speed at which an organism grows, maximizing the production of energy carrying metabolites (such as ATP), and minimizing the overall usage of enzymes [16].

Flux predictions

Flux is a commonly used concept in physics where it is defined as the rate of flow of a magnitude or property through a defined area [17]. In the realm of CB metabolic models, this term is used to indicate the rate of conversion of one metabolite to another per unit of biomass (usually given in mmol gdw^{-1} h^{-1}, where gdw denotes grams of cell dry weight). For transport reactions, there is no metabolite conversion and the term flux refers to the rate of transportation between cellular or sub-cellular compartments. Fluxes can have positive or negative values in CB metabolic models, depending on whether a forward or reverse reaction is predicted. The metabolic state, flux state or flux distribution of an organism is defined as the whole of all fluxes throughout metabolism [18, 19]. Constraints can be placed on some of these fluxes (e.g. the uptake and secretion rates) to limit the model. These constraints reflect the limitations of enzymes, transport proteins or nutrients and lead, upon optimization of one or multiple objectives, to meaningful flux distributions.

Objective functions

An important assumption of CB metabolic models is that optimization principles underpin metabolic states. In other words, the model assumes that a cell 'strives to achieve a metabolic objective' [20]. CB metabolic models are underdetermined and can be solved mathematically, which requires the optimization of one or multiple objective functions. Most genome-scale CB metabolic models contain one or multiple biomass functions. A biomass function is an integral part of a CB metabolic model and entails the amounts (in mmol) of metabolites that are required to form one gram dry weight of biomass and as such represents growth of the organism. The amounts of the individual metabolites are usually based on literature about the organism and vary for different reconstructions. Maximization of the flux through the biomass function thus leads to a prediction of the metabolic state when maximal growth is achieved, given a defined set of available nutrients.

Schuetz and colleagues [16] used a model of the central carbon metabolism of *Escherichia coli* to systematically compare flux distributions, resulting from 11 objective functions, to ^{13}C-determined *in vivo* flux distributions from six growth conditions. They concluded that no single objective best describes all conditions and the most relevant objective for each condition has to be identified.

Solution space

The solution space of a CB metabolic model (represented as a dashed cube in Figure 2.1) is defined as the range in which fluxes can vary while leading to the optimal value of the objective function. An inherent property of CB metabolic models is the fact that, even after optimizing a given objective function, the solution space remains largely undetermined. This region of feasible metabolic flux distributions grows larger with increasing model size and reflects the metabolic flexibility of living organisms. Once the solution space has been defined, Markov chain Monte Carlo sampling [21] or variations thereof [22] can be used to obtain probability distributions for the fluxes and extract descriptors (such as means and standard deviations) for these distributions. Such an approach gives an indication of which fluxes can be accurately determined under a given set of constraints, and which cannot. Moreover, an estimation of the significance of the change of each flux between different conditions can be provided.

Predictions of specific growth rates

CB metabolic models can quantitatively predict specific growth rates, or growth yields. Therefore, a comparison between predicted and experimentally determined values, such as the specific growth rate, provide the means to test the accuracy of the model. Constraints are set on the set of experimentally measured uptake and/or secretion rates while the uptake of other available metabolites (if any) is left unconstrained. Subsequently, the biomass function is set as the objective to maximize, which results in a predicted maximal specific growth rate.

These quantitative validations are limited since only one predicted 'flux' value, the specific growth rate, is compared to experimental data. Due to the inherent uncertainty provided by the size of the solution space, not all metabolic fluxes can be predicted with equal accuracy. However, many of these fluxes can still be predicted within a narrow range. Comparing multiple predicted fluxes to experimentally measured or experimentally inferred fluxes provides a much more solid and quantitative validation of CB metabolic models.

The importance of updating models

An example that illustrates the importance of updating CB metabolic models is the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate catalyzed by PfkA and/or PfkB in Mtb. Within different models, the enzymes and their interaction in catalyzing this ATP driven reaction are annotated differently. In one Mtb model this reaction can only occur if PfkA and PfkB are both present, while in another this reaction can occur if either PfkA or PfkB is present [23, 24]. However, Phong and colleagues [25] showed that only PfkA catalyzes the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate whereas PfkB does not. Thus, clearly

both models should be updated. This is one of the examples that show that it is important not only to create consistent models, but also to continuously update them. CB metabolic models organize and integrate the knowledge on metabolism and transport into a well-defined network. Therefore, CB metabolic models enable to systematically explore the metabolic capacities of organisms under a broad range of conditions and allow assessing the effect of perturbations (genetic or environmental) on the underlying metabolic network. On the basis thereof, these analyses subsequently enable generating experimentally testable hypotheses, making predictions over a range of conditions, and to provide invaluable insights that cannot be obtained if not from a systems perspective.

Merging of metabolic models

Two or more independently created CB metabolic models of the same organism will likely contain many common reactions and metabolic pathways. Owing to the specific emphasis and expertise of the model builders, it is also likely that both models would describe different parts of metabolism or the same pathways with different detail level. To preserve the knowledge in these models, a logical step is combining them into one comprehensive or consensus model. Merging two or more CB metabolic models describing the same organism might seem, at first sight, a straightforward task. Nevertheless, it can prove quite time consuming and full of unexpected challenges, such as those associated with the so-called namespace problem, derived from using different names for the metabolites [26]. This complicates the automatic identification of compounds common to both models. This implies that manual curation is still required to identify similar reactions and remove discrepancies.

Topological and elemental balancing inconsistencies

In a CB metabolic model all reactions must be stoichiometrically balanced so that there is no net internal production of any metabolite. Software tools, such as the COBRA Toolbox [27] include functionalities to inspect the model and detect unbalanced reactions. These tools require all metabolites in the model to be annotated with their chemical formula, which is the case for iNJ661 but not for GSMN-TB 1.1. Moreover, GSMN-TB 1.1 does not explicitly contain water or protons (apart from transport reactions and respiration), making it impossible to verify whether the reactions are elementally balanced.

Futile cycles are metabolic routes with no net gain. The existence of futile cycles in a metabolic network expands the solution space and complicates flux predictions. In some cases, these cycles are inherent to the biology of the studied organism. However, they can also appear as a result of an overlooked doubling of reactions, or by wrongly assigned reaction directionality. These types of futile cycles are

harmless from the model point of view, as long as they do not lead to net production or degradation of metabolites. Otherwise, they render the model unbalanced and model predictions can become unreliable.

History of CB metabolic models of *Mycobacterium tuberculosis*

Figure 2.2 shows the timeline of the successive CB models of Mtb metabolism that have been reconstructed since 2005. The very first CB metabolic model described the synthesis of triacylglycerol from glucose in human adipose tissue in 1986 [28]. Nearly two decades later, in 2005, the first CB metabolic model of Mtb appeared [14]. This model (MAP) was a detailed description of the mycolic acid synthesis pathway. Mycolic acids are long chain fatty acids that are unique to mycobacteria and essential for their survival [29]. In 2007, two genome-scale CB metabolic models of Mtb were independently published. Even though both models, GSMN-TB (Genome Scale Metabolic Network Tuberculosis) [30] and iNJ661 (*in silico* Neema Jamshidi, 661 genes) [23], describe the same organism, there are a number of substantial differences between them. GSMN-TB is arguably more complete than iNJ661, as it contains more genes (726 as compared to 661) and it also accounts for the methylcitrate cycle, which is critical for intracellular growth of Mtb [31]. iNJ661 has a more detailed annotation containing chemical formulas for each metabolite (except for some groups of metabolites that are lumped together and protein-metabolite complexes) and it is topologically more consistent, since it contains no duplicated reactions or metabolites. In 2009, Colijn and colleagues [32] metabolically interpreted gene expression data to predict the impact of 75 different drugs, combinations of drugs and media compositions on the mycolic acid synthesis capacity of Mtb. The mycolic acid synthesis pathway is described with greater detail within MAP than in GSMN-TB. Therefore, all mycolic acid reactions in GSMN-TB were replaced with the mycolic acid reactions from MAP creating a more comprehensive model (indicated by MMF-RmwBo in Figure 2.2). In the beginning of 2010, Fang and colleagues [33] used a semi-automatic method to create a model more compatible with *in vivo* conditions, iNJ661v, which optimally reproduced *in vivo* gene essentiality measurements. For completeness, this model was supplemented with reactions and metabolites from GSMN-TB and with the methylcitrate cycle. In the same year, Bordbar and colleagues [7] created the first macrophage-Mtb combined model. This dual model combined iNJ661 with a cell-specific alveolar macrophage model derived from the first human metabolic reconstruction [34]. High-throughput host gene expression data from *ex vivo* infected macrophages were integrated in the model to distinguish three different forms of tuberculosis: latent, pulmonary and meningeal. In 2011 Chindelevitch and colleagues developed MetaMerge, an algorithm to combine two CB metabolic models, and used it to merge iNJ661 and GSMN-TB [35]. The joining of both models

by MetaMerge is an automated process, therefore manual curation is still required to select correct reactions from highly similar reactions derived from both models and to identify metabolites that could not automatically be assigned to a database identifier, or whose chemical formula could not be determined. In 2013, an improved and extended version of GSMN-TB, GSMN-TB 1.1 appeared [24]. GSMN-TB 1.1 contains the cholesterol degradation pathway and additional corrections to the original GSMN-TB model.
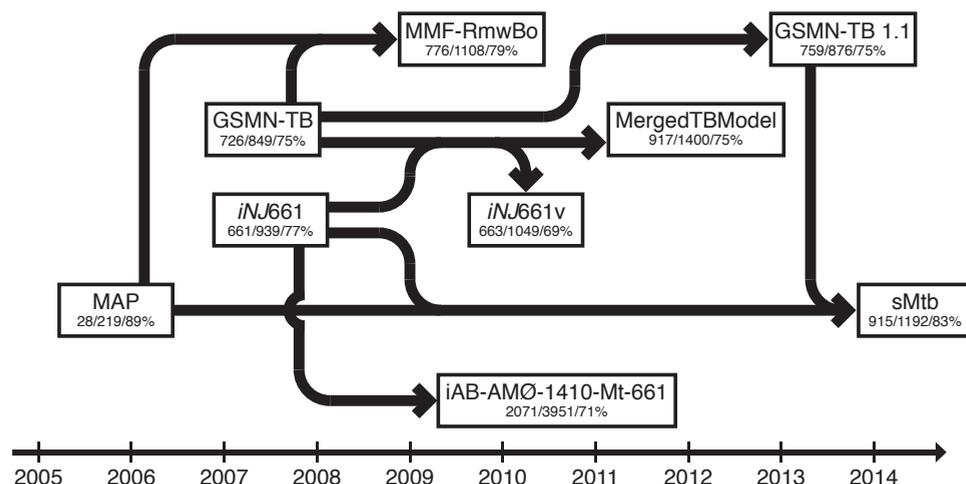

Figure 2.2 - Time line of CB metabolic models of Mtb.
The numbers below every model name denote the number of genes, the number of reactions and the percentage of gene-associated reactions in the model. MAP: Mycolic acid pathway [14], GSMN-TB: Genome-scale metabolic network of *M. tuberculosis* [30], iNJ661: *in silico* Neema Jamshidi 661 genes [23], MMF-RmwBo: Merged McFadden-Ramam with biomass objective [32], iNJ661v: *in vivo* compatible model based on iNJ661 [33], iAB-AMØ-1410Mt-661: *in silico* Aarash Bordbar alveolar macrophage 1410 genes *Mycobacterium tuberculosis* 661 genes [7], MergedTBModel: Merged *Mycobacterium tuberculosis* model [35], GSMN-TB 1.1: A curated and extended version of GSMN-TB [24], sMtb: *in silico Mycobacterium tuberculosis*.

Biomass functions for *in vitro* Mtb

iNJ661 and GSMN-TB 1.1 are reconstructed independently and therefore not only differ in network topology, but also differ in the biomass reactions. The chemical formulas of all biomass precursors in a CB metabolic model, multiplied with their stoichiometric coefficients, should add up to one gram dry weight of biomass. This is the case for the biomass functions of iNJ661 and sMtb and the contribution of each subgroup of metabolites to the total biomass can be calculated (Table 2.1). However, the weight percentage of nucleic acids in iNJ661 seems to be up to five-fold higher than those used in GSMN-TB 1.1. This difference can be attributed to differences in two studies reporting on nucleic acid dry weight percentages [36, 37]. Unfortunately, there are no chemical formulas provided in GSMN-TB 1.1, which complicates the identification of the exact nature of some compounds, such as 'DIM' (dimycocerosate)

or 'PIMS' (phosphatidyl myo-inositol mannosides), and such a classification of the relative contribution of subgroups of metabolites to biomass cannot be obtained. Biomass functions are often used to validate CB metabolic models by comparing predicted specific growth rates with experimentally obtained specific growth rates. iNJ661 was validated in such a way, for growth on three media differing in carbon and nitrogen sources [23]. Similarly GSMN-TB, the predecessor of GSMN-TB 1.1 was validated using experimentally measured specific growth rates for various measured glycerol uptake rates [30].

For sMtb, the biomass composition is based on the average composition measured at two different growth rates [36], although adaptations would be required for those conditions where experimental evidence shows altered compositions. Objective functions for dormant mycobacteria are most likely very different from those of actively replicating mycobacteria.

Table 2.1 - Weight percentages of different biomass components.

Metabolite group	Percentage of iNJ661 biomass (%wt/wt)[a]	Percentage of sMtb biomass (%wt/wt)[a]
Amino acids	27	22
Nucleic acids	26	5
Sugars and Carbohydrates	21	26
Lipids	25	39
Other	1	8

[a] The weight percentage of each subgroup of metabolites is calculated by multiplying the stoichiometric coefficients for each metabolite in each subgroup by their molecular weights and dividing the total in each subgroup by the total weight of all metabolites.

Biomass functions for *in vivo* Mtb

Biomass composition of Mtb is not constant over different conditions. For instance it is known that Mtb accumulates triacylglycerol under *in vitro* conditions that produce a state which mimics the dormant state in the host [38], and that the synthesis of a specific class of iron chelating molecules, called mycobactin siderophores, is required for iron acquisition [39]. These adaptations effectively change the biomass composition. Moreover, *in vivo* Mtb is under constant stress caused by the host immune system, in particular oxidative stress by reactive oxygen and nitrogen species produced by the host [40]. The damaging effects of these reactive species must be compensated, again changing the growth requirements, which should be reflected in the optimization objective(s) when *in vivo* metabolic states are simulated.

A consensus metabolic model of Mtb (sMtb)

The mere existence of eight different genome-scale metabolic models of Mtb, of which most are extensions of previous ones, reflects the importance of keeping CB metabolic models up to date. Two major independently created CB models of Mtb metabolism have thus far not been merged and manually curated. These two models: GSMN-TB 1.1 and iNJ661 differ in size and cover partly overlapping parts of Mtb metabolism. Metabolites are annotated differently for both models. Model iNJ661 contains for the metabolites: abbreviations, full names, chemical formulas and charges, whereas GSMN-TB 1.1 only contains abbreviations and full names. Both models use different abbreviations and few metabolite names appear the same in both models. Neither model contains references to persistent chemical databases, such as ChEBI [41], PubChem [42] or KeGG [43] or database-independent identifiers, such as SMILES [44]. There are large parts of metabolism covered by GSMN-TB 1.1 that are not covered by iNJ661 and *vice versa*. In addition, the mycolic acid synthesis pathway is described in more detail by model MAP than either iNJ661 or GSMN-TB 1.1. Therefore we have constructed sMtb, a manually curated merged model of MAP, iNJ661 and GSMN-TB 1.1 that is currently the most comprehensive genome-scale metabolic model of Mtb. sMtb is provided in the Supplementary Material in SBML formats, level 2 and 3 and as a spreadsheet. Unlike previously published CB metabolic Mtb models, sMtb contains chemical formulas, references to KeGG, PubChem, ChEBI and SMILES for all metabolites. These references permit automated reasoning and allow all reactions to be elementally balanced. The metabolic network of sMtb contains 1192 reactions, 915 genes, and 929 metabolites. It includes a number of important extensions to previous models, such as the mycolic acid synthesis [29], dimycocerosate ester biosynthesis [45] and cholesterol degradation [8] pathways that have been updated according to the latest insights. In sMtb 84% of the reactions are associated with the corresponding genes, whereas in GSMN-TB 1.1 and iNJ661 these percentages are only 75% and 77%, respectively. A high percentage of gene-associated reactions in a CB metabolic model is a signature of a reliable network topology. However, it is not a guarantee, because the gene essentiality predictions of GSMN-TB 1.1 are better than those of iNJ661 (Table 2.2). This does not necessarily mean that the network topology of GSMN-TB 1.1 is better than that of iNJ661, it could also be due to the more accurate biomass objective of GSMN-TB 1.1 that is designed to describe *in vitro* growth.

Table 2.2 - Validation of network topology and biomass function by gene essentiality.
Note that due to rounding, the totals may not add up to 100%.

Model	iNJ661		GSMN-TB 1.1[a]		sMtb	
Objective for *in vitro* growth	no		yes		yes	
True Positives	132	20%	175	23%	215	23%
True Negatives	288	44%	395	52%	522	57%
False Positives	59	9%	45	6%	45	5%
False Negatives	182	28%	144	19%	133	15%
Correct predictions	420	64%	570	75%	737	80%
Sensitivity	42%		55%		62%	
Specificity	83%		90%		92%	
Accuracy	64%		75%		80%	

[a] TP, TN, FP, FN and Correct predictions percentages from [24].

Prediction of gene essentiality

Gene essentiality predictions depend, among other factors, on the available nutrients, the topology of the metabolic network, the quality of the annotation and the chosen objective function. These predictions are suitable to test the topology of a metabolic network, however, they are by no means a quantitative validation of flux distribution predictions. Genes are deleted from the model one at the time and all the reactions that are dependent on the enzyme encoded by the gene are constrained to carry no flux. If the value of the objective function (often maximization of biomass production) is significantly or totally reduced by these constraints, the gene is predicted to be essential. These predictions are thus condition specific and differ for the various models. We have used iNJ661, GSMN-TB 1.1 and sMtb to predict genes that upon *in silico* deletion would result in a decrease of the specific growth rate by 95% or more (see Supplementary Methods). Those genes were said to be essential and compared to an *in vitro* gene essentiality dataset generated via deep sequencing [5]. It can be seen in Table 2.2 that sMtb performs best in predicting *in vitro* gene essentiality, with an accuracy of 80% as compared to 75% for GSMN-TB 1.1 and 64% for iNJ661.

However, as the chosen threshold changes, so do the sensitivity (also called true positive rate) and the false positive rate (1 - specificity). The relationship between the false positive rate and the true positive rate for the gene essentiality predictions by the various models for different threshold values is given in a Receiver Operating Characteristic (ROC) curve (Supplementary Figure 2.2). The corresponding Area Under the Curve (AUC) represents the chance that a randomly chosen experimentally observed essential gene is predicted as such and is

commonly used for model comparison. For iNJ661, GSMN-TB 1.1 and sMtb this chance equals 0.65, 0.78 and 0.80 respectively. In all three cases the p-values (all lower than 10^{-5}) associated with the AUC show that these areas are significantly different from 0.5, which would correspond to a random prediction.

Central carbon metabolic flux predictions compared to ^{13}C data

To validate CB metabolic models, ideally the predicted metabolic states would be compared to measured metabolic states. Although metabolic states cannot directly be measured, they can be inferred by isotopic labeling experiments. Flux distributions obtained from Mtb CB genome-scale metabolic models have thus far not been compared to *in vitro* ^{13}C inferred fluxes as has been done for other organisms, such as *E. coli* [16, 46].

We compared the ability to correctly predict metabolic flux distributions for the three CB metabolic models: iNJ661, GSMN-TB 1.1 and sMtb. *In vitro* results for Mtb and the attenuated TB vaccine strain *Mycobacterium bovis* Bacillus Calmette-Guérin (BCG) were obtained from Beste et al. [47]. BCG has a high degree of genome identity to Mtb and is therefore often used as an Mtb surrogate [48-50]. The three CB metabolic models GSMN-TB 1.1, iNJ661 and sMtb all contain biomass functions that are based on both BCG and Mtb biomass composition. Therefore, metabolic fluxes from both Mtb and BCG are used. Beste and colleagues measured the specific glycerol consumption rate, the specific Tween 80 consumption rate and the specific CO_2 production rate at two different dilution rates: 0.01 h^{-1} and 0.03 h^{-1} for BCG and 0.01 h^{-1} for Mtb [47]. These experiments were done in a chemostat, therefore the dilution rate equals the specific growth rate. Tween 80 is a fatty acid ester of sorbitan polyethoxylate. Mycobacteria have phospholipase A activity that releases fatty acids from Tween [51]. In the case of Tween 80, oleic acid is released. Therefore, the specific consumption rate of Tween 80 can be simulated as the specific consumption rate of oleic acid (for more details see Supplementary Methods).

Table 2.3 - Growth related ATP coefficients and specific growth rate predictions for the various models.

Model	Growth related coefficient (mmol gdw^{-1})	Specific growth rate prediction (h^{-1})		
		0.01 BCG	0.03 BCG	0.01 Mtb
iNJ661	60	0.0137	0.0155	0.0077
GSMN-TB 1.1	47 (+9[a])	0.0037	0.0070	0.0037
sMtb	57	0.0151	0.0260	0.0129

[a] Excluding ATP costs for protein, RNA, and DNA synthesis. The sum of these costs equals approximately 9 mmol gdw^{-1} h^{-1}.

Non-growth associated maintenance is expressed as a conversion of ATP to ADP and quantifies the energy required by Mtb to maintain itself in a given environment. All models gave the best specific growth rate prediction when the non-growth associated maintenance was set to 0 mmol gdw^{-1} h^{-1} (Supplementary Figure 2.1). However, a small amount of energy for maintenance is always required to sustain an organism in its environment, therefore a small arbitrary maintenance flux of 0.1 mmol gdw^{-1} h^{-1} was included in each model before predicting the optimal specific growth rate to compare with the measured values (Table 2.3).

As can be seen in Figure 2.3, predicted fluxes and ^{13}C inferred *in vitro* fluxes in general do not completely agree. The different pathways in central carbon metabolism are separated in Figure 2.4 and the predictions of the different models are given. Metabolic pathway representations of the metabolic state predictions are given in Supplementary Figures 2.3, 2.4, and 2.5. All models predict a low flux through the pentose phosphate pathway, even though ^{13}C inferred fluxes show otherwise for BCG at a specific growth rate of 0.03 h^{-1}, but show completely different behaviors for the tricarboxylic acid cycle and the glyoxylate shunt (Figure 2.4). The discrepancies between ^{13}C inferred fluxes and the flux predictions by the various models show that the predictions of the models become worse as the distance (i.e. the number of reactions) from the glycerol entry point, where glycerol is converted to glycerol-3-phosphate, increases. The predictions for pathways such as the TCA cycle and glyoxylate shunt are worse than those for glycolysis and glycerol uptake, because they are further 'downstream' of the glycerol entry point in the models and thus more options exist for the flux to be rerouted towards alternative parts of the metabolic network that are not shown in the network depicted in Figure 2.3. Model sMtb does relatively well at flux predictions for glycolysis and the TCA cycle. In contrast to iNJ661 and GSMN-TB 1.1, it is the only model that predicts a flux from pyruvate to acetyl-CoA for BCG at a specific growth rate of 0.03 h^{-1} and Mtb at a specific growth rate of 0.01 h^{-1}. The standard deviations for most predicted fluxes are relatively small (given by error bars in Figure 2.4), implying that the predictions are precise but not accurate. This could be partly due to the applied sampling method to determine means and standard deviations (see Supplementary Methods), but it could also be caused by a bimodal distribution of flux solutions instead of a normal distribution, which would limit the usefulness of concepts such as means and standard deviations. Another point to consider regarding flux predictions is that although the flux predictions of all three models can be improved, ^{13}C fluxes are also inferred from a model, using measured metabolites, which makes it more complicated to point out whether the predicted fluxes, inferred fluxes, or both can be improved.
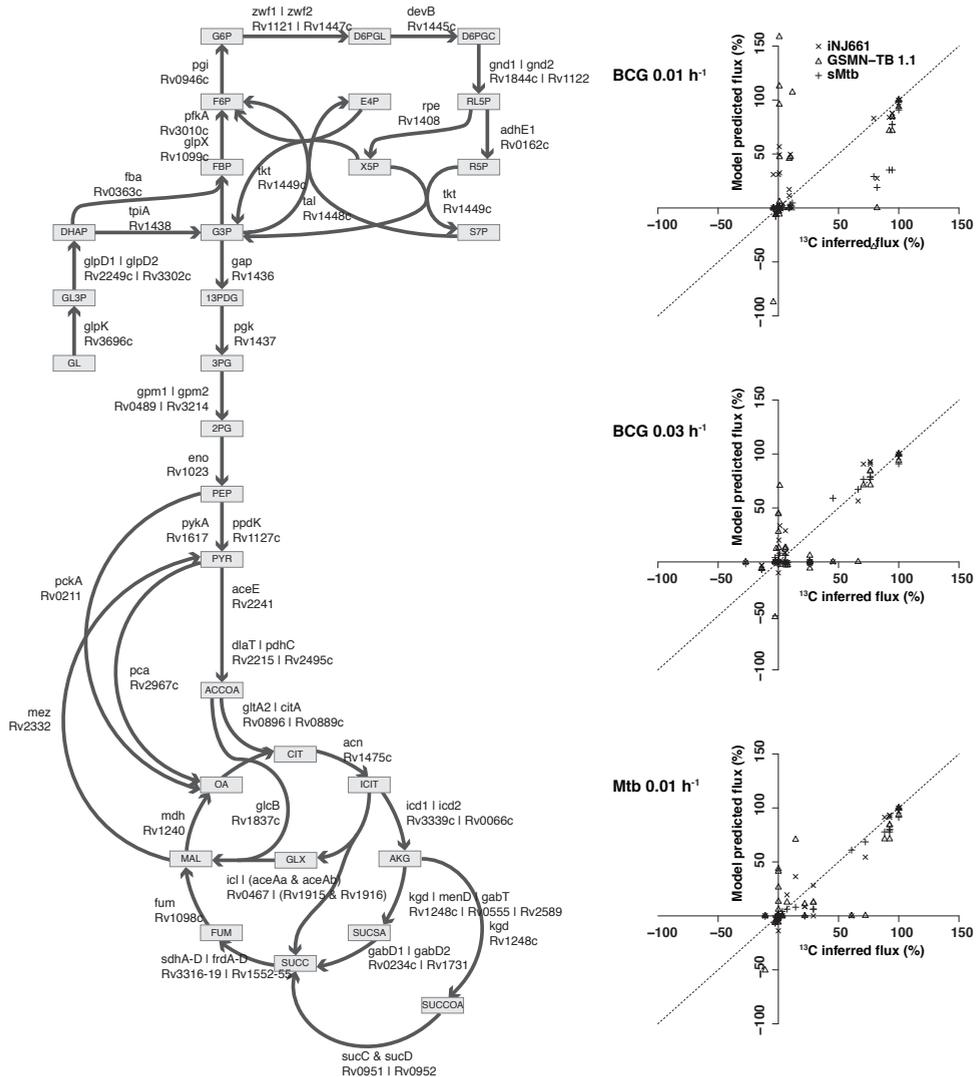
Figure 2.3 - Central carbon metabolism and agreement between ^{13}C inferred and predicted flux.
Central carbon metabolism of BCG and Mtb is given on the left. For each reaction a gene name and a locus tag is given corresponding to the gene(s) encoding the enzyme(s) catalyzing the reaction. Isozymes are indicated separated by a '|' while subunits are separated by a '&' symbol. The graphs on the right indicate the agreement between ^{13}C inferred [47] fluxes and predicted fluxes by iNJ661 (crosses), GSMN-TB 1.1 (triangles) and sMtb (plusses). The fluxes are given as a percentage of the glycerol uptake rate. Negative percentages denote a reversed flux direction. The black dashed line represents perfect agreement.

Metabolite abbreviations: GL, glycerol; G6P, D-glucose 6-phosphate; F6P, D-frucose 6-phosphate; FBP, D-fructose 1,6-bisphosphate; G3P, D-glyceraldehyde 3-phosphate; 13PDG, 3-phospho-D-glyceroyl phosphate; 3PG, 3-phospho-D-glycerate; 2PG, 2-phospho-D-glycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; D6PGL, D-glucono-1,5-lactone 6-phosphate; D6PGC, 6-phospho-D-gluconate; RL5P, D-ribulose 5-phosphate; X5P, D-xylulose 5-phosphate; R5P, D-ribose 5-phosphate; S7P, sedoheptulose 7-phosphate; E4P, D-erythrose 4-phosphate; ACCOA, acetyl-CoA; ICIT, isocitrate; AKG, 2-oxoglutarate; SUCSA, succinic semialdehyde; SUCCOA, succinyl-CoA; FUM, fumarate; MAL, malate; OA, oxaloacetate; GLX, glyoxylate.
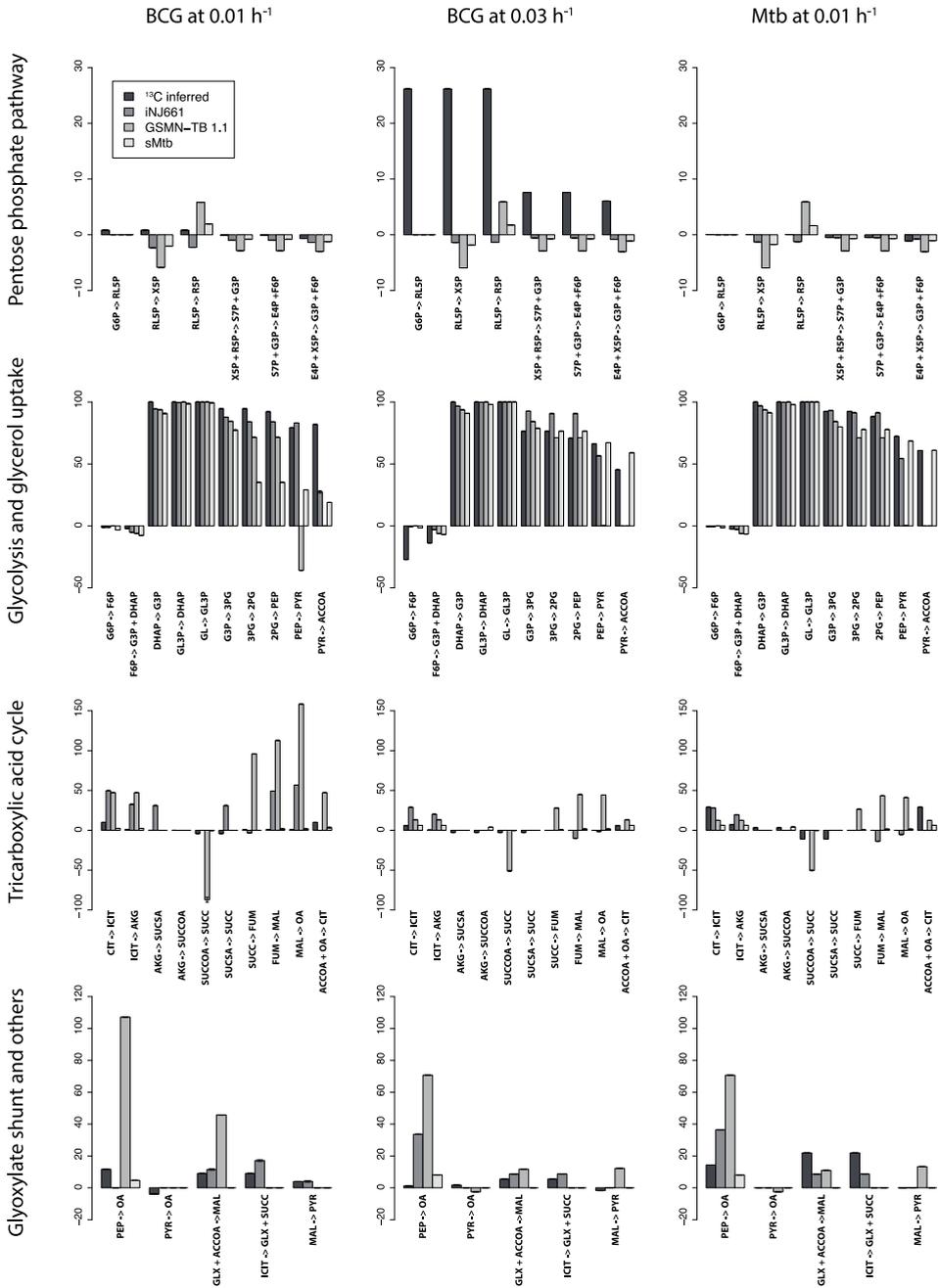
Figure 2.4 - ^{13}C inferred fluxes and predicted fluxes for various parts of central carbon metabolism. ^{13}C Inferred [47] (black) and predicted fluxes for iNJ661 (dark grey), GSMN-TB 1.1 (grey) and sMtb (light grey), given as a percentage of the glycerol uptake rate, for the various parts of central carbon metabolism for BCG grown at 0.01 h^{-1} and 0.03 h^{-1} and Mtb grown at 0.01 h^{-1}. The error bars indicate the standard deviations.

Nevertheless, sMtb shows the highest agreement between inferred and predicted fluxes, closely followed by iNJ661 (Table 2.4). The more accurately reflected cellular behavior under *in vitro* conditions by sMtb as compared to iNJ661 and GSMN-TB 1.1 increases the confidence of predictions of cellular behavior under *in vivo* conditions by sMtb. Therefore, sMtb provides a more accurate platform for drug target discovery than was available before.

Table 2.4 - Pearson's correlation coefficient for inferred and predicted fluxes for BCG and Mtb at various growth rates.

Model	Pearson's correlation coefficient			
	BCG $\mu = 0.01$ h^{-1}	BCG $\mu = 0.03$ h^{-1}	Mtb $\mu = 0.01$ h^{-1}	Average
iNJ661	0.87	0.90	0.94	0.90
GSMN-TB 1.1	0.42	0.80	0.90	0.66
sMtb	0.90	0.95	0.98	0.94

Drug-phenotype predictions

We tested the three models on their ability to assess the effectiveness of anti-TB drugs with known metabolic targets. Table 2.5 provides an overview of the predicted phenotypes after drug application by inactivating the specific enzyme and the corresponding reaction(s) *in silico* (see Supplementary Methods). sMtb predicts the highest number of non-viable phenotypes caused by anti-TB drugs, closely followed by iNJ661 and GSMN-TB 1.1. Nevertheless, these predictions are based on growth on Roison's minimal medium [47], which does not represent *in vivo* conditions. Moreover, *in vitro* biomass reactions are used for both GSMN-TB 1.1 and sMtb. Setting the models such that they simulate *in vivo* conditions would alter these drug-phenotype predictions. However, this is complicated due to the fact that iNJ661 does not contain a cholesterol degradation pathway, which has been shown to be important for intracellular growth and survival [8, 52-56]. Mtb infection is a complex interplay between the pathogen and its host that involves cellular changes in both organisms [57]. Therefore, modeling both host and pathogen metabolism simultaneously is required for an accurate representation of infection.

Table 2.5 - Drugs with known metabolic targets [6, 58-60] and the percentage of the specific growth rates obtained after *in silico* gene knockouts of these targets.

Target	Drug	Percentage of the wild type specific growth rate obtained		
		iNJ661	GSMN-TB 1.1	sMtb
InhA	Isoniazid, ethionamide	0 %	100 %	0 %
KasA	Isoniazid	0 %	100 %	0 %
DfrA	Isoniazid	0 %	100 %	100 %
EmbB	Ethambutol	0 %	0 %	0 %
Alr	Cycloserine	0 %	0 %	0 %
DdlA	Cycloserine	0 %	0 %	0 %
FolP1	Para-amino salicylic acid	100 %	100 %	100 %
AtpE	TMC207	63 %	57 %	60 %
DprE1	BTZ043	100 %	0 %	0 %
KasB	Thiolactomycin	0 %	100 %	0 %
FabH	Thiolactomycin	100 %	0 %	100 %
MmaA4	Thiacetazone	100 %	0 %	0 %
Total percentage of non-viable phenotypes[a]		58%	50%	67%

[a] If the predicted specific growth rate of an *in silico* knockout mutant equals 5% or less of the *in silico* wild-type specific growth rate prediction, the knockout mutant is classified as a non-viable phenotype.

While CB metabolic models unfortunately cannot directly predict which molecules are effective drugs, they can predict which metabolic enzymes make for suitable drug targets. Whether or not such enzymes can be effectively inhibited depends on the characteristics of the enzyme itself. Databases such as TuberQ [61] can provide a druggability analysis for an enzyme predicted to be a suitable drug target, thereby verifying if the enzyme can effectively be targeted. An approach to select suitable drug targets will be more effective if essentiality analysis is combined with additional systems level information such as information on the accumulation of stable toxic intermediates. For examples, the cholesterol degradation pathway in Mtb [8] contains a large number of enzymes, many of them essential for cholesterol degradation and thus possible drug targets. However, accumulation of stable toxic intermediates such as cholest-4-en-3-one and catechol derivatives can occur if the enzymes HsaC, KshA, Cyp125 and Cyp142 are non-functional [62, 63]. The accumulation of such intermediates can be fatal to Mtb, increasing the potential of these enzymes as drug targets. A similar approach can be taken by designing

replacement substrates for enzymes that serve as inhibitors of subsequent enzymes after being activated by the target enzyme [64].

Perhaps one of the biggest advantages of using CB metabolic models to find drug targets is that it enables the prediction of metabolic rearrangements after constraining the flux through reactions that are known to be affected by a given drug. This can highlight the possible 'escape routes' that Mtb possesses. Bhat and colleagues [12] used such an approach which is further discussed in the part: Discovering new drug targets and combinations of drugs.

sMtb overall performance

Model iNJ661 predicts metabolic states relatively well as compared to GSMN-TB 1.1 (Table 2.4, Figures 2.3 and 2.4), but on the other hand the gene essentiality predictions of GSMN-TB 1.1 are better (accuracy of 75%) than those of iNJ661 (accuracy of 64%). The consensus genome-scale CB metabolic model sMtb is the most comprehensive, manually curated genome-scale CB model of Mtb to date. It represents the strengths of iNJ661 and GSMN-TB 1.1 and not only gives accurate qualitative predictions, such as gene essentiality predictions (Table 2.2) and drug-phenotype predictions (Table 2.5), but also accurate quantitative predictions, such as the specific growth rate (Table 2.3) and the metabolic states (Figure 2.4; Supplementary Figures 2.3, 2.4, and 2.5). The overall improved performance of sMtb is essential for obtaining meaningful and accurate predictions of the metabolic state in conditions that are experimentally inaccessible. Moreover, the improved annotation of sMtb regarding its metabolites is a critical point, as it enables future refinements and extensions by other researchers with relative ease.

However, even though sMtb performs better in overall predictions of *in vitro* metabolic states, there is room for improvement, especially regarding the metabolic state predictions of the pentose phosphate pathway and the glyoxylate shunt. Options to achieve these better predictions would be to supply a more accurate objective, or to improve the underlying metabolic network of sMtb.

Understanding Mtb metabolism and designing intervention strategies: challenges and outlook

In an attempt to mimic metabolic states of Mtb in various environments more accurately, CB metabolic models can be constrained with various types of -omics data. Unlike flux measurements, gene expression data can be relatively straightforwardly obtained using RNA sequencing or microarray technologies. CB metabolic models can also act as scaffolds for other types of -omics data, such as

proteomics. These data types have the added advantage of being (almost) genome-scale and can be integrated into CB metabolic models, creating condition-specific models with increased predictive power. Such condition-specific models are important to provide reliable metabolic state predictions in *in vivo* conditions where uptake rates and metabolic objectives are unclear, with the ultimate goal of designing novel intervention strategies.

Integration of expression data

Alternative methods have been developed to integrate either gene or protein expression data into CB models, see [20, 65-67] for recent reviews. A systematic evaluation of these methods, comparing performance and robustness using alternative models and data sets [68] shows that no method outperforms the others in all the tested scenarios. Here, we will focus on the methods that have been applied to explore mycobacterial metabolism.

E-Flux [32] constrains the maximum flux through a reaction using the measured gene expression levels. Whenever the expression level of an enzyme-coding gene is low, tight constraints are imposed on the maximal flux through the corresponding reaction. The rationale is that mRNA levels can be used as an approximation to the amounts of protein available, and these in turn can be used as an approximation to the upper bound on reactions rates. This algorithm was tested using two models, MAP and MMF-RmwBo. The Boshoff Mtb gene expression compendium [69] contains over 400 microarray experiments measuring the transcriptional adaptations of Mtb to 75 different drugs, drug combinations and growth conditions. E-Flux was used to predict the impact of each of these conditions and drugs on the biosynthesis of mycolic acids. This approach correctly predicted the specificity of seven of the eight known inhibitors of mycolic acid biosynthesis included in the data compendium. Additionally, it was also able to identify a small number of non-specific potential inhibitors and enhancers of mycolic acid biosynthesis.

While E-Flux uses transcript data to improve the predictions of metabolic fluxes, Fang and colleagues [70] proposed an *in silico* approach to create state-specific models by integrating gene expression data. Their method relies on comparing gene expression levels between a metabolically well-characterized reference state and the perturbed state of interest. This method uses the flux distribution in the reference state and imposes soft constraints on the fluxes according to the observed changes in gene expression to characterize the perturbed metabolic state. Changes in gene expression data for wild type Mtb H37Rv, as well as for the ΔdosR deletion mutant, associated with the transfer from normoxic to hypoxic conditions were combined with iNJ661v to produce condition specific models for both strains. These models correctly predicted the essentiality of dosR for the adaptation to hypoxia. Additionally, the model also predicted the altered biomass composition of Mtb in

hypoxic conditions (linked to the increased production of cell-wall metabolites) and the critical contribution of the reductive side of the tricarboxylic acid cycle to the adaptation to low oxygen environments. The condition-specific models can also serve to specifically identify drug targets for the latent stages of the disease. The algorithms described so far provide as primary output models of metabolism with altered constraints that can be used to further characterize the metabolic responses. Differential Producibility Analysis (DPA) [71] on the other hand, aims at extracting metabolic signals from expression data. DPA uses the model to identify genes affecting the production of each metabolite in the network, then expression data is used to obtain and average expression values of each set of metabolite associated genes. These values are then used to identify the metabolites associated with increased and decreased gene expression. DPA was used to analyze the metabolic state of Mtb *in vivo* (with expression data obtained from sputum samples of TB patients and from pathogens replicating in mouse macrophages) [72, 73] and in various *in vitro* conditions (such as growth on different carbon sources or exposure to different stress sources) [69, 74]. The analysis showed that one of the main adaptations to the macrophage environment is the downregulation of genes influencing metabolites in central metabolism, and the simultaneous upregulation of genes linked to cell wall synthesis

Integration of regulatory information
Probabilistic regulation of metabolism (PROM) [19] is an algorithm that attempts to link regulatory and metabolic networks. The transcriptional regulatory network of Mtb [75] and the Boshoff Mtb compendium [69] were used to build a probabilistic model of gene regulation. The probabilities were then integrated into the iNJ661 model as constraints on reactions of which the flux could vary according to the state of the transcription factor regulating the expression of the enzyme-coding gene. PROM correctly predicted the phenotype of 23 out of the 24 studied transcription factor knock out mutants. The increased knowledge on the regulatory networks in Mtb [76] opens new ways to consider not only genes primarily related to metabolism but also to their regulators, thereby increasing the potential to discover new drug targets.

Growth related ATP coefficients and non-growth associated maintenance
The biomass reaction describes the assembly of biomass precursors into new cells. Each biomass precursor has a defined coefficient denoting the amount (in mmol) required to form one gram dry weight of biomass. The assimilation of these precursors requires energy, in the form of ATP to ADP conversion that is introduced through a growth related ATP coefficient in the biomass function (also called growth associated maintenance). This coefficient is very similar for iNJ661, GSMN-TB 1.1

and sMtb (Table 2.3). The growth related ATP coefficient of iNJ661 equals 60 mmol gdw^{-1} and that of GSMN-TB 1.1 equals 47 mmol gdw^{-1} plus an additional 8.8 mmol gdw^{-1} associated with protein formation. Both models thus have a similar value for growth-associated maintenance. Unlike the growth related ATP coefficient, non-growth associated maintenance is independent of the biomass composition. Instead, it depends on the environment and on the metabolic pathways utilized for growth [77]. It is assumed that non-growth associated maintenance, in the form of ATP to ADP conversion, is a fixed value independent of the specific growth rate. Here, we have set the non-growth associated maintenance to a small value so that the three models give the best predictions of the specific growth rate (see Supplementary Figure 2.1)

Non-growth associated maintenance is a useful parameter when trying to simulate *in vivo*, e.g. phagosomal, growth. The phagosome is a hostile environment and the energy required for non-growth associated maintenance will be relatively high, compared to *in vitro* growth conditions. Moreover, the specific growth rate will be limited in the phagosome. A high non-growth associated maintenance requirement and a low specific growth rate cannot be simulated effectively using a model that contains a regular biomass reaction, which includes a growth related ATP coefficient, but no non-growth associated maintenance cost.

Objective and constraints for *Mycobacterium tuberculosis* in the host

When using CB genome-scale metabolic models of Mtb as opposed to non-pathogenic microorganisms grown in an *in vitro* condition, it is not straightforward to select an optimization objective. The primary objective of the pathogen might be focused on survival instead of growth. In addition, the host-pathogen interaction is a complex and time-dependent dynamic process, where they mutually influence each other. Hence, CB metabolic models, which rely on the steady state assumption, might not be realistic for many pathogens. Mtb is known for its ability to remain dormant in the host for years. In those cases, the host's immune system prevents the pathogen from spreading and Mtb is contained within solid granulomas [2]. It is estimated that 2 billion people worldwide are latently infected [1]. The relative metabolic activity at the latent infection stage however, is very low. There is thus a stark need to understand the mechanisms underlying dormancy and predict its dynamics and the switch to active state. Modeling accurately and realistically this infection stage is hence of utmost importance. A key factor determining the accuracy of CB metabolic models in an infection setting is the identification of a suitable objective function representing dormant Mtb. Shi and colleagues created an objective function representing non-growing cells, based on the minimal cell wall composition deduced from gene expression data [78]. They compared predicted flux changes between growing and non-growing cells with qPCR data

and found consistency between fluxes and gene expression for critical pathways of central metabolism. A limitation of this approach is that the metabolic model is based on transcript abundance data [78]. A leap forward would be to investigate the biomass composition of Mtb in an *in vivo* or *ex vivo* situation.

Knowledge (or the lack thereof) of the availability of nutrients in the host environment is another factor that determines the quality of the model predictions. Bordbar and colleagues constructed a macrophage-Mtb model, iAB-AMØ-1410-Mt-661, they estimated that the carbon sources available in the phagosome were glycerol and even long chain fatty acids (myristic acid, palmitic acid and stearic acid) [7]. Recent insights have changed this picture and highlighted the importance of cholesterol [8], aspartate [79], and other nutrients [9] in the phagosomal environment. Knowing the precise composition and availability of such nutrients will enable much more accurate predictions of the *in vivo* metabolic state and of the *in vivo* essentiality of gene products.

Annotation of combinatorial proteins

Little is known about transport proteins of Mtb despite the abundance of genomic data [80]. Transport proteins are at the boundaries of the metabolic networks and therefore function as gatekeepers for fluxes. Not only is it important to know which compounds Mtb can take up, but it is also important to know whether these transporters are channels, symporters or antiporters. In addition, quantitative predictions also require the knowledge of the energy requirements of the transporters. A better annotation of transport proteins of Mtb is therefore required.

Cofactor limitation

Beste and colleagues mimicked the cofactor requirements of the enzymes by forcing the reactions catalyzed by these enzymes to use a small arbitrary amount of cofactor [30]. Quantitative predictions are most likely not accurate due to the arbitrarily chosen amount of cofactor used in any reaction. Nevertheless, such an approach could be extended to simulate cofactor limitations. Iron availability is assumed to be reduced in the phagosome [72], thus introducing ways to mimic this iron scarcity in the CB models, will lead to more accurate descriptions of the bacterial metabolism during the infection process.

Discovering new drug targets and combinations of drugs

Fang and colleagues integrated a dynamic cell population growth model and an enzyme inhibition model with a modified version of iNJ661 [81]. The integrated model was able to reproduce *in vitro* experimentally measured dose-response curves of 3-nitropropionate, an inhibitor of the glyoxylate shunt and the methylcitrate cycle.

Simulating single or double gene knock out mutants to discover potential drug targets and synergistic combinations, greatly depends on the network topology, the objective function, and the substrate(s) available to the bacteria. The difference in specific growth rate predictions between the wild-type and simulated single or double knock out mutants, is mainly attributable to the rates at which substrates are taken up and metabolites are secreted, and not to the compounds available. Synergistic combinations of drug targets can also be found by gradually decreasing flux through the first potential target, which can be found for example, through a classic gene essentiality approach, and afterwards identifying those parts of the metabolism that are forced to carry a relatively higher flux. Bhat and colleagues applied a similar strategy and studied the effect of varying inhibition by isoniazid, a front-line drug, on the metabolic state [12]. By gradually limiting the flux through the target of isoniazid, InhA, they found that the flux through various pathways was induced compared to the unperturbed state. These pathways could then potentially be analyzed to identify suitable targets for drugs administered in combination with isoniazid.

These examples show the potential of using CB models to systematically probe the metabolic space of Mtb, generate novel insights and pin-point possible targets for interventions, with drugs or otherwise.

Combinatorial models and host drug targets

The integrated human alveolar macrophage-Mtb model iAB-AMØ-1410-Mt-661 combines the Mtb metabolic model iNJ661 and the first reconstruction of human metabolism, RECON 1. Recently, the human model was updated to the consensus reconstruction, RECON 2 [82], which in turn can be combined with sMtb to create an updated macrophage-Mtb model. It is crucial for such a model to contain an accurate description of the phagosomal environment and its contents, as this provides the framework for the host-pathogen interaction and can have a large impact on the predictions of the metabolic state for both organisms. Although drug target discovery is generally focused on the pathogen, there are also opportunities to look at the host metabolism for drug targets. An example of a host-targeted drug is thioridazine, which is postulated to inhibit efflux of potassium and calcium from the phagolysosome required for its acidification [83]. The phagosomal environment steers the pathogen metabolism, thus drugs targeting primarily the host and altering this environment will result in metabolic changes in Mtb as well. This could result in a state that renders the bacteria more susceptible to subsequent anti-TB drugs. A combined model could provide additional host drug targets, however a thorough understanding of the functioning and composition of the phagosome is required. An experimentally validated and accurate macrophage-Mtb model has much potential for drug target discovery, especially for the identification of synergistic drug targets, both in the host and Mtb itself or a combination of both.

Conclusions

The quality and predictive power of genome-scale reconstructions of the metabolism and transport of Mtb is gradually increasing. Our current model, sMtb, outperforms considerably previously published models in *in vitro* metabolic state predictions (Table 2.4, Figure 2.3 and 2.4) and specific growth rate predictions (Table 2.3) as well as *in vitro* gene essentiality predictions (Table 2.2) and drug-phenotype predictions (Table 2.5). However, there is still ample room for improvement. The predictions of flux through the pentose phosphate pathway can be improved for all models, while flux through the glyoxylate shunt is still best predicted by iNJ661. Better metabolic state predictions can be obtained through an improved network topology, by improving the determination of the biomass composition under different conditions, and by defining more accurately the objective function, as Schuetz and colleagues did [16] for a small *E. coli* model. Different combinations of the growth related ATP coefficient and the non-growth associated maintenance also have an impact on the metabolic state predictions, but these are hard to measure and their values can vary even for well-known organisms [84, 85]. Nevertheless they can be valuable parameters to fit CB metabolic models to ^{13}C data, thereby improving their predictive power.

A CB metabolic model with sufficient *in vitro* predictive power forms the foundation for reliable *in vivo* metabolic state predictions. Nevertheless, the *in vivo* metabolic state of Mtb is arguably not in steady state and relatively little is known about the 'objective' of Mtb in the host. Efforts on both the experimental and modeling side of Mtb metabolism continuous to shed light on its *in vivo* metabolic state(s) and paves the way for the discovery of new (synergistic) drug targets and possible new intervention strategies. The long term vision is that such a metabolic model will be one of the modules of a larger multi-scale modeling framework that connects a variety of models at different scales, each describing a particular subset of the behavior of Mtb in infection settings. This will thus ultimately contribute to the grander vision of a model-based 'Virtual Patient', with enormous potential to Health and Medicine.

Availability of supporting data

Supplementary Files can be accessed via: Rienk A. Rienksma, Maria Suárez-Diez, Lucie Spina, Peter J. Schaap, Vitor A.P. Martins dos Santos. Systems-level modeling of mycobacterial metabolism for the identification of new (multi-)drug targets. Seminars in Immunology 26 (2014).

Acknowledgements

2

References

1. WHO, *Global Tuberculosis Report 2013*. 2013, France: World Health Organization.
2. Gengenbacher, M. and S.H.E. Kaufmann, Mycobacterium tuberculosis: *success through dormancy*. FEMS Microbiology Reviews, 2012. **36**(3): p. 514-532.
3. Ma, Z., C. Lienhardt, H. McIlleron, A.J. Nunn, and X. Wang, *Global tuberculosis drug development pipeline: the need and the reality*. The Lancet, 2010. **375**(9731): p. 2100-2109.
4. Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Genes required for mycobacterial growth defined by high density mutagenesis*. Molecular Microbiology, 2003. **48**(1): p. 77-84.
5. Griffin, J.E., J.D. Gawronski, M.A. Dejesus, T.R. Ioerger, B.J. Akerley, and C.M. Sassetti, *High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism*. PLoS Pathog, 2011. **7**(9): e1002251.
6. Sacchettini, J.C., E.J. Rubin, and J.S. Freundlich, *Drugs versus bugs: in pursuit of the persistent predator* Mycobacterium tuberculosis. Nat Rev Micro, 2008. **6**(1): p. 41-52.
7. Bordbar, A., N.E. Lewis, J. Schellenberger, B.O. Palsson, and N. Jamshidi, *Insight into human alveolar macrophage and* M. tuberculosis *interactions via metabolic reconstructions*. Mol Syst Biol, 2010. **6**: p. 422.
8. Wipperman, M.F., N.S. Sampson, and S.T. Thomas, *Pathogen roid rage: Cholesterol utilization by* Mycobacterium tuberculosis. Critical Reviews in Biochemistry and Molecular Biology, 2014: p. 1-25.
9. Beste, Dany J.V., K. Nöh, S. Niedenführ, Tom A. Mendum, Nathaniel D. Hawkins, Jane L. Ward, Michael H. Beale, W. Wiechert, and J. McFadden, *[13]C-Flux Spectral Analysis of Host-Pathogen Metabolism Reveals a Mixed Diet for Intracellular* Mycobacterium tuberculosis. Chemistry & Biology, 2013. **20**(8): p. 1012-1021.
10. Gouzy, A., G. Larrouy-Maumus, T.-D. Wu, A. Peixoto, F. Levillain, G. Lugo-Villarino, J.-L. Gerquin-Kern, L.P.S. de Carvalho, Y. Poquet, and O. Neyrolles, Mycobacterium tuberculosis *nitrogen assimilation and host colonization require aspartate*. Nat Chem Biol, 2013. **9**(11): p. 674-676.
11. Gouzy, A., G. Larrouy-Maumus, D. Bottai, F. Levillain, A. Dumas, J.B. Wallach, I. Caire-Brandli, C. de Chastellier, T.D. Wu, R. Poincloux, R. Brosch, J.L. Guerquin-Kern, D. Schnappinger, L.P. Sorio de Carvalho, Y. Poquet, and O. Neyrolles, Mycobacterium tuberculosis *exploits asparagine to assimilate nitrogen and resist acid stress during infection*. PLoS Pathog, 2014. **10**(2): p. e1003928.
12. Bhat, A., R. Vashisht, and N. Chandra, *Modeling metabolic adjustment in* Mycobacterium tuberculosis *upon treatment with isoniazid*. Systems and Synthetic Biology, 2010. **4**(4): p. 299-309.
13. Haggart, C.R., J.A. Bartell, J.J. Saucerman, and J.A. Papin, *Whole-genome metabolic network reconstruction and constraint-based modeling*. Methods in enzymology, 2011. **500**: p. 411-433.

2

14. Raman, K., P. Rajagopalan, and N. Chandra, *Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs.* PLoS Comput Biol, 2005. **1**(5): p. e46.

15. Kim, H.U., S.Y. Kim, H. Jeong, T.Y. Kim, J.J. Kim, H.E. Choy, K.Y. Yi, J.H. Rhee, and S.Y. Lee, *Integrative genome-scale metabolic analysis of* Vibrio vulnificus *for drug targeting and discovery.* Molecular Systems Biology, 2011. **7**(1).

16. Schuetz, R., L. Kuepfer, and U. Sauer, *Systematic evaluation of objective functions for predicting intracellular fluxes in* Escherichia coli. Mol Syst Biol, 2007. **3**: p. 119.

17. Bird, B.R., W.E. Stewart, and E.N. Lightfoot, *Transport Phenomena.* Revised 2nd Edition ed. 2007, United States of America: John Wiley & Sons.

18. Lewis, N.E., H. Nagarajan, and B.O. Palsson, *Constraining the metabolic genotype–phenotype relationship using a phylogeny of* in silico *methods.* Nat Rev Micro, 2012. **10**(4): p. 291-305.

19. Chandrasekaran, S. and N.D. Price, *Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in* Escherichia coli *and* Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences, 2010. **107**(41): p. 17845-17850.

20. Bordbar, A., J.M. Monk, Z.A. King, and B.O. Palsson, *Constraint-based models predict metabolic and associated cellular functions.* Nat Rev Genet, 2014. **15**(2): p. 107-20.

21. Schellenberger, J. and B.O. Palsson, *Use of randomized sampling for analysis of metabolic networks.* J Biol Chem, 2009. **284**(9): p. 5457-61.

22. Bordel, S., R. Agren, and J. Nielsen, *Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes.* PLoS Comput Biol, 2010. **6**(7): p. e1000859.

23. Jamshidi, N. and B.O. Palsson, *Investigating the metabolic capabilities of* Mycobacterium tuberculosis *H37Rv using the* in silico *strain iNJ661 and proposing alternative drug targets.* BMC Syst Biol, 2007. **1**: p. 26.

24. Lofthouse, E.K., P.R. Wheeler, D.J.V. Beste, B.L. Khatri, H. Wu, Tom A. Mendum, A.M. Kierzek, and J. McFadden, *Systems-based approaches to probing metabolic variation within the* Mycobacterium tuberculosis *complex.* PLoS ONE, 2013. **8**(9): p. e75913.

25. Phong, W.Y., W. Lin, S.P.S. Rao, T. Dick, S. Alonso, and K. Pethe, *Characterization of phosphofructokinase activity in* Mycobacterium tuberculosis *reveals that a functional glycolytic carbon flow is necessary to limit the accumulation of toxic metabolic intermediates under hypoxia.* PLoS ONE, 2013. **8**(2): p. e56037.

26. Oberhardt, M.A., J. Puchałka, V.A.P. Martins dos Santos, and J.A. Papin, *Reconciliation of Genome-Scale Metabolic Reconstructions for Comparative Systems Analysis.* PLoS Comput Biol, 2011. **7**(3): p. e1001116.

27. Becker, S.A., A.M. Feist, M.L. Mo, G. Hannum, B.O. Palsson, and M.J. Herrgard, *Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox.* Nat. Protocols, 2007. **2**(3): p. 727-738.

28. Fell, D.A. and J.R. Small, *Fat synthesis in adipose tissue - An examination of stoichiometric constraints.* Biochemical Journal, 1986. **238**(3): p. 781-786.

29. Marrakchi, H., M.-A. Lanéelle, and M. Daffé, *Mycolic Acids: Structures, Biosynthesis, and Beyond.* Chemistry & Biology, 2014. **21**(1): p. 67-85.

30. Beste, D.J., T. Hooper, G. Stewart, B. Bonde, C. Avignone-Rossa, M.E. Bushell, P. Wheeler, S. Klamt, A.M. Kierzek, and J. McFadden, *GSMN-TB: a web-based genome-scale network model of* Mycobacterium tuberculosis *metabolism.* Genome Biol, 2007. **8**(5): p. R89.

31. Griffin, J.E., A.K. Pandey, S.A. Gilmore, V. Mizrahi, J.D. McKinney, C.R. Bertozzi, and C.M. Sassetti, *Cholesterol catabolism by* Mycobacterium tuberculosis *requires transcriptional and metabolic adaptations.* Chem Biol, 2012. **19**(2): p. 218-27.

32. Colijn, C., A. Brandes, J. Zucker, D.S. Lun, B. Weiner, M.R. Farhat, T.Y. Cheng, D.B. Moody, M. Murray, and J.E. Galagan, *Interpreting expression data with metabolic flux models: predicting* Mycobacterium tuberculosis *mycolic acid production.* PLoS Comput Biol, 2009. **5**(8): p. e1000489.

33. Fang, X., A. Wallqvist, and J. Reifman, *Development and analysis of an* in vivo-*compatible metabolic network of* Mycobacterium tuberculosis. BMC Syst Biol, 2010. **4**: p. 160.

34. Duarte, N.C., S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, and B.Ø. Palsson, *Global reconstruction of the human metabolic network based on genomic and bibliomic data.* Proceedings of the National Academy of Sciences, 2007. **104**(6): p. 1777-1782.

35. Chindelevitch, L., S. Stanley, D. Hung, A. Regev, and B. Berger, *MetaMerge: scaling up genome-scale metabolic reconstructions with application to* Mycobacterium tuberculosis. Genome Biol, 2012. **13**(1): p. r6.

36. Beste, D.J., J. Peters, T. Hooper, C. Avignone-Rossa, M.E. Bushell, and J. McFadden, *Compiling a molecular inventory for* Mycobacterium bovis *BCG at two growth rates: Evidence for growth rate-mediated regulation of ribosome biosynthesis and lipid metabolism.* J Bacteriol, 2005. **187**(5): p. 1677-84.

37. Youmans, A.S. and G.P. Youmans, *Ribonucleic acid, deoxyribonucleic acid, and protein content of cells of different ages of* Mycobacterium tuberculosis *and the relationship to immunogenicity.* Journal of Bacteriology, 1968. **95**(2): p. 272-279.

38. Daniel, J., H. Maamar, C. Deb, T.D. Sirakova, and P.E. Kolattukudy, Mycobacterium tuberculosis *uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy-like phenotype in lipid-loaded macrophages.* PLoS Pathog, 2011. **7**(6): p. e1002093.

39. McMahon, M.D., J.S. Rush, and M.G. Thomas, *Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in* Mycobacterium tuberculosis. J Bacteriol, 2012. **194**(11): p. 2809-18.

40. Voskuil, M.I., I. Bartek, K. Visconti, and G.K. Schoolnik, *The response of* Mycobacterium tuberculosis *to reactive oxygen and nitrogen species.* Frontiers in Microbiology, 2011. **2**.

41. Brooksbank, C., G. Cameron, and J. Thornton, *The European Bioinformatics Institute's data resources: towards systems biology.* Nucleic Acids Research, 2005. **33**(suppl 1): p. D46-D53.

42.	Acland A, A.R., Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bryant SH, Canese K, Church DM, Clark K, DiCuccio M, Dondoshansky I, Federhen S, Feolo M, Geer LY, Gorelenkov V, Hoeppner M, Johnson M, Kelly C, Khotomlianski V, Kimchi A, Kimelman M, Kitts P, Krasnov S, Kuznetsov A, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Karsch-Mizrachi I, Murphy T, Ostell J, O'Sullivan C, Panchenko A, Phan L, Pruitt DP, Rubinstein W, Sayers EW, Schneider V, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Siyan K, Slotta D, Soboleva A, Starchenko G, Tatusova TA, Trawick B, Vakatov D, Wang Y, Ward M, Wilbur W, Yaschenko E, Zbicz K, *Database resources of the National Center for Biotechnology Information.* Nucleic Acids Research, 2013. **41**(D1): p. D8-D20.

43.	Kanehisa, M., S. Goto, M. Hattori, K.F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa, *From genomics to chemical genomics: new developments in KEGG.* Nucleic Acids Research, 2006. **34**(suppl 1): p. D354-D357.

44.	Weininger, D., *SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules.* Journal of Chemical Information and Computer Sciences, 1988. **28**(1): p. 31-36.

45.	Daffé, M., D.C. Crick, and M. Jackson, *Genetics of Capsular Polysaccharides and Cell Envelope (Glyco)lipids.* Microbiology Spectrum, 2014. **2**(4).

46.	Chen, X., A.P. Alonso, D.K. Allen, J.L. Reed, and Y. Shachar-Hill, *Synergy between ^{13}C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in* E. coli. Metab Eng, 2011. **13**(1): p. 38-48.

47.	Beste, D.J., B. Bonde, N. Hawkins, J.L. Ward, M.H. Beale, S. Noack, K. Noh, N.J. Kruger, R.G. Ratcliffe, and J. McFadden, *^{13}C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in mycobacteria which requires isocitrate lyase and carbon dioxide fixation.* PLoS Pathog, 2011. **7**(7): p. e1002091.

48.	Garnier, T., K. Eiglmeier, J.C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P.R. Wheeler, J. Parkhill, B.G. Barrell, S.T. Cole, S.V. Gordon, and R.G. Hewinson, *The complete genome sequence of* Mycobacterium bovis. Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7877-82.

49.	Cole, S.T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, and B.G. Barrell, *Deciphering the biology of* Mycobacterium tuberculosis *from the complete genome sequence.* Nature, 1998. **393**(6685): p. 537-544.

50.	Brosch, R., S.V. Gordon, T. Garnier, K. Eiglmeier, W. Frigui, P. Valenti, S. Dos Santos, S. Duthoy, C. Lacroix, C. Garcia-Pelayo, J.K. Inwald, P. Golby, J.N. Garcia, R.G. Hewinson, M.A. Behr, M.A. Quail, C. Churcher, B.G. Barrell, J. Parkhill, and S.T. Cole, *Genome plasticity of BCG and impact on vaccine efficacy.* Proc Natl Acad Sci U S A, 2007. **104**(13): p. 5596-601.

51. Parker, S.K., K.M. Curtin, and M.L. Vasil, *Purification and characterization of mycobacterial phospholipase A: An activity associated with mycobacterial cutinase.* Journal of Bacteriology, 2007. **189**(11): p. 4153-4160.

52. Yang, X., N.M. Nesbitt, E. Dubnau, I. Smith, and N.S. Sampson, *Cholesterol Metabolism Increases the Metabolic Pool of Propionate in* Mycobacterium tuberculosis. Biochemistry, 2009. **48**(18): p. 3819-3821.

53. Kendall, S.L., P. Burgess, R. Balhana, M. Withers, A. Ten Bokum, J.S. Lott, C. Gao, I. Uhia-Castro, and N.G. Stoker, *Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2.* Microbiology, 2010. **156**(Pt 5): p. 1362-71.

54. Kendall, S.L., M. Withers, C.N. Soffair, N.J. Moreland, S. Gurcha, B. Sidders, R. Frita, A. Ten Bokum, G.S. Besra, J.S. Lott, and N.G. Stoker, *A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in* Mycobacterium smegmatis *and* Mycobacterium tuberculosis. Molecular Microbiology, 2007. **65**(3): p. 684-699.

55. Nesbitt, N.M., X. Yang, P. Fontán, I. Kolesnikova, I. Smith, N.S. Sampson, and E. Dubnau, *A Thiolase of* Mycobacterium tuberculosis *Is Required for Virulence and Production of Androstenedione and Androstadienedione from Cholesterol.* Infection and Immunity, 2010. **78**(1): p. 275-282.

56. Uhía, I., B. Galán, F.J. Medrano, and J.L. García, *Characterization of the KstR-dependent promoter of the gene for the first step of the cholesterol degradative pathway in Mycobacterium smegmatis.* Microbiology, 2011. **157**(9): p. 2670-2680.

57. Galagan, J.E., *Genomic insights into tuberculosis.* Nat Rev Genet, 2014. **15**(5): p. 307-320.

58. Andries, K., P. Verhasselt, J. Guillemont, H.W.H. Göhlmann, J.-M. Neefs, H. Winkler, J. Van Gestel, P. Timmerman, M. Zhu, E. Lee, P. Williams, D. de Chaffoy, E. Huitric, S. Hoffner, E. Cambau, C. Truffot-Pernot, N. Lounis, and V. Jarlier, *A Diarylquinoline Drug Active on the ATP Synthase of* Mycobacterium tuberculosis. Science, 2005. **307**(5707): p. 223-227.

59. Makarov, V., G. Manina, K. Mikusova, U. Möllmann, O. Ryabova, B. Saint-Joanis, N. Dhar, M.R. Pasca, S. Buroni, A.P. Lucarelli, A. Milano, E. De Rossi, M. Belanova, A. Bobovska, P. Dianiskova, J. Kordulakova, C. Sala, E. Fullam, P. Schneider, J.D. McKinney, P. Brodin, T. Christophe, S. Waddell, P. Butcher, J. Albrethsen, I. Rosenkrands, R. Brosch, V. Nandi, S. Bharath, S. Gaonkar, R.K. Shandil, V. Balasubramanian, T. Balganesh, S. Tyagi, J. Grosset, G. Riccardi, and S.T. Cole, *Benzothiazinones Kill* Mycobacterium tuberculosis *by Blocking Arabinan Synthesis.* Science, 2009. **324**(5928): p. 801-804.

60. Raman, K., K. Yeturu, and N. Chandra, *targetTB: A target identification pipeline for* Mycobacterium tuberculosis *through an interactome, reactome and genome-scale structural analysis.* BMC Systems Biology, 2008. **2**(1): p. 109.

61. Radusky, L., L.A. Defelipe, E. Lanzarotti, J. Luque, X. Barril, M.A. Marti, and A.G. Turjanski, *TuberQ: a* Mycobacterium tuberculosis *protein druggability database.* Database, 2014. **2014**.

62. Chang, J.C., M.D. Miner, A.K. Pandey, W.P. Gill, N.S. Harik, C.M. Sassetti, and D.R. Sherman, igr *Genes and* Mycobacterium tuberculosis *cholesterol metabolism.* J Bacteriol, 2009. **191**(16): p. 5232-9.

63. Petrusma, M., L. Dijkhuizen, and R. van der Geize, *Structural features in the KshA terminal oxygenase protein that determine substrate preference of 3-ketosteroid 9α-hydroxylase enzymes.* J Bacteriol, 2012. **194**(1): p. 115-21.

64. Chakraborty, S., T. Gruber, C.E. Barry, H.I. Boshoff, and K.Y. Rhee, *Para-Aminosalicylic Acid Acts as an Alternative Substrate of Folate Metabolism in* Mycobacterium tuberculosis. Science, 2013. **339**(6115): p. 88-91.

65. Blazier, A.S. and J.A. Papin, *Integration of expression data in genome-scale metabolic network reconstructions.* Front Physiol, 2012. **3**: p. 299.

66. Hyduke, D.R., N.E. Lewis, and B.O. Palsson, *Analysis of omics data with genome-scale models of metabolism.* Molecular BioSystems, 2013. **9**(2): p. 167-174.

67. Reed, J.L., *Shrinking the Metabolic Solution Space Using Experimental Datasets.* PLoS Comput Biol, 2012. **8**(8): p. e1002662.

68. Machado, D. and M. Herrgard, *Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism.* PLoS Comput Biol, 2014. **10**(4): p. e1003580.

69. Boshoff, H.I., T.G. Myers, B.R. Copp, M.R. McNeil, M.A. Wilson, and C.E. Barry, 3rd, *The transcriptional responses of* Mycobacterium tuberculosis *to inhibitors of metabolism: novel insights into drug mechanisms of action.* J Biol Chem, 2004. **279**(38): p. 40174-84.

70. Fang, X., A. Wallqvist, and J. Reifman, *Modeling Phenotypic Metabolic Adaptations of* Mycobacterium tuberculosis *H37Rv under Hypoxia.* PLoS Comput Biol, 2012. **8**(9): p. e1002688.

71. Bonde, B.K., D.J. Beste, E. Laing, A.M. Kierzek, and J. McFadden, *Differential producibility analysis (DPA) of transcriptomic data with metabolic networks: deconstructing the metabolic response of* M. tuberculosis. PLoS Comput Biol, 2011. **7**(6): p. e1002060.

72. Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan, and G.K. Schoolnik, *Transcriptional Adaptation of* Mycobacterium tuberculosis *within Macrophages: Insights into the Phagosomal Environment.* J Exp Med, 2003. **198**(5): p. 693-704.

73. Garton, N.J., S.J. Waddell, A.L. Sherratt, S.-M. Lee, R.J. Smith, C. Senner, J. Hinds, K. Rajakumar, R.A. Adegbola, G.S. Besra, P.D. Butcher, and M.R. Barer, *Cytological and Transcript Analyses Reveal Fat and Lazy Persister-Like Bacilli in Tuberculous Sputum.* PLoS Med, 2008. **5**(4): p. e75.

74. Beste, D.J.V., E. Laing, B. Bonde, C. Avignone-Rossa, M.E. Bushell, and J.J. McFadden, *Transcriptomic Analysis Identifies Growth Rate Modulation as a Component of the Adaptation of Mycobacteria to Survival inside the Macrophage.* Journal of Bacteriology, 2007. **189**(11): p. 3969-3976.

75. Balázsi, G., A.P. Heath, L. Shi, and M.L. Gennaro, *The temporal response of the* Mycobacterium tuberculosis *gene regulatory network during growth arrest.* Molecular Systems Biology, 2008. **4**(1).

76. van Dam, J.C.J., P.J. Schaap, V.A.P. Martins dos Santos, and M. Suarez-Diez, *Integration of heterogeneous molecular networks to unravel gene-regulation in* Mycobacterium tuberculosis. BMC Syst Biol, 2014. **8**(111).

77. Teusink, B., A. Wiersma, L. Jacobs, R.A. Notebaart, and E.J. Smid, *Understanding the adaptive growth strategy of* Lactobacillus plantarum *by* in silico *optimisation.* PLoS Comput Biol, 2009. **5**(6): p. e1000410.

78. Shi, L., C.D. Sohaskey, C. Pfeiffer, P. Datta, M. Parks, J. McFadden, R.J. North, and M.L. Gennaro, *Carbon flux rerouting during* Mycobacterium tuberculosis *growth arrest.* Mol Microbiol, 2010. **78**(5): p. 1199-215.

79. Gouzy, A., Y. Poquet, and O. Neyrolles, *A central role for aspartate in* Mycobacterium tuberculosis *physiology and virulence.* Front Cell Infect Microbiol, 2013. **3**: p. 68.

80. Niederweis, M., *Nutrient acquisition by mycobacteria.* Microbiology, 2008. **154**(3): p. 679-692.

81. Fang, X., A. Wallqvist, and J. Reifman, *A systems biology framework for modeling metabolic enzyme inhibition of* Mycobacterium tuberculosis. BMC Syst Biol, 2009. **3**: p. 92.

82. Thiele, I., N. Swainston, R.M.T. Fleming, A. Hoppe, S. Sahoo, M.K. Aurich, H. Haraldsdottir, M.L. Mo, O. Rolfsson, M.D. Stobbe, S.G. Thorleifsson, R. Agren, C. Bolling, S. Bordel, A.K. Chavali, P. Dobson, W.B. Dunn, L. Endler, D. Hala, M. Hucka, D. Hull, D. Jameson, N. Jamshidi, J.J. Jonsson, N. Juty, S. Keating, I. Nookaew, N. Le Novere, N. Malys, A. Mazein, J.A. Papin, N.D. Price, E. Selkov Sr, M.I. Sigurdsson, E. Simeonidis, N. Sonnenschein, K. Smallbone, A. Sorokin, J.H.G.M. van Beek, D. Weichart, I. Goryanin, J. Nielsen, H.V. Westerhoff, D.B. Kell, P. Mendes, and B.O. Palsson, *A community-driven global reconstruction of human metabolism.* Nat Biotech, 2013. **31**(5): p. 419-425.

83. Amaral, L. and M. Viveiros, *Why thioridazine in combination with antibiotics cures extensively drug-resistant* Mycobacterium tuberculosis *infections.* International Journal of Antimicrobial Agents, 2012. **39**(5): p. 376-380.

84. Varma, A., B.W. Boesch, and B.O. Palsson, *Stoichiometric interpretation of* Escherichia coli *glucose catabolism under various oxygenation rates.* Applied and Environmental Microbiology, 1993. **59**(8): p. 2465-2473.

85. Varma, A. and B.O. Palsson, *Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type* Escherichia coli *W3110.* Applied and Environmental Microbiology, 1994. **60**(10): p. 3724-3731.
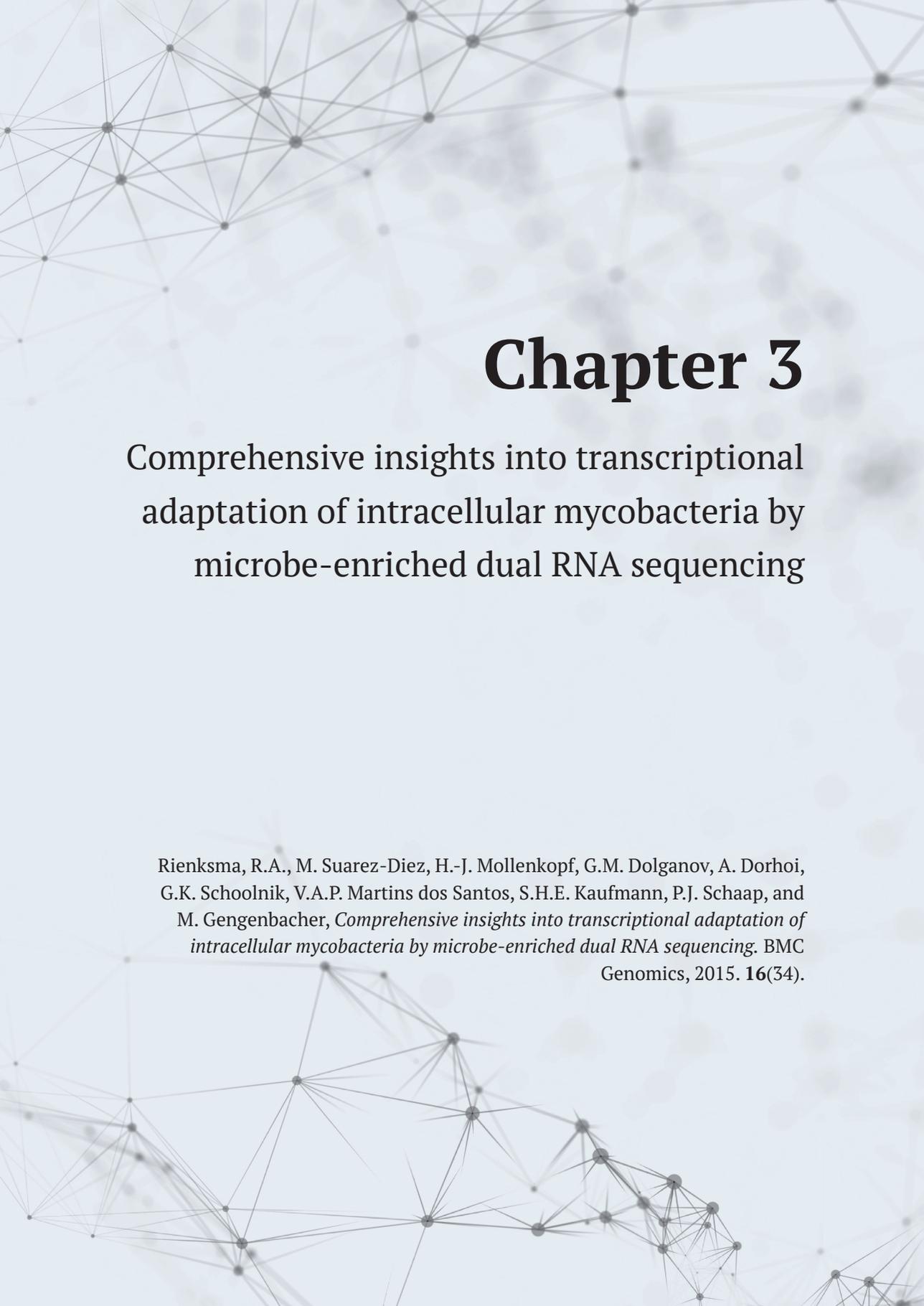
Chapter 3

Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing

Abstract

The human pathogen *Mycobacterium tuberculosis* has the capacity to escape eradication by professional phagocytes. During infection, *M. tuberculosis* resists the harsh environment of phagosomes and actively manipulates macrophages and dendritic cells to ensure prolonged intracellular survival. In contrast to other intracellular pathogens, it has remained difficult to capture the transcriptome of mycobacteria during infection due to an unfavorable host-to-pathogen ratio.

We infected the human macrophage-like cell line THP-1 with the attenuated *M. tuberculosis* surrogate *M. bovis* Bacillus Calmette–Guérin (M. bovis BCG). Mycobacterial RNA was up to 1000-fold underrepresented in total RNA preparations of infected host cells. We employed microbial enrichment combined with specific ribosomal RNA depletion to simultaneously analyze the transcriptional responses of host and pathogen during infection by dual RNA sequencing. Our results confirm that mycobacterial pathways for cholesterol degradation and iron acquisition are upregulated during infection. In addition, genes involved in the methylcitrate cycle, aspartate metabolism and recycling of mycolic acids were induced. In response to *M. bovis* BCG infection, host cells upregulated *de novo* cholesterol biosynthesis presumably to compensate for the loss of this metabolite by bacterial catabolism.

Dual RNA sequencing allows simultaneous capture of the global transcriptome of host and pathogen, during infection. However, mycobacteria remained problematic due to their relatively low number per host cell resulting in an unfavorable bacterium-to-host RNA ratio. Here, we use a strategy that combines enrichment for bacterial transcripts and dual RNA sequencing to provide the most comprehensive transcriptome of intracellular mycobacteria to date. The knowledge acquired into the pathogen and host pathways regulated during infection may contribute to a solid basis for the deployment of novel intervention strategies to tackle infection.

Keywords: *Mycobacterium bovis* BCG, THP-1 cells, infection, host-microbe interaction, transcriptome, dual RNA sequencing, microbe enrichment.

Comprehensive insights into transcriptional adaptation of intracellular
mycobacteria by microbe-enriched dual RNA sequencing

3

Background

Tuberculosis (TB) is an infectious disease caused by the airborne pathogen *Mycobacterium tuberculosis* and accounts for 1.3 million fatalities annually [1]. Unlike non-pathogenic microbes that are eliminated inside the maturing phagosome of immune cells such as macrophages, *M. tuberculosis* brings phagosome maturation to a halt and manages to cope with various host threats including acidification, reactive radicals and nutrient limitation [2]. Studying the transcriptome of intracellular pathogens, in particular *M. tuberculosis*, during infection remained difficult due to a low bacteria-to-host RNA ratio. For different pathogens the number of organisms per host cell spans several orders of magnitudes ranging from 1 to 10 for *M. tuberculosis* and up to 1000 for *Chlamydia* [3, 4].

The first insights into the intracellular life of *M. tuberculosis* provided by comparative microarray analysis, revealed a switch from aerobic to anaerobic respiration, induction of the dormancy regulon *dosR* and iron scavenging as well as upregulation of β-oxidation of fatty acids upon infection [5]. Similar technologies and quantitative real-time PCR were applied to broaden our understanding of specific aspects of intracellular *M. tuberculosis* [6-9]. Microarray probes have the disadvantage of unspecific cross-hybridization between pathogen and host [4], and most often such probes are not optimized for minimal cross-reactivity with other species. Cappelli and colleagues [8] estimated that non-specific signals account for up to 12.5% of all signals. Additionally, transcription of non-coding regions and missed or miss-annotated genes often remain disregarded due to a limited array design. Quantitative real-time PCR has mostly been applied to small subsets of genes, since detection of each transcript requires a pair of specific oligonucleotides [6-9].

Dual RNA sequencing (dual RNA-seq) is a relatively novel technique to study gene expression profiles. This technique allows unbiased and simultaneous sequencing of transcriptomes of multiple organisms and therefore is a superb technology to study intracellular pathogens during infection of host cells. The sequencing reads can subsequently be matched *in silico* to the respective organism. Without prior knowledge of sample content, its composition can be deduced from dual RNA-seq datasets without targeting specific species [10]. Most importantly, dual RNA-seq captures the transcriptome in its entirety thereby overcoming the limitations of microarrays discussed above. First application of this technology to study M. *avium* subsp. *paratuberculosis* during macrophage infection has shed new light on mycobacterial iron acquisition [11].

The attenuated TB vaccine strain *M. bovis* Bacillus Calmette–Guérin (*M. bovis* BCG) has been widely used in research as surrogate for pathogenic *M. tuberculosis* due to a high degree of genome identity [12-14]. In this study, we investigated the transcriptional adaptation of *M. bovis* BCG 24 hours after infection of the human macrophage-like cell line THP-1 by dual RNA-seq. The underrepresentation of bacterial RNA in preparations of total RNA from infected host cells requires high sequencing depth to gain statistical significance and adequate pathogen coverage, leading to increased costs. Mangan and colleagues developed a method entailing differential lysis with guanidine thiocyanate to enrich for mycobacteria from infected macrophages, thus avoiding massive underrepresentation of bacterial RNA as compared to total RNA preparations of infected cells [15]. This method has been used for *in vivo* transcriptome studies using microarrays [6, 16]. Here we present a strategy that combines bacterial enrichment for bacterial transcripts and dual RNA-seq, which we evaluate against non-enriched samples.

Results

Twenty-four hours post-infection, THP-1 cells were harvested and total RNA was isolated. Additionally, two out of three infected THP-1 samples were enriched for *M. bovis* BCG bacilli, using the procedure described in Materials and Methods. The analysis of the 50-bp RNA-derived paired-end sequencing data is illustrated in Figure 3.1. Two out of the three datasets derived from the non-enriched infections (IF1/2) were compared to a reference sample with uninfected THP-1 cells (THP) and differentially expressed THP-1 genes were identified. For differential *M. bovis* BCG gene and small RNA expression analysis, the datasets derived from enriched infections (IF1/2ER) were compared to a reference culture of exponentially growing *M. bovis* BCG (EGB). A spike-in sample (SPI) was used to estimate the percentage of infected cells and to correlate the reads of spiked-in *M. bovis* BCG with the *M. bovis* BCG culture and the non-enriched THP-1 infections with *M. bovis* BCG. An overview of the primary sequencing data is depicted in Table 3.1.

Pathogen specific enrichment strategy is effective
It has been estimated that a minimum of 2–5 million reads from a ribosomal RNA-depleted library is required to adequately cover the gene expression profile of a pathogen in a dual RNA-seq experiment [17-19]. Datasets IF0, IF1 and IF2, derived from non-enriched infections contained 0.4, 1.6 and 1.1 million 50-bp reads that aligned to the *M. bovis* BCG genome, which was too low for significant coverage of the gene expression profile. Subsequently, an enrichment strategy for *M. bovis* BCG was applied to overcome this obstacle, thereby increasing the coverage of

intracellular *M. bovis* BCG transcripts. These enriched datasets (IF1ER and IF2ER)
contained 6.1 and 3.3 million 50-bp reads that aligned to the *M. bovis* BCG genome
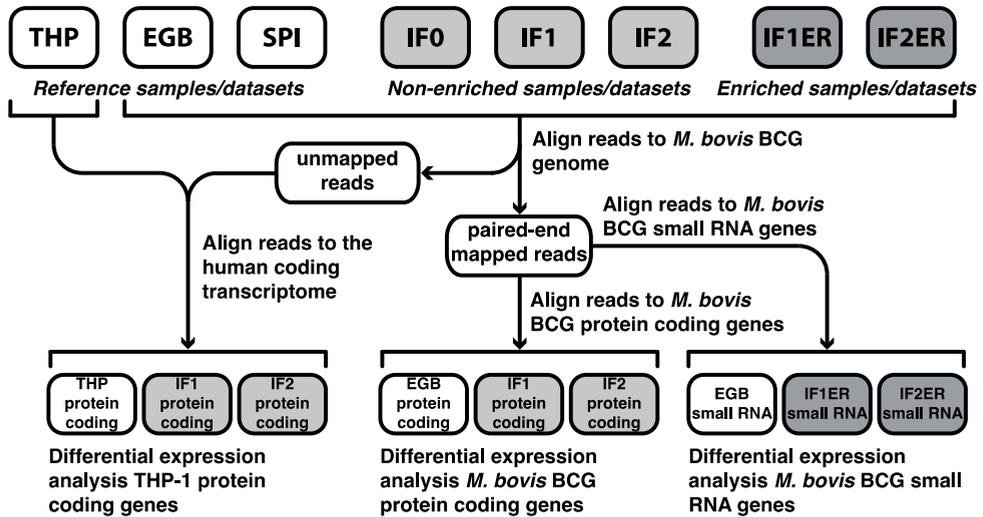(Table 3.1).

Figure 3.1 - Schematic overview of RNA sequencing data analysis.
A total of eight datasets were processed by aligning the 50-bp paired-end sequencing reads to the human transcriptome, the *M. bovis* BCG genome, and/or the *M. bovis* BCG gene and small RNA sequences. Six of these datasets were used for differential gene and/or small RNA expression analysis. (THP: Reference dataset for the THP-1 transcriptome, EGB: Reference dataset for the (exponentially growing) *M. bovis* BCG transcriptome, SPI: Spike-in dataset, IF0/1/2: Datasets of *M. bovis* BCG bacilli infecting THP-1 cells, IF1/2ER: Datasets of *M. bovis* BCG cells infecting THP-1 cells enriched for *M. bovis* BCG bacilli)

Table 3.1 - Reads (millions and percentages) mapped on the human transcriptome and the M. bovis BCG genome.

Dataset	Description	Human transcriptome		M. bovis BCG genome		
		M of reads	%	M of reads	%	Total
THP	Uninfected THP-1 cells	30.9	100	–	–	30.9
EGB	*M. bovis* BCG bacilli	–	–	168	100	168
SPI	Mixed THP-1 and *M. bovis* BCG RNA	31.8	91.0	3.16	9.0	35.0
IF0	Infected THP-1 cells replicate 0	21.5	98.0	0.45	2.0	21.9
IF1	Infected THP-1 cells replicate 1	38.2	96.0	1.57	4.0	39.7
IF2	Infected THP-1 cells replicate 2	28.0	96.3	1.07	3.7	29.0
IF1ER	Infected THP-1 cells replicate 1 enriched for *M. bovis* BCG bacilli	18.0	74.7	6.09	25.3	24.0
IF2ER	Infected THP-1 cells replicate 2 enriched for *M. bovis* BCG bacilli	26.0	88.6	3.35	11.4	29.4

The absolute number of *M. bovis* BCG reads of all infected sample preparations was subsequently classified in four different categories: protein-coding RNA, ribosomal RNA, small RNA, and other (Figure 3.2A). We simulated the relationship between the number of identified differentially expressed protein-coding *M. bovis* BCG genes and sequencing depth (Figure 3.2B). For very low numbers of sequencing reads, the number of identified genes increases in a linear way with the library size. With increasing library size the number of correct identifications tends to stabilize (Figure 3.2B). The relative abundance of the four different categories was fairly similar in both, enriched and non-enriched samples, demonstrating that impact of enrichment *per se* on *M. bovis* BCG derived sequencing reads is negligible (Figure 3.2C). The normalized counts of the protein coding *M. bovis* BCG transcripts in the enriched datasets (IF1ER and IF2ER) and the non-enriched datasets (IF1 and IF2) revealed a linear relationship, with Pearson's correlation coefficients of 0.91 and 0.92, respectively (Supplementary File 3.1). We conclude that pathogen enrichment does not introduce any bias to protein-coding gene expression of *M. bovis* BCG. However, the correlation for normalized counts per gene of THP-1 reads between the same datasets is much lower, 0.57 and 0.70, respectively (Supplementary File 3.1). Therefore, the non-enriched datasets (IF1 and IF2) were used for differential gene expression analysis of THP-1 genes. This enrichment procedure thus enabled us to study the intracellular gene expression of *M. bovis* BCG during infection.
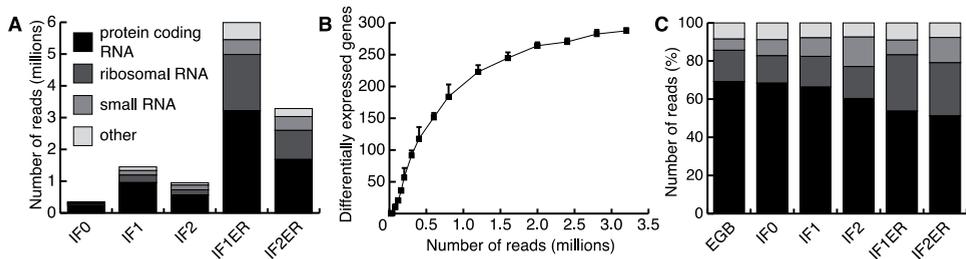

Figure 3.2 - Classification of 50-bp sequencing reads and effect of increasing sequencing depth. **(A)** The total number of 50-bp sequencing reads, matching the paired-end analysis criterion that both reads could be aligned to the *M. bovis* BCG genome, were assigned to four different categories (protein-coding RNA, ribosomal RNA, small RNA, and other). The total of the reads for each sample represents the number of reads aligning to the *M. bovis* BCG genome. **(B)** Simulation of the relation between the number of differentially expressed *M. bovis* BCG genes and sequencing depth. Random subsets of reads were selected from EGB, IF1ER and IF2ER and the mean number (n=5) of reliably identified differentially expressed genes (FDR < 0.05) and the standard deviation (error bars) are given for various sequencing depths. Note that the ratio of a random set to the total set approaches 1 as the size of the random set increases. Therefore, the random samples become more similar to each other and the standard deviation decreases. For reasons of completeness, we have included a standard deviation for every point. **(C)** Classification of the relative number of 50-bp paired-end sequencing reads aligning to the *M. bovis* BCG genome. The legend is the same as in (A).

Comprehensive insights into transcriptional adaptation of intracellular
mycobacteria by microbe-enriched dual RNA sequencing

3

M. bovis BCG response to infection

Twenty-four hours post-infection a clear response of the phagocytosed *M. bovis* BCG bacilli can be observed on the transcriptome level. A total of 367 *M. bovis* BCG genes were differentially expressed (FDR < 0.05), of which 216 were induced and 151 were repressed. A list of all differentially expressed genes of both *M. bovis* BCG and THP-1 cells is provided in Supplementary File 3.2.

M. bovis BCG cholesterol catabolism genes are induced during infection

Cholesterol is a complex lipid that consists of three cyclohexane rings (A, B and C), a cyclopentane ring (D), and an 8-carbon side chain. An incomplete degradation pathway of cholesterol was recently proposed for *M. tuberculosis* [20]. This pathway was extended with the side chain degradation of rings C and D (Supplementary File 3.3) and several genes involved in the pathway were added based on additional literature [21-28]. This extended cholesterol degradation pathway has been previously described in a genome-scale metabolic model of *M. tuberculosis* [29].

We observed a strong increase in expression of almost all genes assigned to cholesterol degradation (Figure 3.3A and Supplementary File 3.3). Initially, cholesterol is taken
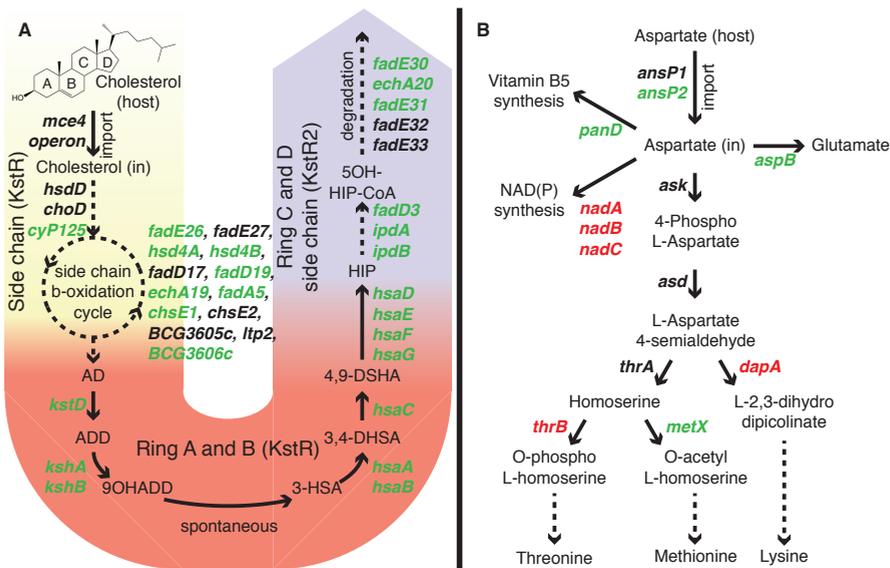

Figure 3.3 - Metabolic processes during infection.
Genes in green are induced upon infection (FDR < 0.05), genes in red are repressed (FDR < 0.05) and genes in black show no differential expression. (A) Cholesterol degradation is divided in three parts: The degradation of the side chain (yellow part), degradation of rings A and B (red part) and the degradation of the side chain of rings C and D (blue part). Dashed arrows represent multiple reactions. The degradation of the rings C and D side chain is based on homologous genes from *Rhodococcus equi*. AD: 4-androstenedione, ADD: 1,4-androstenedione, 9OHADD: 9-hydroxy-1,4, androstene-3-17-dione, 3-HSA: 3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione 3,4-DHSA: 3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione 4,9 DSHA: 4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oic acid, HIP: 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid, 5OH-HIP: 5-hydroxy-methylhexahydro-1-indanone propionate. (B) Aspartate could be imported via AnsP2 and used for the synthesis of vitamin B5, glutamate and methionine. *thrB*, *dapA* and *nadABC* are downregulated, indicating that aspartate is to a lesser extent used to synthesize threonine, lysine and NAD(P).

up by the transport system encoded by the *mce4* gene cluster [30]. The 3β-hydroxyl group is oxidized and isomerized to cholest-4-en-3-one either by the membrane-bound oxidase ChoD or by the dehydrogenase HsdD [21, 31]. No apparent induction of the *mce4* operon, the hydroxysteroid dehydrogenase (HsdD) and cholesterol oxidase (ChoD) coding genes was observed in our datasets. However, the number of transcript reads assigned to the *mce4* operon and to *choD* and *hsdD* indicate that they were expressed in both the infectious and the non-infectious state (Data set S3.1).

Although the degradation of rings A and B is well established, the side chain degradation of rings C and D (Figure 3.3A and Supplementary File 3.3) is less understood in mycobacteria and therefore was reconstructed based on orthology with *Rhodococcus equi genes* [22, 26].
KstR and KstR2 (BCG3639, BCG3621c; Rv3574, Rv3557c) have been previously identified as regulators of cholesterol utilization in mycobacteria [32]. The KstR2 regulon comprises *kstR2* itself and all genes linked to the degradation of the side chain of rings C and D, whereas genes regulated by KstR participate in the degradation of rings A and B and the initial degradation of the cholesterol side-chain (Figure 3.3A and Supplementary File 3.3). In our datasets the expression of *kstR2* was strongly induced upon infection, whereas *kstR* remained unchanged (Data set S3.1). To verify these findings and the expression of other genes we selected a subset of 14 genes, of which 3 encode small RNAs, and designed primers (Supplementary File 3.5) to use for qRT-PCR. Among the selected genes, 5 genes are involved in cholesterol catabolism and 2 genes encode enzymes of the methylcitrate cycle (Supplementary File 3.6). The qRT-PCR results confirmed the integrity of our RNA-seq data.

We analyzed the behavior of the genes in the cholesterol degradation pathway in a compendium of expression data collected for *M. tuberculosis*. Although no condition associated with cholesterol utilization has been included in the compendium, many conditions in our compendium lead to differential expression of genes regulated by KstR and KstR2 (Figure 3.4). Yet, a reduced set of KstR2-regulated genes (*fadD3*, *fadE31* and *ipdA*) exists, which seems to be specifically induced upon infection and most likely specifically reacts to only this kind of perturbation. The specific induction renders FadD3, FadE31 and IpdA of potential interest for therapeutic intervention. Bioinformatics analysis using the consensus IdeR binding motif [33] and the KstR2 binding motif [32] revealed that these regions overlap (Figure 3.5).
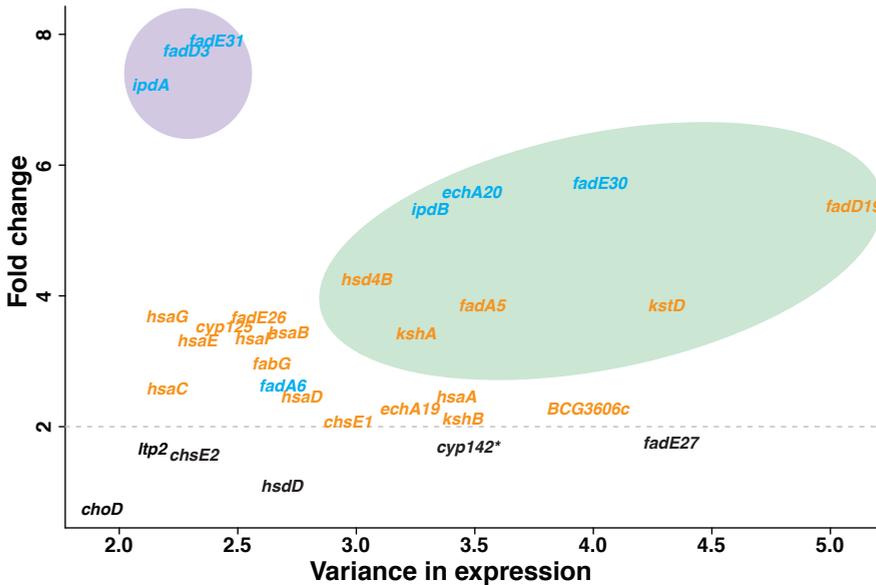
Figure 3.4 - Variance in expression levels (compendium) compared to the fold changes upon infection. Genes in blue are KstR2-regulated whereas orange genes denote KstR-regulated genes. Genes in black show no significant change (FDR > 0.05). Green/violet ellipses denote areas of high/low variability in the expression compendium. *In *M. tuberculosis* H37Rv, *cyp142* has the same function as *cyp125*. In *M. bovis* BCG, *cyp142* encodes an inactive product.
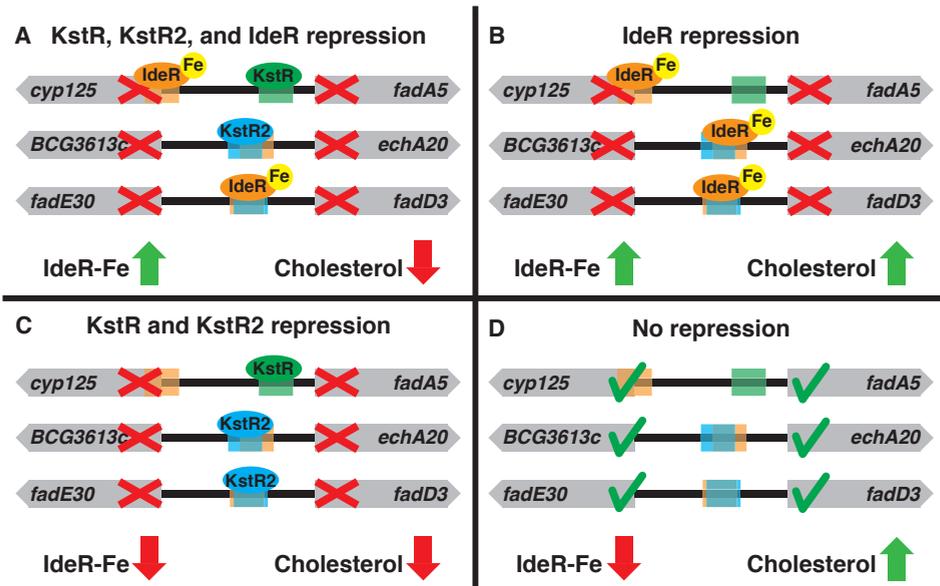
Figure 3.5 - Regulation of the cholesterol degradation pathway by IdeR, KstR and KstR2. Sequences similar to the IdeR binding boxes appear in the upstream regions of genes in the cholesterol degradation pathway in close proximity to (and sometimes overlapping with) the KstR and KstR2 binding regions. (A, B and C) Under either normal iron availability or lack of cholesterol either IdeR or KstR/KstR2 represses the expression of genes in this pathway. (D) Only under low iron availability (relieving IdeR repression) and in presence of cholesterol (relieving KstR and KstR2 repression), can the genes in the cholesterol degradation pathway be expressed.

Griffin and co-workers [34] found that although propionyl-CoA can be derived from other host metabolites, the requirement for the methylcitrate cycle is largely attributable to the degradation of host cholesterol. The induction of the methylcitrate cycle and the slight repression of *icd1* (BCG3409c; Rv3339c), encoding an isocitrate dehydrogenase, suggests that the oxidative part of the citric acid cycle is bypassed in favor of this pathway (Supplementary File 3.4). This emphasizes that cholesterol is the main carbon source for intra-phagosomal *M. bovis* BCG.

Expression profile suggests *M. bovis* BCG recycles mycolic acids

Mycobacterial fatty acids are precursors for mycolic acids and are synthesized by at least two fatty acid synthases: FAS-I and FAS-II [35]. FAS-I consists of a single multifunctional enzyme, encoded by *fas* (BCG2545c; Rv2524c), and elongates fatty acids at the beginning of the mycolic acid synthesis pathway, while FAS-II consists of multiple enzymes and elongates fatty acids created by FAS-I. The mycobacterial genes *umaA1*, *cmaA2*, *hadA*, and *mmaA3* (BCG0509, BCG0546c, BCG0684, BCG0692c; Rv0469, Rv0503c, Rv0635, Rv0643) encode enzymes that further process FAS-II products (Figure 3.6). Previous reports suggested that FadE23 and FadE24 (BCG3163, BCG3162; Rv3140, Rv3139) might be involved in recycling of mycolic acids [24]. Taken together, the expression patterns observed in our study (Figure 3.6) indicate that new acids are rather generated by remodeling existing mycolic acids and host fatty acids than synthesized *de novo*.
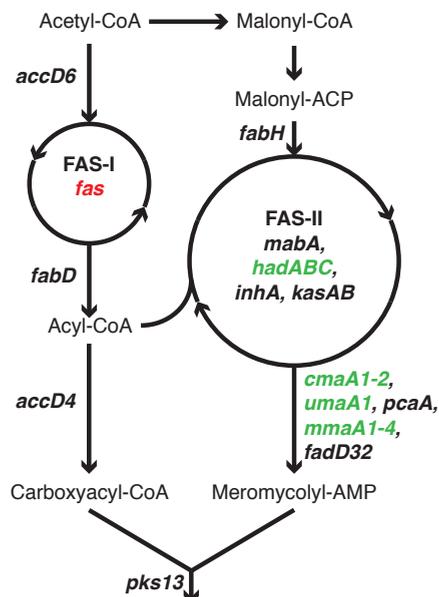

Figure 3.6 - M. bovis BCG gene expression pattern of mycolic acid synthesis
Genes involved in fatty acid synthase II (FAS-II) and downstream of FAS-II are induced (green), while fatty acid synthase I (FAS-I) is repressed (red).

Expression pattern of intracellular *M. bovis* BCG suggests utilization of host aspartate

Gouzy and colleagues showed that nitrogen incorporation from exogenous aspartate is required for host colonization by *M. tuberculosis* [36]. We observed changes in the gene expression pattern upon infection, regarding aspartate metabolism (Figure 3.3B). Intriguingly, the gene encoding the unique aspartate importer AnsP1 (BCG2144; Rv2127) showed no significant change in expression, while its homolog *ansP2* (BCG0385c; Rv0346c) showed a two-fold induction (Data set S3.1). Gouzy and colleagues found that an *M. tuberculosis ansP2*-knock-out (KO) mutant was able to grow on aspartate as sole nitrogen source [36]. Moreover, an *ansP1* mutant showed no growth defect in either resting or activated macrophages. The lack of induction of the sole asparaginase gene *ansA* (BCG1590c; Rv1538c), that can catalyze the conversion of asparagine to aspartate, suggests that, in addition to its reported asparagine transport capacity [37], mycobacterial AnsP2 imports aspartate from the phagosome during infection. Some of the genes that encode aspartate-utilizing enzymes are induced, such as *panD* and *aspB* (BCG3665c, BCG3629; Rv3601c, Rv3565). In particular, AspB was predicted to transfer nitrogen from aspartate to glutamate, which serves as a central nitrogen carrier for alternative metabolic pathways [38], suggesting that *M. bovis* BCG utilizes host aspartate as nitrogen source during infection.

The repression of *de novo* NAD(P) synthesis genes *nadA*, *nadB* and *nadC* (BCG1632, BCG1633, BCG1634; Rv1594, Rv1595, Rv1596) and the absence of significant changes in expression of *pncA*, *pncB*, *nadD* and *nadE* (BCG2062c, BCG1392c, BCG2437c, BCG2457c; Rv2043c, Rv1330c, Rv2421c, Rv2438c) involved in NAD(P) synthesis and salvage [39] (Data set S3.1), indicates that bacterial NAD(P) may become limited during infection. The transcripts of enzymes catalyzing branching reactions towards threonine, methionine and lysine showed an unexpected pattern (Figure 3.3B): both *dapA* (BCG2769c; Rv2753c) and *thrB* (BCG1356; Rv1296), involved in initiation of threonine and lysine biosynthesis respectively were repressed, while *metX* (BCG3411; Rv3341), encoding an enzyme that initiates methionine biosynthesis, was induced. We conclude that host aspartate utilized by *M. bovis* BCG might largely be converted into methionine rather than threonine and lysine.

The induction of *sodA* (BCG3909; Rv3846) (Data set S3.1), encoding superoxide dismutase that destroys harmful radicals, confirms that *M. bovis* BCG counteracts reactive oxygen intermediates (ROI) produced by the host cell [2, 40]. Interestingly, aspartate has the capacity to quench ROI by intramolecular oxidation of the sulphur atom [41]. Although experimental prove has yet to be provided, it is attractive to speculate that mycobacteria produce methionine during infection to support counteraction to ROI.

M. bovis BCG iron scavenging; siderophore synthesis, secretion and import

Mycobactins comprise an essential class of mycobacterial siderophore molecules to access iron of the host. These molecules are synthesized by an array of mycobactin enzymes, consisting of several proteins organized in a megasynthase [42]. The mycobactin megasynthase genes *mbtB–F* were induced upon infection and so were the majority of additional genes involved in mycobactin biosynthesis: *mbtG/I/J/K/N* (BCG2392c, BCG2400c, BCG2399, BCG1409c, BCG1408; Rv2378c, Rv2385, Rv2386c, Rv1347c, Rv1346) (Data set S3.1).

The type VII secretion system ESX-3 is essential for mycobactin-mediated iron acquisition and *in vivo* survival [43]. The ESX-3 secretion system is regulated by ZuR (BCG2373; Rv2359) [44] and consists of 11 genes [45] of which 7 were induced upon infection (Data set S3.1). The repression of *zuR*, resulted in the induction of ESX-3. A siderophore transport system of *M. tuberculosis* consisting of MmpL4 and MmpS4 (BCG0489c, BCG0490c; Rv0450c, Rv0451c) is required for infection of mice [46]. Both *mmpL4* and *mmpS4* and two other genes encoding an inner membrane transporter for mycobactin *irtA/B* [47] (BCG1410/1411; Rv1348/1349) were induced (Data set S3.1). Of the bacterioferritins BfrA/B (iron storage proteins induced by IdeR), only *bfrB* (BCG3904; Rv3841) showed a significant decrease. A possible explanation could be the reduced availability of iron in the host, and thus less iron storage capacity is required.

M. bovis BCG small RNAs

Small RNAs have only recently been discovered in Mycobacteria [48, 49]. Although their function is mostly unclear, they can be present in large quantities [50]. In our study, differential expression was observed for 19 small RNAs (FDR < 0.05). High transcript levels of the small RNAs MTS0997, MTS1338 and MTS2823 were reported in chronically *M. tuberculosis*-infected mouse lungs [50]. We observed a significant (FDR < 0.05) induction of MTS2823 as well, although the fold change is small (logFC = 1.49) compared to other reports. The expression of MTS1338 was repressed in our study, and showed a small fold change (logFC = -2.03). DosR (BCG3156c; Rv3133c) induces the latter in *M. tuberculosis* upon hypoxia and infection [51]. The low expression (below 100 CPM) and lack of induction of *dosR* in our datasets, may explain why MTS1338 remained unchanged. We verified the RNA-seq data by qRT-PCR (Supplementary File 3.6). Whether the contrary expression of MTS1338 in *M. bovis* BCG and in *M. tuberculosis* during infection is critical for virulence remains to be defined.

Host immune response to *M. bovis* BCG is AIM2 dependent

A pathway enrichment analysis using InnateDB [52] revealed that the THP-1 cells show distinct signs of infection (Table 3.2) since we identified numerous enriched

Comprehensive insights into transcriptional adaptation of intracellular
mycobacteria by microbe-enriched dual RNA sequencing

3

pathways involved in immune response such as IFN-α/β signaling, IFN-γ signaling
and RIG-I/MDA5-mediated induction of IFN-α/β pathways. Interferons (IFNs) are
synthesized by the host upon infection and trigger the activation of its immune
system. IFNs can be divided in three classes: type I IFNs (IFN-α, IFN-β, IFN-ε and
IFN-ω), type II IFNs (IFN-γ) and type III IFNs [53].

Shah and colleagues [54] showed that virulent mycobacteria, such as *M. tuberculosis*
inhibit IFN-β production and signaling, resulting in the inhibition of the activation
of AIM2 (interferon-inducible protein). AIM2 is part of the inflammasome that
recognizes cytosolic bacterial and viral DNA, thereby contributing to the host's
defense. In contrast to virulent mycobacteria, nonvirulent mycobacteria such as
M. smegmatis, induce AIM2 [54]. *M. bovis* BCG seems to respond similarly to other
nonvirulent mycobacteria, as the transcription of the gene encoding AIM2 is highly
induced (Data set S1) as well as the IFN-α/β signaling pathway and the cytosolic
DNA-sensing pathway (Table 3.2).

Table 3.2 - Induced THP-1 pathways upon *M. bovis* BCG infection.
The ten most significantly induced pathways are shown.

Pathway name	Number of genes annotated in pathway	Number of induced genes	P-Value
IFN-α/β signaling	36	23	$< 1.0 \times 10^{-5}$
IFN-γ signaling	28	16	$< 1.0 \times 10^{-5}$
RIG-I/MDA5-mediated induction of IFN-α/β pathways	12	8	$< 1.0 \times 10^{-5}$
Cytosolic DNA-sensing pathway	23	8	0.00018
Cholesterol biosynthesis	14	6	0.0033
Hepatitis C	71	14	0.00073
Staphylococcus aureus infection	11	5	0.00078
Steroid biosynthesis	11	5	0.00078
Iron uptake and transport	7	4	0.00096
Toxoplasmosis	60	12	0.00152

**Host genes involved in glycolysis and ketogenesis are induced upon
mycobacterial infection**

Phagocytosis of pathogenic mycobacteria triggers the accumulation of lipid bodies
in the host cell described as foamy phenotype [55]. Secretion of mycobacterial
ESAT-6 is required to mediate this process by stimulating the uptake of glucose
into the host cell, which might lead to increased glycolytic activity and elevated
levels of acetyl-CoA, which in turn leads to the generation of D-3-hydroxybutyrate
via ketogenesis [55]. Although *M. bovis* BCG lost the ESX-1 locus, a major virulence
determinant of pathogenic mycobacteria that encodes the effector proteins ESAT-6

and CFP-10 [13, 56], we observe increased expression of several glycolytic enzymes including HK3, GPI, PFKP, FBPI, GAPDH, and PGAM1 (Supplementary File 3.2). Moreover, we found genes of the ketogenesis pathway induced: ACAT2, HMGCS1, BDH2 and HMGCR. The latter gene encodes HMG-CoA reductase, which catalyzes the conversion of hydroxy-β-methylglutaryl CoA and leads to the synthesis of cholesterol and other sterols. BDH2 encodes 3-hydroxybutyrate dehydrogenase, which catalyzes the reversible conversion of acetoacetate to D-3-hydroxybutyrate. Subsequent steps that lead to mycobacteria-induced formation of lipid bodies in the host cell involve the activation of the anti-lipolytic G protein-coupled receptor GPR109A, which triggers adenylyl cyclase. The resulting decrease in host cyclic AMP levels leads to a decrease in phosphorylation of stored host lipids by protein kinase A (PKA), rendering them less vulnerable to lipolysis by hormone sensitive lipase (HSL), thus promoting the formation of lipid bodies [55]. Consistent with the attenuated *M. bovis* BCG strain, we do not observe a change in the expression of GPR109A, adenylyl cyclase, PKA and HSL, indicating that this part of the pathway leading to the formation of lipid bodies is not active in the host or that this response is regulated post-transcriptionally and therefore remains invisible using a transcriptome approach.

Taken together, despite the absence of the ESX-1 locus in *M. bovis* BCG, the host response regarding the initial steps of lipid body formation is similar to that of *M. tuberculosis*. Several pathogen factors including mycolic acids that were demonstrated to induce the foamy phenotype in macrophages [57], may therefore be required to reprogram the host for lipid build-up.

Cholesterol is synthesized and iron losses are compensated by the host upon infection

Four out of seven genes in the THP-1 iron uptake and transport pathway were induced. The enzyme encoded by the induced gene HMOX1 encodes heme oxygenase 1, assigned to iron uptake and transport, catalyzes the rate-limiting step of heme degradation and is required to confer host resistance to mycobacterial infection in mice [58]. Among the other induced genes were FTH1 and FTL, encoding the heavy and light polypeptide of ferritin. This suggests that the THP-1 cells compensate for the loss of iron caused by *M. bovis* BCG, by taking up extra iron and degrading heme.

The THP-1 cholesterol biosynthesis pathway was induced, as six genes of this pathway, including HMGCR encoding the rate-limiting enzyme for cholesterol biosynthesis, were upregulated. We conclude that infected macrophages synthesize cholesterol to compensate for loss of this molecule caused by mycobacterial catabolism.

Comprehensive insights into transcriptional adaptation of intracellular
mycobacteria by microbe-enriched dual RNA sequencing

3

Discussion

This study describes a deep sequencing approach towards the elucidation of mycobacterial and host cell gene expression profiles during intracellular infection. Initially we employed standard deep sequencing settings for eukaryotes to resolve the transcriptional profile of intracellular mycobacteria. Although this set-up allowed, to some extent, analyzing gene expression in *M. bovis* BCG, mycobacterial transcript coverage was insufficient. Increasing the sequencing depth was hindered by high sequencing costs, thereby preventing the method to become a broad application. Thus, we decided to enrich for mycobacteria during sample preparation of infected THP-1 cells. Indeed, the strategy applied greatly increased the coverage of the intracellular *M. bovis* BCG transcriptome. Although mycobacterial gene expression can be analyzed in non-enriched samples (215 genes, FDR < 0.05), implementation of enrichment greatly expanded the number of reliably identified differentially expressed genes by 71% (367 genes, FDR < 0.05). Moreover, the sequencing depth simulation (Figure 3.2B) revealed that enrichment allowed identification of differentially expressed genes that would have been missed otherwise. Repasy and colleagues showed that for an *in vivo* infection setting with mice and *M. tuberculosis* the MOI ranges from 1 to 5 [3]. Therefore, when studying an *in vivo* infection, the enrichment might not be sufficient to obtain a similar sequencing depth as obtained in our *in vitro* study, for which an MOI of 10 was used. To determine whether enrichment introduced any bias into the datasets, we analyzed non-enriched and enriched samples of two independent biological infection replicates. Although the non-enriched and enriched samples comprised different numbers of intracellular *M. bovis* BCG reads, the transcriptomes of respective samples of both infection experiments correlated well as indicated by high linearity and correlation coefficients (Supplementary File 3.1). Additionally, the correlation of intracellular *M. bovis* BCG expression between the biological replicates of both enriched datasets was comparable with the non-enriched datasets (0.93 and 0.94, respectively), verifying that the enrichment procedure was repeatable and robust, and did not introduce any bias to the intracellular mycobacterial transcriptome. For host expression we identified a lower correlation between the non-enriched and enriched samples (0.57 for IF1 and IF1ER, and 0.70 for IF2 and IF2ER) (Supplementary File 3.1). Hence, the datasets of non-enriched samples were favored for analysis of the host transcripts in order to preserve accuracy of the transcriptional landscape during infection.

Our method is dependent on differential susceptibility to lysis of host and microbial cells and not on a molecular sequence capture or depletion method as previously described [11]. This has the added advantage that small RNAs can be detected and analyzed for differential expression. Additionally, our method

is independent of mRNA polyadenylation, as e.g. MICROBEnrich simultaneously captures and removes polyadenylated mRNAs along with rRNAs. We consider this a critical point because previous findings support the existence of polyadenylated tracts in mRNA of mycobacteria [59] and other bacteria [60, 61].

It has been shown previously that mycolic acid liposomes are phagocytized by murine macrophages, changing the morphology of the macrophages to foam-like cells accumulating cholesterol [57]. Although the cholesterol degradation pathway is induced at 24 hours post-infection (Figure 3.3A and Supplementary File 3.3), the *mce4* gene cluster and *hsdD* and *choD* were not induced. Nonetheless, these genes were expressed in both conditions and respective proteins could be already present before infection, ready for a situation when cholesterol becomes available. For *hsdD* and *choD*, it is tempting to speculate that alternative genes with a similar function exist, since these genes have been found dispensable for cholesterol degradation in mycobacteria [21].

Gene regulation of mycobacterial cholesterol catabolism involves a complex interplay between KstR, KstR2, and IdeR. KstR and KstR2 are the prime regulators of the genes depicted in Figure 3.3A. A clear distinction between their targets becomes apparent with KstR regulating the degradation of rings A and B and KstR2 regulating the degradation of rings C and D. In addition, genes in these regulons (*fadA5*, *fadD3*, *fadE30*, *echA20*) contain IdeR-binding sites in their upstream regions [33] and in some cases a profound overlap between these binding sites was observed. This implies that these genes can only be expressed upon removal of both types of repressors: IdeR under normal iron availability and KstR and KstR2 during cholesterol shortage (Figure 3.5). The upregulation of IdeR-dependent iron uptake systems can be caused not only by low iron availability inside the macrophage but also as a response to the NO-induced damage caused to iron-containing proteins [5]. HsaC, KshA, and the cytochromes Cyp125 and Cyp142 are iron-containing enzymes (and hence are susceptible to NO-induced damage) and lack of functionality of these enzymes leads to the accumulation of stable toxic catabolic intermediates, such as catechol derivatives and cholest-4-en-3-one [62-65]. The IdeR control of this pathway ensures that it is only expressed when the corresponding repair/replacement systems for iron-containing proteins are in place, therefore minimizing the risk of toxic intermediate accumulation.

It has been shown that aspartate functions as a major nitrogen reservoir in the host [36]. In line with this finding, we observed induction of aspartate utilizing enzymes (Figure 3.3B). Interestingly, we did not detect induction of the gene encoding the aspartate transporter *ansP1*, but induction of its homolog, *ansP2*. Earlier reports demonstrated that an *ansP1*-KO mutant fails to import aspartate *in vitro*, but shows wild-type behavior in either resting or activated macrophages, even though aspartate is a major nitrogen source in the host. Moreover, the absence of *ansA*

Comprehensive insights into transcriptional adaptation of intracellular
mycobacteria by microbe-enriched dual RNA sequencing

3

induction suggests AnsP2 functions as an aspartate importer during infection in addition to its role as asparagine importer [37].

Rodríguez and colleagues cultivated *M. tuberculosis* H37Rv on even long-chain fatty acids and analyzed the transcriptome by RNA sequencing, observing a dormancy-related phenotype [66]. Although there are similarities between their *in vitro* model and our results, the induction of cholesterol catabolism, the methylcitrate cycle and aspartate metabolism are not captured using such a method, highlighting the differences between using an *in vitro* model based on even long chain fatty acids as opposed to studying intracellular infection directly.

Conclusions

Dual RNA-seq allowed elucidation of the complex interplay between *M. bovis* BCG and THP-1 macrophages. The comparison of non-enriched and enriched ribosomal RNA-depleted sequencing libraries of two biological replicates from identical infection cultures, showed high correlation of sequencing reads without technical bias. Taken together, microbe-enriched dual RNA-seq is a powerful technology that enables the assessment of the global transcriptome of "low-number" intracellular microbes and their host as demonstrated by the simultaneous induction of *M. bovis* BCG cholesterol degradation genes and host cholesterol synthesis genes.

Methods

Bacterial strains and growth conditions

*M. bovis BCG SSI 1331 (*American Type Culture Collection, #35733*)* was grown in Middlebook 7H9 medium (Becton Dickinson) supplemented with 0.05% Tween 80, 0.2% glycerol, 10% albumin-dextrose-catalase supplement (Becton Dickinson) (7H9-ADC) or on Middlebrook 7H11 agar (Becton Dickinson) containing 0.2% glycerol and 10% oleic acid-albumin-dextrose-catalase enrichment (Becton Dickinson). Mycobacterial cultures were grown to the mid-log phase in 1 L roller bottles (450 cm^2) at 37°C and 2 rpm. For CFU enumeration, serial dilutions were performed in phosphate-buffered saline containing 0.05% Tween 80 and plated on Middlebrook 7H11 agar. Plates were incubated at 37°C for 3–4 weeks prior to counting.

Infection of the human macrophage-like cell line THP-1

The THP-1 cell line (American Type Culture Collection #TIB-202) was maintained in Roswell Park Memorial Institute medium 1640 supplemented with 10% fetal calf

serum, 2 mM glutamine, 1 mM sodium pyruvate and 0.05 mM 2-mercapto-ethanol in a humidified 5% carbon dioxide atmosphere at 37°C. Estimated $4×10^6$ cells/well in a 6-well plate were differentiated for 24 h using culture medium containing 40 ng/ml phorbol 12-myristate 13-acetate. Cells were then washed with fresh culture medium and incubated for 48 h. *M. bovis* BCG was pelleted (3,200 rpm, RT, 10 min), washed twice with phosphate buffered saline and resuspended in THP-1 culture medium. THP-1 cells were infected with *M. bovis* BCG at a multiplicity of infection of 10. The *M. bovis* BCG culture medium suspension was added to the differentiated cells, centrifuged for 3 min at 800 rpm and subsequently incubated for 4 h. The infection mix was removed, cells were washed twice with pre-warmed PBS and incubated with fresh culture medium for 20 h prior to RNA extraction.

RNA isolation and mycobacterial RNA enrichment of infected cells
Total RNA from mycobacterial cultures was prepared as previously described [67]. Extraction of total RNA from THP-1 cells was prepared with TRIzol reagent using glycogen as a carrier according to the suppliers' recommendation (Life Technologies). Total RNA from *M. bovis* BCG infected THP-1 cells was isolated by abrasive particles in a reciprocal shaker with TRIzol [67]. Enrichment of mycobacteria from infected THP-1 cells was carried out by differential lysis of host and mycobacterial cells by guanidine thiocyanate (GITC).
Infected cells were washed with PBS at RT. Cold 4M GITC was added to the monolayer and the cells were transferred to a 1.5-ml screw cap tube. After centrifugation the pellet was resuspended in residual GITC and mixed with 1 ml TRIzol containing 20 μg/ml linear acrylamide, followed by incubation for 5 min at RT. Bacteria were disrupted by bead beating (FastPrep Instrument; two cycles of 30s at maximum speed with cooling on ice between cycles). The sample was centrifuged for 1 min at 4°C/13,000 rpm and the supernatant was transferred to a 2-ml screw cap tube containing 200 μl chloroform, mixed and incubated at RT for 5 min. After centrifugation at 4°C/13,000 rpm for 10 min, RNA was extracted from the aqueous phase using the Qiagen RNeasy mini kit including an on-column DNase digestion (Qiagen). Quality and quantity of total RNA were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies) and a NanoDrop 1000 spectrophotometer (Kisker).

Calculation of spike-in concentration
Total RNA of THP-1 cells and *M. bovis* BCG cultures was determined on the basis of cell counts and RNA isolation yield. As the intracellular copy number for a pathogen varies from species to species we assumed the most minimal infection rate of one mycobacterium per host cell. Therefore a cellular multiplicity of infection of 1:1 was used for mixing host and pathogen RNA, resulting in a ratio of 1000:1 total RNA, derived from the proportion of RNA abundance per cell between host and pathogen.

Comprehensive insights into transcriptional adaptation of intracellular
mycobacteria by microbe-enriched dual RNA sequencing

3

RNA sequencing

The RNA-seq libraries were prepared according to the TruSeq RNA Sample Preparation v2 Guide (Illumina) without fragmentation and without size selection. Using an Agilent Bioanalyzer high sensitivity DNA kit, we confirmed that the average final library size was already approximately 350 bp without fragmentation; therefore the library insert fragmentation time at 94 °C was set to 0 minutes, as no additional fragmentation step was required. Up to 98% of bacterial RNA consists of ribosomal RNA, which can prevent adequate coverage of a bacterial transcriptome, using RNA sequencing [68, 69]. Therefore, the Gram-Positive Bacteria Ribo-Zero (Epicentre) rRNA Magnetic Removal Kit was used to remove bacterial rRNA from mycobacterial total RNA. For THP-1 total RNA the Ribo-Zero Magnetic Kit Human/ Mouse/Rat was used, while depletion of rRNA from infections without enrichment as well as the spike-in experiment was done by a two-step procedure with both kits. The mycobacterial enriched total RNA preparations of infected cells were depleted with the Ribo-Zero Magnetic Gold Kit Epidemiology. All cDNA libraries were checked for quality using the DNA-1000 kit (Agilent) on a 2100 Bioanalyzer and quantified with the Qubit 2.0 Fluorometer (Life Technologies). Libraries of each consecutive experiment were pooled as 4-plex and on-board loaded with a Hi-Seq 1500 instrument. The sequencing reaction was carried out as Rapid Run using a TruSeq Rapid PE Cluster Kit and a TruSeq Rapid SBS Kit and 2 × 51 cycles including 7 cycles indexing in order to obtain 50-bp paired-end reads.

Data analysis pipeline

A total of eight datasets were created: Uninfected THP-1 cells (dataset: THP), exponentially growing *M. bovis* BCG (dataset: EGB), a spike-in sample consisting of THP-1 RNA and *M. bovis* BCG RNA in a 1000:1 ratio (dataset: SPI), *M. bovis* BCG-infected THP-1 cells with a multiplicity of infection of 10 *M. bovis* BCG bacilli per THP-1 host cell (dataset IF0), two additional independent biological replicates of *M. bovis* BCG-infected THP-1 cells (datasets: IF1 and IF2), and two datasets prepared with the enrichment method using the two biological replicates of *M. bovis* BCG-infected THP-1 cells (datasets: IF1ER and IF2ER).

All 50-bp paired-end reads from datasets IF1/2ER, IF0/1/2, EGB were aligned, using megablast, against the complete genome sequence of *M. bovis* BCG Pasteur, obtained from the NCBI bacterial genome database (ftp.ncbi.hih.gov/genomes/ Bacteria) and to the human transcriptome obtained from the Ensemble database (www.ensembl.org). If both 50-bp reads of a given pair could not be aligned to the human transcriptome, or the *M. bovis* BCG genome, they were discarded from further analysis. The 50-bp paired-end reads aligning to the *M. bovis* BCG genome were subsequently aligned to protein-coding genes, ribosomal genes, and small RNA (sRNA) genes. The protein-coding gene sequences and the ribosomal

gene sequences were obtained from the NCBI database. The sRNA-coding gene sequences were obtained from the Bacterial Small Regulatory RNA Database [70] and supplemented with sRNA sequences from *M. tuberculosis* [51].

Gene counting procedure

A single count was assigned to a transcript if a complete pair of reads aligned to the *M. bovis* BCG genome. If only one read of a pair of reads aligned to a given gene, also a single count was assigned to this transcript, assuming that the other read could align to an intergenic region, or another gene, due to the existence of operons. If a pair of reads would align to two different genes, a count was assigned to both genes, resulting in a total of two counts per pair of reads. For THP-1 cells, all reads assigned to different splice variants of the same gene were counted (one count for each aligned read) and summed.

Differential gene expression analysis

The R package edgeR [71] was used for differential gene expression analysis of *M. bovis* BCG and THP-1 genes. Gene and protein functions have been extensively studied in *M. tuberculosis* and therefore the *M. tuberculosis* (strain H37Rv) orthologs of each *M. bovis* BCG gene were included. Low expression tags receiving less than 100 counts per million in two or more of the three datasets: EGB, IF1/2ER (or EGB, IF1/2), were excluded from differential expression analysis as described in the edgeR manual. Tags receiving more than 100 counts per million in the EGB dataset and less than 100 counts per million in either one or both replicates of IF1/2ER or IF1/2 were still used to account for large decreases in expression upon infection. EdgeR is a Bioconductor package designed to identify significant changes between two or more groups, given that at least one of the groups has replicated measurements [71], which is the case for our experimental setup (Figure 3.1). The edgeR algorithm models the read counts associated to a gene using a negative binomial probability distribution. The variance of this distribution takes into account both the stochasticity of the sequencing process and the variability associated to biological variation. The common Biological Coefficient of Variance (BCV) measures the average dispersion in gene expression values associated to biological variability. We used EdgeR to compute the BCV of the corresponding samples prior to computing differential expression. For the BCG genes, we obtain a BCV of 0.16 from the IF1/2ER samples whereas for the THP-1 genes we obtained a BCV of 0.20 from the IF1/2 samples. We assumed that the BCV in datasets EGB and THP to be smaller than those in datasets IF1/2ER and IF1/2, due to the less stable and controlled conditions arising from infection as compared to standard culturing procedures. To compute differential gene expression for *M. bovis* BCG and THP-1 cells, we took a conservative approach and assigned the same BCV to

Comprehensive insights into transcriptional adaptation of intracellular
mycobacteria by microbe-enriched dual RNA sequencing

3

datasets EGB and THP as those obtained from IF1/2ER and IF1/2, respectively. Differential expression was then computed for each gene using a pairwise exact testing procedure. The algorithm of Benjamini and Hochberg was used to control false discovery rates [72]. Protein-coding genes and sRNA-coding genes with an FDR < 0.05 were assumed to be differentially expressed.

For THP-1 cells, gene counts from dataset THP were compared to those of datasets IF1/2. Differential gene expression analysis was performed similar to *M. bovis* BCG. The low expression tags cutoff of 100 counts per million was normalized for the number of genes in the THP-1 genome. Afterwards, we used InnateDB and pathway enrichment analysis to identify induced host pathways [52].

Sequencing depth simulation
Subsets of 50-bp sequencing reads, where both pairs aligned to the complete genome of *M. bovis* BCG, were taken randomly from dataset EGB and IF1ER. These subsets were of the same size (3.35 M) as the total number of sequencing reads from dataset IF2ER aligning to the *M. bovis* BCG genome in pairs (Figure 3.2A). From these two subsets, and from the paired-end sequencing reads of IF2ER aligning to the *M. bovis* BCG genome, random sets of reads were chosen of different sizes. For each different size, five random sets were generated. Differential gene expression for *M. bovis* BCG was determined using the method described above (data analysis pipeline) for every set. The mean and standard deviation of the number of differentially expressed genes was calculated (Figure 3.2B).

Compendium of expression data
A compendium containing 565 two-color microarrays for *M. tuberculosis* (strain H37Rv) was obtained from literature [73] and most of these (454) captured the effect of 75 drugs targeting metabolic pathways [74, 75] whereas 111 captured stress-induced dormancy in the wild-type and in DosR activation genes in KO mutants [76, 77].

Quantitative Real Time PCR
To verify the expression of several genes and small RNA's, their expression was determined using qRT-PCR. For 13 genes, we designed primers (Supplementary File 3.5) and quantified their relative expression to the reference gene: *rpoB* (BCG0716; Rv0667). This gene encodes the RNA polymerase β-subunit and is thought to be a housekeeping gene, which was used as such in previous qRT-PCR studies [78-80]. Fold changes in gene expression (Supplementary File 3.6) were calculated using the ΔΔCt, as described previously [81].

Availability of supporting data

Raw sequence read data supporting the results of this article are available in the EMBL-EBI European Nucleotide Archive under the Accession No. PRJEB6552, http://www.ebi.ac.uk/ena/data/view/PRJEB6552.
Supplementary Files can be accessed via: Rienk A. Rienksma, Maria Suárez-Diez, Hans-Joachim Mollenkopf, Gregory M. Dolganov, Anca Dorhoi, Gary K. Schoolnik, Vitor A.P. Martins dos Santos, Stefan H. E. Kaufmann, Peter J. Schaap, Martin Gengenbacher. Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. BMC Genomics 16 (2015).

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

References

1. WHO, *Global Tuberculosis Report 2013.* 2013, France: World Health Organization.

2. Gengenbacher, M. and S.H.E. Kaufmann, Mycobacterium tuberculosis: *success through dormancy.* FEMS Microbiology Reviews, 2012. **36**(3): p. 514-532.

3. Repasy, T., J. Lee, S. Marino, N. Martinez, D.E. Kirschner, G. Hendricks, S. Baker, A.A. Wilson, D.N. Kotton, and H. Kornfeld, *Intracellular bacillary burden reflects a burst size for* Mycobacterium tuberculosis in vivo. PLoS Pathog, 2013. **9**(2): p. e1003190.

4. Westermann, A.J., S.A. Gorski, and J. Vogel, *Dual RNA-seq of pathogen and host.* Nat Rev Micro, 2012. **10**(9): p. 618-630.

5. Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan, and G.K. Schoolnik, *Transcriptional Adaptation of* Mycobacterium tuberculosis *within Macrophages: Insights into the Phagosomal Environment.* J Exp Med, 2003. **198**(5): p. 693-704.

6. Rachman, H., M. Strong, U. Schaible, J. Schuchhardt, K. Hagens, H. Mollenkopf, D. Eisenberg, and S.H. Kaufmann, Mycobacterium tuberculosis *gene expression profiling within the context of protein networks.* Microbes Infect, 2006. **8**(3): p. 747-57.

7. Fontan, P., V. Aris, S. Ghanny, P. Soteropoulos, and I. Smith, *Global transcriptional profile of* Mycobacterium tuberculosis *during THP-1 human macrophage infection.* Infect Immun, 2008. **76**(2): p. 717-25.

8. Cappelli, G., E. Volpe, M. Grassi, B. Liseo, V. Colizzi, and F. Mariani, *Profiling of* Mycobacterium tuberculosis *gene expression during human macrophage infection: upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function.* Res Microbiol, 2006. **157**(5): p. 445-55.

9. Dubnau, E., J. Chan, V.P. Mohan, and I. Smith, *Responses of* Mycobacterium tuberculosis *to growth in the mouse lung.* Infect Immun, 2005. **73**(6): p. 3754-7.

10. Kuczynski, J., E.K. Costello, D.R. Nemergut, J. Zaneveld, C.L. Lauber, D. Knights, O. Koren, N. Fierer, S.T. Kelley, R.E. Ley, J.I. Gordon, and R. Knight, *Direct sequencing of the human microbiome readily reveals community differences.* Genome Biology, 2010. **11**(5).

11. Lamont, E.A., W.W. Xu, and S. Sreevatsan, *Host-*Mycobacterium avium *subsp.* paratuberculosis *interactome reveals a novel iron assimilation mechanism linked to nitric oxide stress during early infection.* BMC Genomics, 2013. **14**: p. 694.

12. Garnier, T., K. Eiglmeier, J.C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P.R. Wheeler, J. Parkhill, B.G. Barrell, S.T. Cole, S.V. Gordon, and R.G. Hewinson, *The complete genome sequence of* Mycobacterium bovis. Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7877-82.

3

13. Brosch, R., S.V. Gordon, T. Garnier, K. Eiglmeier, W. Frigui, P. Valenti, S. Dos Santos, S. Duthoy, C. Lacroix, C. Garcia-Pelayo, J.K. Inwald, P. Golby, J.N. Garcia, R.G. Hewinson, M.A. Behr, M.A. Quail, C. Churcher, B.G. Barrell, J. Parkhill, and S.T. Cole, *Genome plasticity of BCG and impact on vaccine efficacy.* Proc Natl Acad Sci U S A, 2007. **104**(13): p. 5596-601.

14. Cole, S.T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, and B.G. Barrell, *Deciphering the biology of* Mycobacterium tuberculosis *from the complete genome sequence.* Nature, 1998. **393**(6685): p. 537-544.

15. Mangan, J.A., I.M. Monahan, and P.D. Butcher, *Gene expression during host—pathogen interactions: Approaches to bacterial mRNA extraction and labelling for microarray analysis*, in *Methods in Microbiology*, N.D. Brendan Wren, Editor. 2002, Academic Press. p. 137-151.

16. Talaat, A.M., R. Lyons, S.T. Howard, and S.A. Johnston, *The temporal expression profile of* Mycobacterium tuberculosis *infection in mice.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(13): p. 4602-4607.

17. Oliver, H.F., R.H. Orsi, L. Ponnala, U. Keich, W. Wang, Q. Sun, S.W. Cartinhour, M.J. Filiatrault, M. Wiedmann, and K.J. Boor, *Deep RNA sequencing of* L. monocytogenes *reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs.* BMC Genomics, 2009. **10**: p. 641.

18. Perkins, T.T., R.A. Kingsley, M.C. Fookes, P.P. Gardner, K.D. James, L. Yu, S.A. Assefa, M. He, N.J. Croucher, D.J. Pickard, D.J. Maskell, J. Parkhill, J. Choudhary, N.R. Thomson, and G. Dougan, *A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus* Salmonella typhi. PLoS Genet, 2009. **5**(7): p. e1000569.

19. Yoder-Himes, D.R., P.S. Chain, Y. Zhu, O. Wurtzel, E.M. Rubin, J.M. Tiedje, and R. Sorek, *Mapping the* Burkholderia cenocepacia *niche response via high-throughput sequencing.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3976-81.

20. Ouellet, H., J.B. Johnston, and P.R. de Montellano, *Cholesterol catabolism as a therapeutic target in* Mycobacterium tuberculosis. Trends Microbiol, 2011. **19**(11): p. 530-9.

21. Brzostek, A., A. Rumijowska-Galewicz, B. Dziadek, E.A. Wojcik, and J. Dziadek, *ChoD and HsdD can be dispensable for cholesterol degradation in mycobacteria.* J Steroid Biochem Mol Biol, 2013. **134**: p. 1-7.

22. Casabon, I., A.M. Crowe, J. Liu, and L.D. Eltis, *FadD3 is an acyl-CoA synthetase that initiates catabolism of cholesterol rings C and D in actinobacteria.* Mol Microbiol, 2013. **87**(2): p. 269-83.

3

23. Thomas, S.T. and N.S. Sampson, Mycobacterium tuberculosis *utilizes a unique heterotetrameric structure for dehydrogenation of the cholesterol side chain.* Biochemistry, 2013. **52**(17): p. 2895-904.

24. Voskuil, M.I., Mycobacterium tuberculosis *cholesterol catabolism requires a new class of acyl coenzyme A dehydrogenase.* J Bacteriol, 2013. **195**(19): p. 4319-21.

25. Wipperman, M.F., M. Yang, S.T. Thomas, and N.S. Sampson, *Shrinking the FadE Proteome of* Mycobacterium tuberculosis: *Insights into Cholesterol Metabolism through Identification of an* $\alpha_2\beta_2$ *Heterotetrameric Acyl Coenzyme A Dehydrogenase Family.* Journal of Bacteriology, 2013. **195**(19): p. 4331-4341.

26. van der Geize, R., A.W. Grommen, G.I. Hessels, A.A. Jacobs, and L. Dijkhuizen, *The steroid catabolic pathway of the intracellular pathogen* Rhodococcus equi *is important for pathogenesis and a target for vaccine development.* PLoS Pathog, 2011. **7**(8): p. e1002181.

27. Driscoll, M.D., K.J. McLean, C. Levy, N. Mast, I.A. Pikuleva, P. Lafite, S.E. Rigby, D. Leys, and A.W. Munro, *Structural and biochemical characterization of* Mycobacterium tuberculosis *CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen.* J Biol Chem, 2010. **285**(49): p. 38270-82.

28. Wipperman, M.F., N.S. Sampson, and S.T. Thomas, *Pathogen roid rage: Cholesterol utilization by* Mycobacterium tuberculosis. Crit Rev Biochem Mol Biol, 2014.

29. Rienksma, R.A., M. Suarez-Diez, L. Spina, P.J. Schaap, and V.A.P. Martins dos Santos, *Systems-level modeling of mycobacterial metabolism for the identification of new (multi-)drug targets.* Seminars in Immunology, 2014. **26**(6): p. 610-622.

30. Pandey, A.K. and C.M. Sassetti, *Mycobacterial persistence requires the utilization of host cholesterol.* Proceedings of the National Academy of Sciences, 2008. **105**(11): p. 4376-4380.

31. de Souza, G.A., N.A. Leversen, H. Malen, and H.G. Wiker, *Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway.* J Proteomics, 2011. **75**(2): p. 502-10.

32. Kendall, S.L., P. Burgess, R. Balhana, M. Withers, A. Ten Bokum, J.S. Lott, C. Gao, I. Uhia-Castro, and N.G. Stoker, *Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2.* Microbiology, 2010. **156**(Pt 5): p. 1362-71.

33. Yellaboina, S., S. Ranjan, V. Vindal, and A. Ranjan, *Comparative analysis of iron regulated genes in mycobacteria.* FEBS Lett, 2006. **580**(11): p. 2567-76.

34. Griffin, J.E., A.K. Pandey, S.A. Gilmore, V. Mizrahi, J.D. McKinney, C.R. Bertozzi, and C.M. Sassetti, *Cholesterol catabolism by* Mycobacterium tuberculosis *requires transcriptional and metabolic adaptations.* Chem Biol, 2012. **19**(2): p. 218-27.

35. Marrakchi, H., M.-A. Lanéelle, and M. Daffé, *Mycolic Acids: Structures, Biosynthesis, and Beyond.* Chemistry & Biology, 2014. **21**(1): p. 67-85.

36. Gouzy, A., G. Larrouy-Maumus, T.-D. Wu, A. Peixoto, F. Levillain, G. Lugo-Villarino, J.-L. Gerquin-Kern, L.P.S. de Carvalho, Y. Poquet, and O. Neyrolles, Mycobacterium tuberculosis *nitrogen assimilation and host colonization require aspartate.* Nat Chem Biol, 2013. **9**(11): p. 674-676.

37. Gouzy, A., G. Larrouy-Maumus, D. Bottai, F. Levillain, A. Dumas, J.B. Wallach, I. Caire-Brandli, C. de Chastellier, T.D. Wu, R. Poincloux, R. Brosch, J.L. Guerquin-Kern, D. Schnappinger, L.P. Sorio de Carvalho, Y. Poquet, and O. Neyrolles, Mycobacterium tuberculosis *exploits asparagine to assimilate nitrogen and resist acid stress during infection.* PLoS Pathog, 2014. **10**(2): p. e1003928.

38. Gouzy, A., Y. Poquet, and O. Neyrolles, *A central role for aspartate in* Mycobacterium tuberculosis *physiology and virulence.* Front Cell Infect Microbiol, 2013. **3**: p. 68.

39. Vilcheze, C., B. Weinrick, K.W. Wong, B. Chen, and W.R. Jacobs, Jr., *NAD⁺ auxotrophy is bacteriocidal for the tubercle bacilli.* Mol Microbiol, 2010. **76**(2): p. 365-77.

40. Lee, W.L., B. Gold, C. Darby, N. Brot, X. Jiang, L.P. de Carvalho, D. Wellner, G. St John, W.R. Jacobs, Jr., and C. Nathan, Mycobacterium tuberculosis *expresses methionine sulphoxide reductases A and B that protect from killing by nitrite and hypochlorite.* Mol Microbiol, 2009. **71**(3): p. 583-93.

41. Weissbach, H., F. Etienne, T. Hoshi, S.H. Heinemann, W.T. Lowther, B. Matthews, G. St. John, C. Nathan, and N. Brot, *Peptide Methionine Sulfoxide Reductase: Structure, Mechanism of Action, and Biological Function.* Archives of Biochemistry and Biophysics, 2002. **397**(2): p. 172-178.

42. McMahon, M.D., J.S. Rush, and M.G. Thomas, *Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in* Mycobacterium tuberculosis. J Bacteriol, 2012. **194**(11): p. 2809-18.

43. Siegrist, M.S., M. Unnikrishnan, M.J. McConnell, M. Borowsky, T.Y. Cheng, N. Siddiqi, S.M. Fortune, D.B. Moody, and E.J. Rubin, *Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition.* Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18792-7.

44. Maciag, A., E. Dainese, G.M. Rodriguez, A. Milano, R. Provvedi, M.R. Pasca, I. Smith, G. Palu, G. Riccardi, and R. Manganelli, *Global analysis of the* Mycobacterium tuberculosis *Zur (FurB) regulon.* J Bacteriol, 2007. **189**(3): p. 730-40.

45. Bitter, W., E.N. Houben, D. Bottai, P. Brodin, E.J. Brown, J.S. Cox, K. Derbyshire, S.M. Fortune, L.Y. Gao, J. Liu, N.C. Gey van Pittius, A.S. Pym, E.J. Rubin, D.R. Sherman, S.T. Cole, and R. Brosch, *Systematic genetic nomenclature for type VII secretion systems.* PLoS Pathog, 2009. **5**(10): p. e1000507.

46. Wells, R.M., C.M. Jones, Z. Xi, A. Speer, O. Danilchanka, K.S. Doornbos, P. Sun, F. Wu, C. Tian, and M. Niederweis, *Discovery of a siderophore export system essential for virulence of* Mycobacterium tuberculosis. PLoS Pathog, 2013. **9**(1): p. e1003120.

47. Ryndak, M.B., S. Wang, I. Smith, and G.M. Rodriguez, *The* Mycobacterium tuberculosis *high-affinity iron importer, IrtA, contains an FAD-binding domain.* J Bacteriol, 2010. **192**(3): p. 861-9.

48. DiChiara, J.M., L.M. Contreras-Martinez, J. Livny, D. Smith, K.A. McDonough, and M. Belfort, *Multiple small RNAs identified in* Mycobacterium bovis *BCG are also expressed in* Mycobacterium tuberculosis *and* Mycobacterium smegmatis. Nucleic Acids Research, 2010. **38**(12): p. 4067-4078.

3

49. Tsai, C.-H., C. Baranowski, J. Livny, K.A. McDonough, J.T. Wade, and L.M. Contreras, *Identification of Novel sRNAs in Mycobacterial Species.* PLoS ONE, 2013. **8**(11): p. e79411.

50. Arnvig, K.B., I. Comas, N.R. Thomson, J. Houghton, H.I. Boshoff, N.J. Croucher, G. Rose, T.T. Perkins, J. Parkhill, G. Dougan, and D.B. Young, *Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of* Mycobacterium tuberculosis. PLoS Pathog, 2011. **7**(11): p. e1002342.

51. Arnvig, K. and D. Young, *Non-coding RNA and its potential role in* Mycobacterium tuberculosis *pathogenesis.* RNA Biol, 2012. **9**(4): p. 427-36.

52. Lynn, D.J., G.L. Winsor, C. Chan, N. Richard, M.R. Laird, A. Barsky, J.L. Gardy, F.M. Roche, T.H. Chan, N. Shah, R. Lo, M. Naseer, J. Que, M. Yau, M. Acab, D. Tulpan, M.D. Whiteside, A. Chikatamarla, B. Mah, T. Munzner, K. Hokamp, R.E. Hancock, and F.S. Brinkman, *InnateDB: facilitating systems-level analyses of the mammalian innate immune response.* Mol Syst Biol, 2008. **4**: p. 218.

53. de Weerd, N.A., S.A. Samarajiwa, and P.J. Hertzog, *Type I interferon receptors: biochemistry and biological functions.* J Biol Chem, 2007. **282**(28): p. 20053-7.

54. Shah, S., A. Bohsali, S.E. Ahlbrand, L. Srinivasan, V.A. Rathinam, S.N. Vogel, K.A. Fitzgerald, F.S. Sutterwala, and V. Briken, *Cutting edge:* Mycobacterium tuberculosis *but not nonvirulent mycobacteria inhibits IFN-β and AIM2 inflammasome-dependent IL-1β production via its ESX-1 secretion system.* J Immunol, 2013. **191**(7): p. 3514-8.

55. Singh, V., S. Jamwal, R. Jain, P. Verma, R. Gokhale, and Kanury V.S. Rao, Mycobacterium tuberculosis-*Driven Targeted Recalibration of Macrophage Lipid Homeostasis Promotes the Foamy Phenotype.* Cell Host & Microbe, 2012. **12**(5): p. 669-681.

56. Bottai, D., M. Di Luca, L. Majlessi, W. Frigui, R. Simeone, F. Sayes, W. Bitter, M.J. Brennan, C. Leclerc, G. Batoni, M. Campa, R. Brosch, and S. Esin, *Disruption of the ESX-5 system of* Mycobacterium tuberculosis *causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation.* Molecular Microbiology, 2012. **83**(6): p. 1195-1209.

57. Korf, J., A. Stoltz, J. Verschoor, P. De Baetselier, and J. Grooten, *The* Mycobacterium tuberculosis *cell wall component mycolic acid elicits pathogen-associated host innate immune responses.* Eur J Immunol, 2005. **35**(3): p. 890-900.

58. Silva-Gomes, S., R. Appelberg, R. Larsen, M.P. Soares, and M.S. Gomes, *Heme Catabolism by Heme Oxygenase-1 Confers Host Resistance to* Mycobacterium Infection. Infection and Immunity, 2013. **81**(7): p. 2536-2545.

59. Rindi, L., N. Lari, M.G. Gil, and C. Garzelli, *Oligo(dT)-Primed Synthesis of cDNA by Reverse Transcriptase in Mycobacteria.* Biochemical and Biophysical Research Communications, 1998. **248**(2): p. 216-218.

60. Sarkar, N., *Polyadenylation of mRNA in prokaryotes.* Annual Review of Biochemistry, 1997. **66**(1): p. 173-197.

61. Mohanty, B.K. and S.R. Kushner, *Bacterial/archaeal/organellar polyadenylation.* Wiley Interdiscip Rev RNA, 2011. **2**(2): p. 256-76.

62. Chang, J.C., M.D. Miner, A.K. Pandey, W.P. Gill, N.S. Harik, C.M. Sassetti, and D.R. Sherman, *igr Genes and* Mycobacterium tuberculosis *cholesterol metabolism.* J Bacteriol, 2009. **191**(16): p. 5232-9.

63. Yam, K.C., I. D'Angelo, R. Kalscheuer, H. Zhu, J.X. Wang, V. Snieckus, L.H. Ly, P.J. Converse, W.R. Jacobs, Jr., N. Strynadka, and L.D. Eltis, *Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of* Mycobacterium tuberculosis. PLoS Pathog, 2009. **5**(3): p. e1000344.

64. Ouellet, H., S. Guan, J.B. Johnston, E.D. Chow, P.M. Kells, A.L. Burlingame, J.S. Cox, L.M. Podust, and P.R. de Montellano, Mycobacterium tuberculosis *CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one.* Mol Microbiol, 2010. **77**(3): p. 730-42.

65. Petrusma, M., L. Dijkhuizen, and R. van der Geize, *Structural features in the KshA terminal oxygenase protein that determine substrate preference of 3-ketosteroid 9α-hydroxylase enzymes.* J Bacteriol, 2012. **194**(1): p. 115-21.

66. Rodríguez, J.G., A.C. Hernández, C. Helguera-Repetto, D. Aguilar Ayala, R. Guadarrama-Medina, J.M. Anzóla, J.R. Bustos, M.M. Zambrano, J. González-y-Merchand, M.J. García, and P. Del Portillo, *Global Adaptation to a Lipid Environment Triggers the Dormancy-Related Phenotype of* Mycobacterium tuberculosis. mBio, 2014. **5**(3).

67. Dietrich, G., U.E. Schaible, K.-D. Diehl, H.-J. Mollenkopf, S. Wiek, J. Hess, K. Hagens, S.H.E. Kaufmann, and B. Knapp, *Isolation of RNA from mycobacteria grown under* in vitro *and* in vivo *conditions.* FEMS Microbiology Letters, 2000. **186**(2): p. 177-180.

68. Peano, C., A. Pietrelli, C. Consolandi, E. Rossi, L. Petiti, L. Tagliabue, G. De Bellis, and P. Landini, *An efficient rRNA removal method for RNA sequencing in GC-rich bacteria.* Microbial Informatics and Experimentation, 2013. **3**(1): p. 1.

69. Giannoukos, G., D. Ciulla, K. Huang, B. Haas, J. Izard, J. Levin, J. Livny, A. Earl, D. Gevers, D. Ward, C. Nusbaum, B. Birren, and A. Gnirke, *Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes.* Genome Biology, 2012. **13**(3): p. r23.

70. Li, L., D. Huang, M.K. Cheung, W. Nong, Q. Huang, and H.S. Kwan, *BSRD: a repository for bacterial small regulatory RNA.* Nucleic Acids Res, 2013. **41**(Database issue): p. D233-8.

71. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.

72. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.

73. van Dam, J.C.J., P.J. Schaap, V.A.P. Martins dos Santos, and M. Suarez-Diez, *Integration of heterogeneous molecular networks to unravel gene-regulation in* Mycobacterium tuberculosis. BMC Syst Biol, 2014. **8**(111).

Comprehensive insights into transcriptional adaptation of intracellular
mycobacteria by microbe-enriched dual RNA sequencing

3

74. Boshoff, H.I., T.G. Myers, B.R. Copp, M.R. McNeil, M.A. Wilson, and C.E. Barry, 3rd, *The transcriptional responses of* Mycobacterium tuberculosis *to inhibitors of metabolism: novel insights into drug mechanisms of action.* J Biol Chem, 2004. **279**(38): p. 40174-84.

75. Karakousis, P.C., E.P. Williams, and W.R. Bishai, *Altered expression of isoniazid-regulated genes in drug-treated dormant* Mycobacterium tuberculosis*.* J Antimicrob Chemother, 2008. **61**(2): p. 323-31.

76. Deb, C., C.M. Lee, V.S. Dubey, J. Daniel, B. Abomoelak, T.D. Sirakova, S. Pawar, L. Rogers, and P.E. Kolattukudy, *A novel* in vitro *multiple-stress dormancy model for* Mycobacterium tuberculosis *generates a lipid-loaded, drug-tolerant, dormant pathogen.* PLoS One, 2009. **4**(6): p. e6077.

77. Honaker, R.W., R.L. Leistikow, I.L. Bartek, and M.I. Voskuil, *Unique roles of DosT and DosS in DosR regulon induction and* Mycobacterium tuberculosis *dormancy.* Infect Immun, 2009. **77**(8): p. 3258-63.

78. Adékambi, T., M. Drancourt, and D. Raoult, *The* rpoB *gene as a tool for clinical microbiologists.* Trends in Microbiology, 2009. **17**(1): p. 37-45.

79. Badejo, A.C., A.O. Badejo, K.H. Shin, and Y.G. Chai, *A Gene Expression Study of the Activities of Aromatic Ring-Cleavage Dioxygenases in* Mycobacterium gilvum *PYR-GCK to Changes in Salinity and pH during Pyrene Degradation.* PLoS ONE, 2013. **8**(2): p. e58066.

80. Lim, S.Y., B.J. Kim, M.K. Lee, and K. Kim, *Development of a real-time PCR-based method for rapid differential identification of Mycobacterium species.* Letters in Applied Microbiology, 2008. **46**(1): p. 101-106.

81. Yuan, J., A. Reed, F. Chen, and C.N. Stewart, *Statistical analysis of real-time PCR data.* BMC Bioinformatics, 2006. **7**(1): p. 85
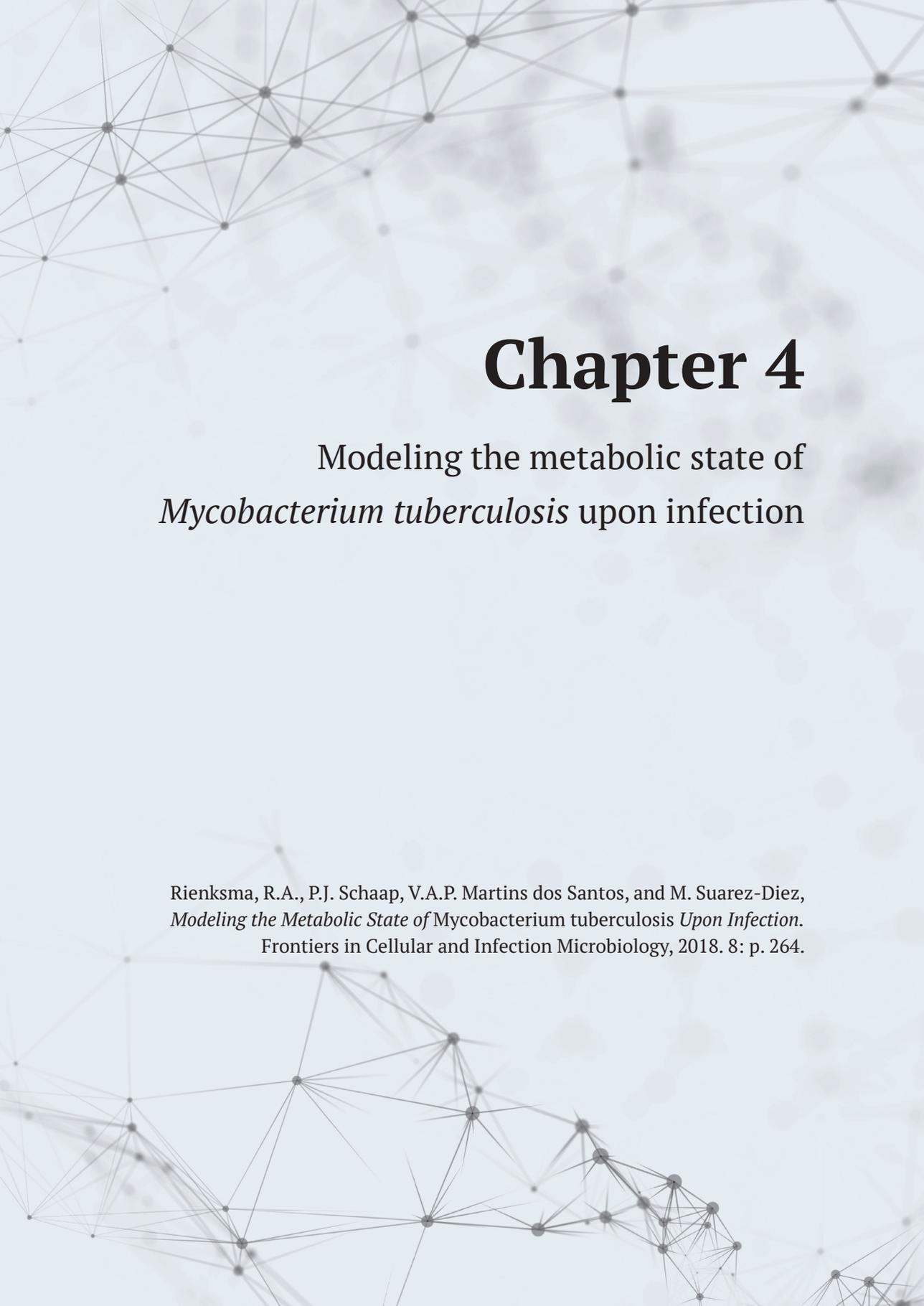
Chapter 4

Modeling the metabolic state of
Mycobacterium tuberculosis upon infection

Abstract

Genome-scale metabolic models of *Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis, have been envisioned as a platform for drug discovery. By systematically probing the networks that underpin such models, the reactions that are essential for Mtb are identified. A majority of these reactions are catalyzed by enzymes and thus represent candidate drug targets to fight an Mtb infection. Nevertheless, this is complicated by the limited knowledge on the environment that Mtb encounters during infection.

Modeling the behavior of the bacteria during infection requires knowledge of the so-called biomass reaction that represents bacterial biomass composition. This composition varies in different environments or bacterial growth phases. Accurate modeling of the metabolic state requires a precise biomass reaction for the described condition. In recent years, additional insights in the in-host environment occupied by Mtb have been gained as transcript abundance data of interacting host and pathogen have become available. Therefore, we used transcript abundance data and developed a straightforward and systematic method to obtain a condition-specific biomass reaction for Mtb during *in vitro* growth and during infection of its host. The method described herein is virtually free of any pre-set assumptions on uptake rates of nutrients, making it suitable for exploring environments with limited accessibility. The condition-specific biomass reaction represents the 'metabolic objective' of Mtb in a given environment (in-host growth and growth on defined medium) at a specific time point, and as such allows modeling the bacterial metabolic state in these environments.

Five different biomass reactions were used predict nutrient uptake rates and gene essentiality. Predictions were subsequently compared to available experimental data. Our results show that nutrient uptake can accurately be predicted. Gene essentiality can also be predicted but accurate predictions remain difficult to obtain. In conclusion, a viable strategy to model Mtb metabolism in hard-to-access environments that is virtually free of pre-set assumptions is provided.

Keywords: metabolic model, *Mycobacterium tuberculosis*, systems biology, host-pathogen interaction, condition specific, flux balance analysis.

Background

Constraint-based genome-scale metabolic models (GSMs) enable prediction of metabolic states. A metabolic state is defined as a vector of all fluxes or conversion rates (in mmol h^{-1}) throughout metabolism per weight unit of biomass (usually 1 gram dry weight, gDw). GSMs comprise linear equations describing conversions among metabolites, uptake or secretion processes, and transport processes over different compartments. These equations are referred to as flux balance constraints and are founded on an underlying metabolic network wherein all metabolites are interconnected by conversion and transport reactions. The flux balance constraints are captured in a stoichiometric matrix [1]. GSMs may comprise additional constraints as well, such as reversibility and capacity constraints. The whole of all possible fluxes that satisfy all constraints of a GSM is referred to as the solution space [2]. Additional constraints present an opportunity to further limit the size of the solution space, which results in a more accurate calculation of the metabolic state. A suitable way to increase the amount of constraints is to measure uptake and/or secretion rates of metabolites/nutrients. Knowledge of a few of these rates can considerable shrink the solution space [3].

Given the stoichiometric matrix, the most straightforward approach for calculating a metabolic state is to simulate conditions on which the organism is in a steady state physiological condition, meaning that there is no net intracellular accumulation of metabolites. Under this assumption, it is possible to construct a Flux Balance Analysis (FBA) problem. FBA finds the optimal (maximum or minimum) value of a selected function, the so-called objective function, while satisfying all constraints. Solution of the FBA problem leads to a vector of reaction fluxes that represents a calculated metabolic state of the organism. This calculated metabolic state is more likely to represent the actual metabolic state as the solution space is shrunk by additional constraints [4, 5].

The metabolic state is, among others, dependent on the objective function. Metabolic states have been accurately predicted for several bacteria in recent years, using objective functions such as maximizing the flux through the biomass reaction to represent growth rate, maximizing ATP production or minimizing enzyme usage among others [6].

However, in some conditions measuring uptake and/or secretion rates can be notoriously difficult, if not impossible. Such is the case for intracellular *Mycobacterium tuberculosis* (Mtb), a pathogenic bacterium able to withstand the harsh environment of the phagosome. Mtb is even capable of halting the maturation of the phagosome inside immune cells and providing a niche for the bacterium to thrive [7, 8]. Genome-scale metabolic models of Mtb, have been envisioned as a platform for drug discovery [9, 10].

In addition to uptake rates, other measurements can serve to estimate or approach (a part of) the metabolic state of a cell, such as transcript profiles [11]. For Mtb, a major difficulty with these measurements is the large size difference between the eukaryotic host cell and the prokaryotic pathogen, which results in metabolites and transcripts from the host vastly outnumbering those of the pathogen [12]. With regard to transcript abundance experimental methods have been developed to increase the ratio of pathogen mRNA to host mRNA [13]. This enrichment in pathogen transcripts renders differences between intracellular and extracellular pathogen transcript abundance apparent. We recently published a dataset of *Mycobacterium bovis* BCG and THP-1 cells using a dual RNA-sequencing strategy [14]. However, such an enrichment method is not available for metabolites, which are more closely related to fluxes as compared to transcripts. Moreover, metabolites, unlike transcripts, cannot be assigned to host or pathogen unless they only occur in one of said host or pathogen [15].

Transcript abundance data can be used to constrain models in environments where knowledge regarding nutrient availability and objective(s) is limited. Methods such as iMAT [16], MADE [17], GIMME [18], E-flux [19], TRFBA [20] and others [21] limit the solution space by using expression values as a proxy for flux. These methods allow for the explanation of phenomena that cannot be derived solely from the models, such as the prediction of the Crabtree effect in yeast [22]. These model and data integration methods limit the solution space within the ranges of expression data, thereby effectively generating condition-specific models. Shrinking the solution space by limiting fluxes based on gene expression seems an obvious choice, but it is not at all obvious how this should be done. Methods for model and data integration have been thoroughly evaluated [23]. The evaluation showed that no method outperforms the others for all tested models and datasets. Finally, this condition-specific model building can hamper exploration of metabolic states that arise from perturbations of the environment, from which the gene expression data was originally derived. These adapted models would only allow changes to the metabolic state that fit within the boundaries of what was originally measured. Such a rigid model appears a poor choice for predictive modeling.

A modeling approach focused on an accurate description of the objective of Mtb during infection appears to be a better strategy to make new predictions because it does not limit the solution space or metabolic flexibility beforehand. Previous approaches have relied on adapting the biomass reaction to represent the composition on mycobacterial cells during infection. Bordbar and colleagues adjusted the biomass reaction based on differential gene expression [24]. This approach is biased by the biomass reaction that is present in the model prior to the tailoring process and the potential synthesis of other metabolites specifically during infection, is overlooked. Shi and colleagues [25] proposed a biomass reaction comprising trehalose dimycolate,

4

triacylglycerol (TAG) and polyglutamate/glutamine to reflect a minimal cell wall composition. The logical assumption applied was that during a 'non-growth state', Mtb utilizes metabolites produced in pathways of which gene expression is elevated and does not, or to a lesser extent, utilize metabolites produced in pathways of which gene expression is suppressed. Shi and colleagues used qPCR to monitor gene expression [25]. This requires a pre-selection of target genes based on experience and experimental output and does not accommodate unbiased exploration of the transcriptional landscape.

Here, we integrate a constraint-based (CB) model of Mtb metabolism and RNA sequencing data to provide condition-specific biomass reactions during host infection and during growth on Middlebrook 7H9 medium. The genome-scale nature of this approach ensures all known pathways and biomass precursors are taken into account, whereas the nature of the used data (RNAseq) ensures unbiased assumptions on the types and quantities of metabolic precursors. During on-going infection mycobacterial cells might enter a non-growth state on which maximal growth rate is not the metabolic objective. However, still minimal macromolecular components need to be synthesized and energetic requirements need to be fulfilled to ensure survival. The condition-specific biomass reaction representing infection combines both aspects as it reflects the composition of mycobacterial cells during infection, and it also represents the metabolic requirements for its survival and interaction with the host which are incorporated in the RNAseq data as well. To simulate the metabolic state of the bacteria upon infection, flux through the condition-specific biomass reaction is maximized, while the total usage of enzymes is minimized. As Mtb faces several types of stress and adverse conditions imposed by the host's immune system during infection of the host [26], it is assumed that Mtb does not squander its resources, and makes optimal use of available enzymes. From a modeling perspective, this can be seen as a bi-objective optimization problem wherein two competing objectives, i.e. maximization of biomass production on the one hand, and minimization of enzyme usage on the other hand, are simultaneously considered.

The goal of multi-objective optimizations is to find Pareto optimal solutions (also called non-dominated solutions) [27]. A solution is Pareto optimal if no other solution exists that better satisfies all objectives. In other words, a solution is Pareto optimal if an improvement in one objective requires a degradation of another. Multiple methods have been developed to obtain Pareto optimal solutions in multi-objective optimization problems such as the normal constraint method [28] that has been used to explore tradeoffs between hepatic metabolic functions [29]. Here, we tackle the problem by using a weighted sum method in which weight factors are attributed to each objective: f_b and $f_{e,i}$, for biomass and enzyme usage for each of the $i = \{1, \dots , m\}$ reactions, respectively. This approach ensures that the obtained solutions are Pareto optimal [30].

Parsimonious enzyme usage FBA (pFBA) has been proposed to explore the tradeoffs between maximizing growth and minimizing enzyme utilization [24]. In pFBA there is an initial maximization of the biomass reaction followed by a minimization of enzyme usage. In our approach, this Pareto optimal solution would correspond to the extreme case wherein the weight of the biomass objective is much higher than that of the enzyme minimization objectives ($f_b \gg f_{e,i}$). By changing these weight factors, a ratio between these two factors, f_r, is established that enables the prediction of metabolic genes essential to Mtb within the macrophage as well as metabolites that are sequestered by Mtb from the phagosome. Comparison of these predictions with experimentally obtained data [31, 32] reveals that by using a condition-specific objective function inferred from transcript abundance data the metabolic state of Mtb upon infection can be accurately predicted.

Results

We created two condition-specific biomass reactions (CSI and CSM) based on transcript abundance data in two conditions. The term 'biomass reaction' is perhaps not the most suitable term as these reactions not only cover synthesis of metabolites used for biomass production, but also synthesis of excreted enzymes and small molecules, repair of damaged lipid membranes and other metabolites involved in host-pathogen interaction. Even though these processes themselves are largely unknown, transcript abundance data indirectly reflects these processes and combined with a GSM can give a picture of required metabolic precursors for these processes.

Condition-specific biomass reactions

The creation of a condition-specific biomass reaction requires a CB model, a list of available nutrients in the given condition, a list of metabolic precursors for synthesis of macromolecules, and transcript abundance data. We used model sMtb, a comprehensive model of Mtb metabolism [10], with minor corrections and additions (Supplementary Files 4.1, 4.2 and 4.3). Transcript abundance data was obtained from a dual RNA-sequencing experiment wherein transcript abundances of *Mycobacterium bovis* BCG, a close relative of Mtb having a highly similar genome [33], were measured under two conditions [14]. In the first condition *M. bovis* BCG infects THP-1 cells, and in the second condition *M. bovis* BCG grows on Middlebrook 7H9 medium. The sMtb model was used as a platform to integrate the expression data and to calculate two condition-specific biomass reactions of Mtb, CSI (condition-specific infection) and CSM (condition-specific medium), for both aforementioned conditions, respectively. A list of all metabolites known or expected to be present in the phagosome was assembled (Supplementary File 4.4). Availability of these metabolites was simulated

by enabling their free uptake in the model. In addition, a list of all known biomass precursors was generated based on the sMtb model (Supplementary File 4.5).
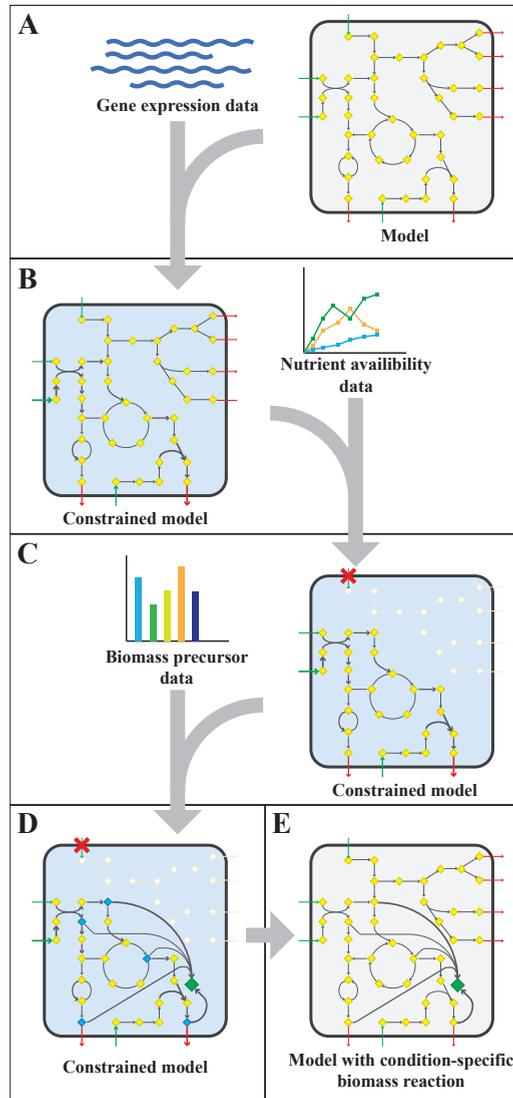
Figure 4.1 - Workflow to create a model with a condition-specific biomass reaction.
(A) An exemplary constraint-based genome-scale metabolic model (GEM) comprising a metabolic network with metabolites (yellow diamonds) and reactions (arrows), including uptake reactions (green arrows) and secretion reactions (red arrows), is depicted in a microorganism (rounded square). Gene expression data (blue wave-shaped lines) is used to constrain maximal flux values according to the E-flux algorithm [19], to obtain a condition-specific GEM having a shrunken solution space. (B) The condition-specific (blue background) GEM is subsequently combined with nutrient availability data (graph) and uptake of unavailable nutrients is constrained to zero. (C) Biomass precursor data (bar plot) is used to pinpoint biomass precursors in the condition-specific GEM with blocked transport reactions (red crosses), and the flux through the flux limiting reaction for each precursor is selected by maximizing flux towards each biomass precursor (blue diamonds) individually. (D) The sum of all precursor fluxes is normalized to one gram biomass dry weight (1 gDw) and a condition-specific biomass reaction (green diamond) is obtained. (E) All constraints placed on the GEM in the previous steps, A to D, are removed and a GEM with a condition-specific biomass reaction is obtained.

The flux towards each biomass precursor was maximized one by one, while limiting the maximum flux through enzymatically-catalyzed reactions based on the transcript abundance for the present condition (Figure 4.1). The ratio of biomass precursors obtained for both conditions represents the two condition-specific biomass reactions (CSM and CSI). The contributions of each class of precursors to these two biomass reactions are shown in Table 4.1 (see Supplementary File 4.5 for a more detailed breakdown). The largest differences in the biomass reactions of both conditions entails the fraction of amino acids, which is approximately doubled in the host as compared to *in vitro* growth on Middlebrook 7H9 medium, which is in accordance with previous predictions [15]. The fraction of carbohydrates on the other hand, is substantially reduced from 20.1% to 9.9%.

Table 4.1 - Composition of the condition-specific biomass reactions.

	Weight percentage (g/gDw)	
	Condition-specific infection, CSI	**Condition-specific medium, CSM**[*]
Amino Acids	33.2	16.1
Nucleic Acids	7.6	8.5
Carbohydrates	9.9	20.1
Lipids	32.5	39.0
Other	16.8	16.2

*Note that due to rounding of the percentages, the total may not add up to 100%.

Simulating Mtb metabolism: balance between growth and enzyme utilization

To predict the *in vivo* metabolic state, reflecting Mtb's intracellular behavior, we compared the performance of five different biomass reactions: the *in vitro* biomass growth reaction (IVB) and a regular biomass growth reaction (REB), both present in sMtb [10], a biomass reaction representing non-replicating cells (NRC) [25], the condition-specific biomass reaction representing growth on Middlebrook 7H9 medium (CSM) and the condition-specific biomass reaction representing growth within the host's phagosome (CSI).

Simulation of the metabolic state of Mtb in the phagosome is complicated by a lack of knowledge on the rate at which nutrients are acquired from the host. However, various studies have shown that the phagosomal environment is likely to be hypoxic [34]. Therefore, we chose to limit the oxygen uptake rate at a relatively low value of 0.01 mmol gDW^{-1} h^{-1} while keeping unrestricted the uptakes of all other nutrients that were assumed to be present in the host. Even with such a restriction, nutrients were predicted to be taken up in unrealistically large quantities. This behavior can be traced back to anaerobic reactions in the model that result in ATP generation, followed by the artificial generation of oxygen at the cost of high amounts of energy in the form of ATP to ADP conversion. In addition, limiting the

oxygen uptake rate all the way to 0 mmol gDW^{-1} h^{-1} resulted in zero flux through the (condition-specific) biomass reaction, and was therefore an unsuitable strategy as well.

To overcome such difficulty, the assumption was made that Mtb utilizes its resources parsimoniously when in a hostile environment. This can be modeled by minimization of enzyme usage while maximizing the flux through the biomass reaction. This bi-objective optimization was performed using a weighted sum method in which the following FBA problem with a weighted objective was solved:

$$w = \max\left\{\left(\sum_{i=1}^{n} -f_{e,i} \cdot |v_{e,i}|\right) + f_b \cdot v_b\right\} \tag{4.1.1}$$

subject to:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b} \tag{4.1.2}$$
$$\mathbf{l} \le \mathbf{v} \le \mathbf{u} \tag{4.1.3}$$

Wherein w is the objective function value, $v_{e,i}$ represents the flux or rate of reaction i catalyzed by at least one enzyme; $f_{e,i}$ represents the weight factor for reaction i; v_b represents the specific growth rate (or biomass reaction flux value), i.e. the flux through one of the five aforementioned (condition-specific) biomass reactions; f_b represents the weight factor for the biomass reaction; n is the total number of reactions catalyzed by at least one enzyme; $\mathbf{S}$ represents the stoichiometric matrix; $\mathbf{v}$ represents a vector with all fluxes (comprising $v_{e,i}$ and v_b); $\mathbf{b}$ represents a vector with zeros; $\mathbf{l}$ represents a vector with lower bounds for all fluxes and $\mathbf{u}$ represents a vector with upper bounds for all fluxes. The weight factor ratio, f_r, between growth and total enzyme utilization is given by:

$$f_r = \frac{f_b}{\sum_{i=1}^{n} \cdot f_{e,i}} \tag{4.2}$$

Each reaction in the model catalyzed by one or multiple enzymes was given the same weight factor (f_e) and the weight factor (f_b) of the (condition-specific) biomass reaction was varied such that log(f_r) varied around a value of 0. A log(f_r) value of 0 entails that the numerator and denominator of equation 4.2 are of equal size and reflects a balanced weight distribution between minimization of enzyme usage (i.e. maximization of the negative values) and maximization of growth. By changing the weight factor ratio, the relative importance of enzyme usage minimization and biomass reaction maximization changes (Figure 4.2). If too much weight is put on the minimization of enzyme usage, i.e. f_r becomes too low, the biomass reaction flux value, v_b, becomes irrelevant and its value drops to zero, this can be seen at the left hand panel of Figure 4.2, where the graphs equal zero. The reason that the

average flux through enzymatically catalyzed reactions, v_e, does not drop to zero when too much emphasis is put on enzyme usage minimization, as can be seen in the middle panel of Figure 4.2, is because there is a small (0.1 mmol gDW^{-1} h^{-1}) growth related maintenance coefficient enforcing a small minimum flux of ATP to ADP conversion.
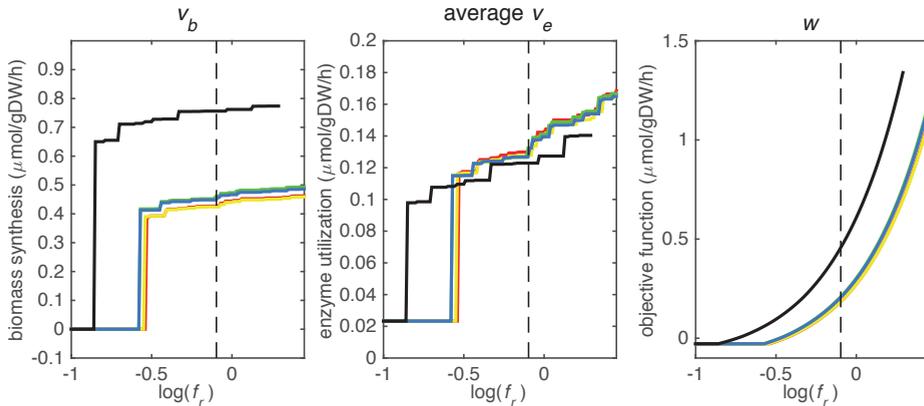

Figure 4.2 - Tradeoff between biomass production (growth rate) and enzyme utilization in the metabolic model.
Predicted values of the flux through the biomass synthesis reaction (left), average flux through all enzymatically catalyzed reactions (middle) and the objective function value (right), i.e. combination of total enzymatic reaction minimization and biomass reaction maximization for various f_r values. Five different biomass reactions are shown: CSI (green), CSM (blue), IVB (yellow), REB (red), and NRC (black). The dashed line indicates $f_r = 0.8$.

Prediction of uptake rates

Figures 4.3-4.5 show predicted uptake rates for the five different biomass reactions. As f_r increases, unrealistically high uptake rates are predicted to overcome the restrictions of the oxygen uptake threshold (0.01 mmol gDW^{-1} h^{-1}). As can be seen in Figure 4.2 (black line), the graph representing NRC biomass reaction (non-replicating cells) is slightly shifted as compared to the other objectives. The reason for this is that the total molecular weight of biomass precursors for this objective as obtained form Shi and colleagues is not normalized to one gram. Its value is actually higher, resulting in a larger objective function value at a smaller f_r value (Figure 4.2, right panel). For the four other biomass reactions a balance exists between maximization of growth and minimization of enzyme usage between approximately $\log(f_r) = -0.5$ and $\log(f_r) = 0.3$ ($0.3 \leqslant f_r \leqslant 2.0$). Beyond $\log(f_r) = 0.3$ the restrictions of the oxygen uptake threshold are overcome, and v_b and v_e values jump to infinite (for the NRC biomass reaction, this point is reached earlier). An appropriate value for f_r was selected from Figures 4.2-4.5 based on the consideration that uptake of asparagine, alanine and glutamate in addition to glycerol-3-phosphate and

CO_2 from the host is likely to occur during infection [31]. In addition, nitric oxide is not produced in high amounts by THP-1 cells, and thus not a likely source of nutrition [35], further justifying an f_r value greater than 0.3 ($\log(f_r) >$ -0.5), when hardly any nitric oxide is predicted to be taken up (Figure 4.5, right panel). At f_r = 0.8 ($\log(f_r)$ = -0.1, dashed vertical lines), uptake of glutamate and glycerol-3-phosphate is predicted for all biomass reactions except for NRC, the biomass reaction describing non-replicating cells. For this biomass reaction uptake of glutamate is not predicted. In addition, at this point (f_r = 0.8) uptake of asparagine is predicted for the condition specific biomass reaction of infection (CSI) and predicted to be likely (the average of minimum and maximum uptake rates is above zero) for the other four objectives. The uptake of alanine at this point is predicted to be likely for all five objective functions.
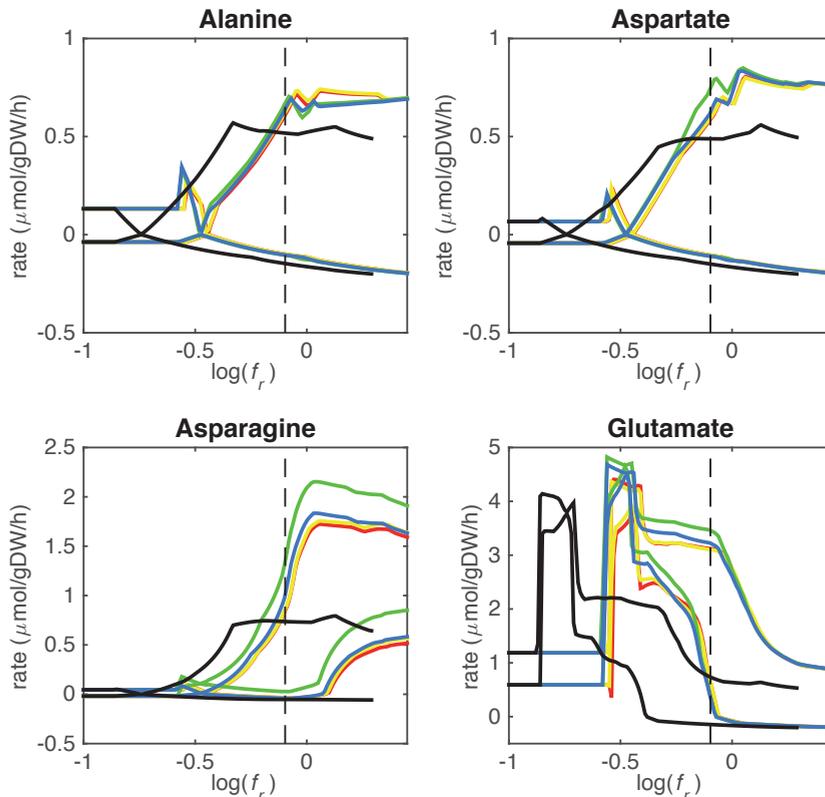
Figure 4.3 - Predicted amino acid uptake rates.
Maximum and minimum predicted uptake rates for alanine, aspartate, asparagine and glutamate using five different biomass reactions: CSI (green), CSM (blue), IVB (yellow), REB (red), and NRC (black) for varying f_r values. The dashed line indicates f_r = 0.8. Two lines of the same color indicate upper and lower limits of the prediction. Note that negative values of uptake rates denote excretion of that metabolite.
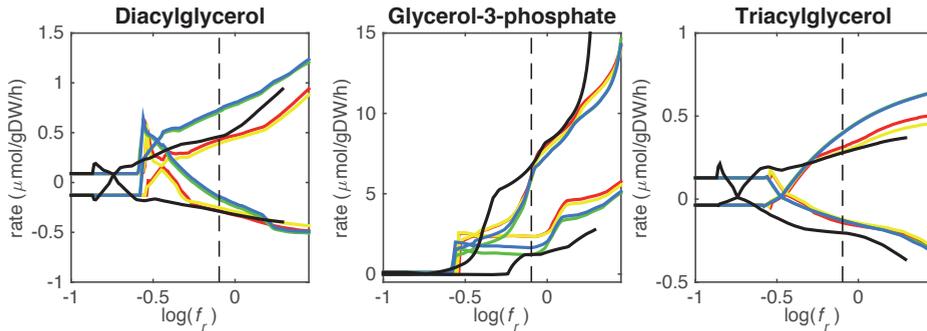
Figure 4.4 - Predicted lipid uptake rates.
Maximum and minimum predicted uptake rates for diacylglycerol, glycerol-3-phosphate, and triacylglycerol using five different biomass reactions: CSI (green), CSM (blue), IVB (yellow), REB (red), and NRC (black) for varying f_r values. The dashed line indicates $f_r = 0.8$. Two lines of the same color indicate upper and lower limits of the prediction. Note that negative values of uptake rates denote excretion of that metabolite.

Figure 4.5 - Predicted oxygen, carbon dioxide, and nitric oxide uptake rates.
Maximum and minimum predicted uptake rates for oxygen, carbon dioxide, and nitric oxide using five different biomass reactions: CSI (green), CSM (blue), IVB (yellow), REB (red), and NRC (black) for varying f_r values. The dashed line indicates $f_r = 0.8$. Two lines of the same color indicate upper and lower limits of the prediction. Note that negative values of uptake rates denote excretion of that metabolite.
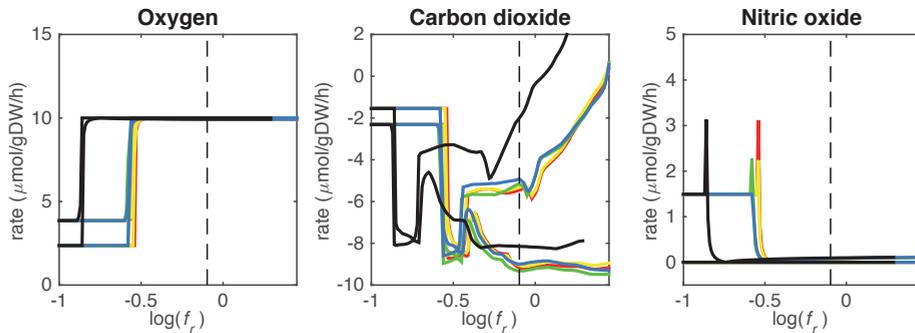
As can be seen in Figures 4.3-4.5, the predicted uptake rates are very similar for all five biomass reactions. Therefore, the biomass reaction itself seems of minor importance for the prediction of uptake rates. The uptake of glutamate appears as especially high for a relatively small f_r value, regardless of the chosen biomass reaction.

Beste and colleagues determined that the amino acids asparagine, alanine and glutamate are likely taken up during infection. Acetate- or acetyl-CoA-derived from β-oxidation of host lipids and CO_2 is utilized intracellularly and glycerol-3-phosphate could be a potential carbon source as well [31]. Regardless of the objective used, sMtb is able to reproduce these observations (Figures 4.3-4.5). In general, glutamate is taken up at low f_r values, while asparagine becomes more important at higher f_r values. The routes of glutamate towards most metabolic

precursors are shorter than those of asparagine, which is predicted to be taken up at a higher f_r value. In this way the change of the uptake rates with the f_r value reflects the metabolic versatility of each component.

Lipid uptake rates show that glycerol-3-phosphate is likely to be taken up, while diacylglycerol and triacylglycerol are possibly taken up. Cholesterol is not predicted to be used as a carbon source at any f_r value, in contrast to mounting evidence that cholesterol plays an important role as a nutrient for Mtb in the host [14, 36]. Currently, the cholesterol degradation pathway of Mtb is partly unknown, therefore only a partial degradation pathway exists in sMtb and the double ringed product (ring C and D of the cholesterol molecule) can only be excreted in sMtb. Partial degradation results in suboptimal yield of energy carrying metabolites derived from the cholesterol molecule compared to other molecules and therefore it is not predicted to be taken up. As knowledge on the cholesterol degradation pathway advances, the complete pathway will eventually be known. Integrating this complete pathway into sMtb will likely yield different results regarding cholesterol uptake.

The prediction of CO_2 uptake is complicated, as it is a nutrient that is excreted and possibly taken up, unlike the other nutrients in Figures 4.3-4.5. With FBA only a prediction of the difference between CO_2 excretion and uptake can be obtained. On average, CO_2 is predicted to be excreted throughout the entire f_r range.

Gene essentiality within the host

Gene essentiality predictions are often used to assess the predictive power of GSMs. Gene essentiality predictions can be simulated with *in silico* gene knock out (KO) mutants and comparing the maximal predicted growth rate of the wild type strain with the KO mutant. A reduction in the predicted specific growth rate of 95% or more is generally accepted as a threshold value for gene essentiality [9, 10, 37]. Here this approach will not provide satisfactory results, as there are too few constraints on the uptake rates of individual nutrients, only on the whole of enzymatically catalyzed reactions, resulting in an excess of unrealistic metabolic routes that could circumvent the deficiency caused by the deletion of the gene. We therefore optimized the aforementioned weighted bi-objective using $f_{e,i} = 0.001$ for all i with and without deleting the corresponding gene. Afterwards, both results were compared and a reduction of the specific growth rate, v_b, by 95% was marked as an essential gene.

These gene essentiality predictions were performed for each of the biomass reactions. We subsequently compared these predictions with experimental data obtained by Mendum and colleagues [32] and the accuracy, sensitivity and specificity of the predictions obtained with each of the five biomass reactions was calculated (Table 4.2).

Table 4.2 - Gene essentiality predictions
Gene essentiality predictions made using sMtb with five objective functions compared with experimental data obtained 3 and 7 days after infection [32].
Abbreviations: TP, true positive, TN, true negative, FP, false positive, FN, false negative, CSI, condition-specific infection reaction, CSM, condition-specific medium reaction, IVB, *in vitro* biomass reaction, REB, regular biomass reaction, NRC, non-replicating cells reaction.

	CSI		CSM		IVB		REB		NRC	
	3ds	**7ds**	**3ds**	**7ds**	**3ds**	**7ds**	**3ds**	**7ds**	**3ds**	**7ds**
TP	47	50	45	47	24	29	45	48	9	10
TN	335	346	343	353	352	365	346	357	419	428
FP	100	97	92	90	83	78	89	86	16	15
FN	222	211	224	214	245	232	224	213	260	251
Accuracy	0.54	0.56	0.55	0.57	0.53	0.56	0.56	0.58	0.61	0.62
Sensitivity	0.17	0.19	0.17	0.18	0.09	0.11	0.17	0.18	0.03	0.04
Specificity	0.77	0.78	0.79	0.80	0.81	0.82	0.80	0.81	0.96	0.97

Discussion

We have created condition-specific biomass reactions based on transcript abundance data, thereby ensuring that the obtained biomass compositions represent the organism's needs in the corresponding conditions. By limiting the availability of nutrients to those known or estimated to be present in the phagosome and restricting the uptake of all other nutrients, we were able to capture the metabolic state of Mtb during infection.

Methods such as iMAT [16], MADE [17] or GIMME[18], aim at developing condition-specific models maximizing the agreement between flux predictions and expression measurements methods. The flexibility of these models is reduced, and this can limit their predictive power. If, for example, certain reactions are perturbed by the effect of drugs, perhaps the system shifts to another metabolic state to accommodate the effect of such perturbation. However, due to the fitting of the gene expression data, it might happen that this effect cannot be accounted for, as the predicted metabolic state is biased to represent the gene expression data. In our approach, we initially constrain the reaction bounds in the model with the gene expression data. The constrained model is used to derive a condition-specific biomass reaction. The obtained coefficients of the biomass precursors contain information on the network wide impact of the gene expression data. The constraints in the model are then removed while the newly defined condition-specific biomass reaction is used to provide an indirect representation of the metabolic state corresponding to the expression data. Our goal was to retain flexibility in the model, while incorporating the experimental data.

We reasoned that the enzymes encoded by transcripts and involved in metabolism, which were present at a given moment in Mtb, should roughly reflect the flux through these enzymes at that specific condition and time point. Even though transcript abundance is not linearly correlated to enzyme abundance or flux (i.e. the reaction rate of an enzyme) [38], for larger systems, such as pathways or the entire metabolism, a correlation is likely to exist. On average, metabolic transcript abundance data should reflect the optimal quantity of a given enzyme that is sufficient to perform its metabolic task. Production of an excess of metabolic enzymes would be a waste of energy, and thus unfavorable for an organism residing in a hostile environment.

The synthesis routes towards amino acids are predicted to carry more flux during host infection as compared to *in vitro* growth, which is in agreement with other predictions [15]. This is represented in Table 4.1 by the higher (doubled) fraction of amino acids required. This suggests that protein synthesis is increased upon infection. Mtb is known to excrete proteins during infection, which could explain this predicted increase [8]. At the same time, the predicted lipid synthesis requirement is lower during infection than during growth on Middlebrook medium, confirming the lipid-rich diet that Mtb encounters in the host environment [8, 34]. Another major difference is the lower carbohydrate synthesis. Following the same reasoning, carbohydrates should be more abundant in the host environment, but it is generally assumed that Mtb has poor access to carbohydrates in this environment [39, 40]. A possible explanation could be that Mtb does not synthesize carbohydrates as the synthesis of other metabolites are preferred within the host as compared to growth on Middlebrook medium.

We have used a bi-objective optimization approach to simultaneously take into account growth requirements and parsimonious enzyme utilization. The tradeoff between both objectives is apparent in Figure 4.2. Still the comparison between the uptake profiles in Figures 4.3-4.5 led us to conclude that a ratio between both objectives, f_r, of 0.8 (corresponding to $\log(f_r) = -0.1$) is likely to represent the metabolic state in the host. This suggests that, under these conditions, growth represents a major sink to cellular resources. Here we have selected an equal $f_{e,i}$ for all enzymatic reactions i, however this could be modified to account for differences in enzymes, such as size (molecular weight), activity or degradation rates.

Finally, it should be borne in mind that the transcriptomics data do not represent later infectious states, but a single time point 24 hours post infection, before the onset of growth arrest. As can be seen from Figures 4.3-4.5, the profiles of uptake rates of different nutrients are quite similar for all five (condition-specific) biomass reactions, even though these reactions are very different. Production of a variety of precursors is apparently possible using a more or less fixed set of nutrients. The predicted combination of nutrients that Mtb acquires during infection is surprising from a modeling point of view. As uptake of one nutrient and subsequent production of energy carrying metabolites (ATP, NADH), biomass precursor(s), and excretion of

byproducts, will always be more favorable than that of another metabolite in terms of its potential to sustain growth. The result is that the one nutrient is always favored above another and uptake of multiple nutrients normally does not occur without setting quantitative arbitrary boundaries on uptake rates. This preferential substrate utilization is often regulated at multiple levels, and it should be considered that this type of models does not explicitly account for regulation. Still, the energy and metabolite precursor gain from each nutrient is very balanced using sMtb and the bi-objective optimization, which indicates that enough regulatory information is retained in the transcript data.

A major advantage of the simulations performed within this study is that virtually no assumptions on quantitative uptake rates are required. The only limitation on uptake rates, apart from not allowing uptake of metabolites that are not known or likely to be available in the phagosome, is set on the uptake of oxygen. The phagosome is likely a hypoxic environment [8, 34] and the oxygen uptake rate was therefore (arbitrarily) set to 1% (0.01 mmol gDW^{-1} h^{-1}) of the rate used in previous predictions on Mtb metabolism [9].

The predictions of essential genes using sMtb and the five different (condition-specific) biomass reactions are not overwhelmingly accurate. In general, the specificity (the correct prediction of non-essential genes) is quite good, but the sensitivity (the correct prediction of essential genes) is very poor. This is rather remarkable, as such a long list of biomass precursors (Supplementary File 4.5) is likely to result in a high number of genes predicted to be essential, as there is ample opportunity to disrupt synthesis routes towards many precursors by an *in silico* knockout. Possibly, there are even more metabolic precursors that should be taken into account when creating biomass reactions for Mtb.

Although the biomass reaction representing non-replicating cells, NRC, has the highest accuracy, its sensitivity is the poorest of all biomass reaction, due to its low number of biomass precursors. If one is interested in developing novel therapeutic intervention strategies, the essential genes are arguably the most interesting. In general, the amount of genes that are predicted to be essential is lower than the measured number. This could imply that the list of 108 biomass precursors is still too short. Given that there are 2500 different lipids identified in Mtb up till now [41], the total number of different metabolites is probably a lot higher. Even if metabolic intermediates are omitted, it is still likely that the total number of biomass precursors is well above 108.

The Mtb genome roughly contains 4000 genes, of which a quarter has an unknown function [42]. Model sMtb currently contains 930 genes, which is approximately one-third of the genome having a known function. Extrapolating these figures would mean that there are still an estimated 300 unknown genes in the Mtb genome that are involved in metabolism. So, an estimated quarter of model sMtb is missing. This

will undoubtedly affect predictions made with sMtb.

Another, more fundamental problem lies in the possibility that Mtb and the host continuously interact and a steady state is not easily obtained [43]. As the foundation of constraint-based metabolic models is the stoichiometric matrix, wherein a steady state (i.e. synthesis and degradation rates for each metabolite are equal) is assumed for all metabolites, a non-steady state situation might negatively impact the predictions made using sMtb.

The poor prediction of genes essential to survival of Mtb within the host is in stark contrast to *in vitro* predictions previously made using sMtb where accuracies of 80% were reached [10]. Remarkably, the biomass reactions seem to have limited influence on gene essentiality predictions within the host. As the general list of biomass precursors of model sMtb is primarily derived from *in vitro* data of Mtb, or close relatives of Mtb, the list of biomass precursors could be overfitted to *in vitro* growth conditions.

In addition, the condition-specific biomass reactions could be incorrectly inferred. As the biomass precursors are maximized individually one at a time, information regarding their interdependency is not taken into account. One could for example envision maximizing the sum of the flux towards all biomass precursors at the same time, while minimizing the difference between the overall flux profile and the gene expression profile, instead of the approach taken here. Nevertheless, such a strategy is at risk of ignoring precursors and corresponding synthesis pathways that are relatively lowly expressed, and ending up with only a few precursors in the biomass reaction.

Another explanation is that important constraints are missing. For example, the influence of metal cofactors such as iron and zinc on the metabolic state is ignored, while these cofactors are crucial for intracellular survival, and many metabolic enzymes do not function without these cofactors [7].

Taken together, the lack of predictive power of sMtb regarding in-host essential gene predictions could be caused by several problems, one of the most fundamental problems being the absence of a steady state situation. The gene essentiality measurements from Mendum and colleagues show a similar picture, as only 78-80% of the metabolic genes essential for survival are shared between 3 days and 7 days after infection [32]. This figure is not strikingly low, but it does point in the direction of a lack of a steady state situation. The effect that a non-steady state situation would have on the predictions of essential genes and the metabolic state is difficult to quantify.

Although Mtb is very similar to *Mycobacterium bovis* BCG, there are obvious differences. First of all, Mtb is highly pathogenic to humans, while *M. bovis* BCG is a relatively safe organism. From a metabolic point of view, both organisms are highly similar, although there are some notable differences [44]. Moreover, it is not

unimaginable that metabolic differences during infection are highlighted as *M. bovis* BCG is eventually eradicated within human immune cells, while Mtb is able to withstand and thrive within such cells. Another aspect is that the gene essentiality measurements are made 3 days and 7 days after infection while the dual RNA-seq data is derived from an experiment 1 day after infection.

We developed a method of modeling the metabolism of *M. tuberculosis* during infection of the host's immune cells. The method has the advantage that, unlike previously applied host-pathogen modeling approaches [24], it is virtually free from any artificially placed constraints on metabolite uptake and secretion rates. In addition, our method does not require a pre-composed biomass reaction. The only requirements are: knowledge of nutrient availability, a genome-scale dataset of transcript abundances (such as an RNA-sequencing dataset), a detailed list of biomass precursors, and a genome-scale constraint-based model of metabolism. A relatively small amount of data is required for this method, and it is therefore suited to explore metabolic states of microorganisms in difficult to access environments where an efficient usage of resources is likely to occur.

Our method allows accurate prediction of nutrients from the host, apart from cholesterol uptake, which was not predicted to take place, likely due to lack of knowledge on the complete degradation pathway. A doubled amino acid synthesis requirement was predicted using our method, suggesting an increased synthesis rate of proteins relative to other metabolic precursors during host infection. Lipid synthesis was predicted to decrease during infection, confirming the predominant lipid diet encountered by Mtb within the host.

Flux predictions obtained with the condition-specific biomass reaction, without any further constrains show poor correlation with the transcriptomics data (lower than 0.1). This value is similar to the values obtained using the other four biomass reactions. Poor correlation between transcriptomics data and proteomics measurements has been shown in a wide number of publications [45-47]. In addition, accurate predictions would also require inclusion of enzyme turnover data [48]. This further confirms that fitting the model to the gene expression data might lead to an over constrained model.

It is important to notice that during the onset of infection not only the bacterium undergoes metabolic changes, but also the host environment it thrives in most likely undergoes changes as the host responds to infection. This interplay between the host and the pathogen has not been taken into account as here only the bacterium is modeled. Another reason for the inaccurate gene essentiality predictions could be that many enzymes play additional roles in the synthesis of precursors that are not required during *in vitro* growth or that the list of precursors is not comprehensive. The latter explanation would be plausible, as the predictions on nutrient uptake are quite accurate, suggesting that nutrient uptake is driven by energy efficiency constraints.

Methods

Mtb metabolic model

We used our genome-scale metabolic model of *M. tuberculosis* called sMtb, *in silico M. tuberculosis* [10]. We made some minor corrections to this model regarding among others the respiration chain, and added six reactions to improve the functioning of the model. This improved sMtb model can be found in Supplementary File 4.1 (as systems biology markup language file, SBML) and in Supplementary File 4.2 (as excel file) together with a small summary of the aforementioned changes in Supplementary File 4.3. A list of metabolites that could be present in the phagosome was collected from literature [31], similarly a list of metabolites in Middlebrook 7H9 medium was collected (Supplementary File 4.4).

Constraining sMtb with gene expression data

Raw sequence read data supporting the results of this article are available in the EMBL-EBI European Nucleotide Archive under the Accession No. PRJEB6552, http://www.ebi.ac.uk/ena/data/view/PRJEB6552 for both *M. bovis* BCG grown on Middlebrook 7H9 medium and *M. bovis* BCG cells infecting THP-1 cells. RNA sequencing reads were aligned to the *M. bovis* BCG genome as described before [14]. For each gene present in sMtb, the number of reads aligning to it was summed. A cutoff value of 100 counts per million (cpm) was used to identify lowly expressed genes, that were assigned a count value of zero. The resulting gene count values were subsequently transferred to their corresponding reactions, summing the counts for reactions catalyzed by isozymes. For reactions catalyzed by a protein complex, the smallest number of counts of every gene that encodes a part of such a complex was assigned to the reaction. For reactions that can be catalyzed by several different protein complexes, the smallest number of counts assigned to one of the genes encoding a part of each complex was identified and subsequently the total of all these smallest numbers of counts was assigned to the reaction. Reactions that received no counts using this method were not allowed to carry any flux. Afterwards, the total number of counts assigned to each reaction was normalized by dividing this total number of counts by the largest number of counts assigned to any reaction in sMtb, resulting in a value ranging between 0 and 1 for each enzyme-catalyzed reaction. This procedure is called the E-flux algorithm and is explained in greater detail by Colijn and colleagues [19].

Obtaining condition-specific biomass reactions
The workflow applied to model sMtb is generally depicted in Figure 4.1. This workflow was applied twice: for gene expression data of *M. bovis* BCG grown on Middlebrook 7H9 medium (medium condition) and for gene expression data of *M. bovis* BCG cells infecting THP-1 cells (infection condition). Firstly, upper bounds on unidirectional (forward) reactions, and upper and lower bounds on bidirectional reactions were replaced by the normalized counts assigned to that reaction (Figure 4.1A). The resulting sMtb model constrained by gene expression data was further constrained by setting uptake rates of unavailable metabolites to zero in the given condition and allowing unconstrained uptake of available nutrients, based on nutrient availability data (Figure 4.1B). The nutrient availability in the phagosome and in Middlebrook 7H9 medium is given in Supplementary File 4.4. Afterwards, a general list of biomass precursors was obtained from sMtb (Supplementary Files 4.1 and 4.2). For each biomass precursor, a sink reaction was added and the flux of each of these sink reactions was individually maximized, effectively maximizing the flux towards the respective precursor (Figure 4.1C). Subsequently, the obtained maximum value for each biomass precursor was normalized such that the total molecular weight of these precursors equaled one gram, resulting in a condition-specific biomass reaction of Mtb during infection, growing in phagosomal conditions, CSI, and a condition-specific biomass reaction of Mtb growing in Middlebrook 7H9 medium, CSM (Figure 4.1D). These condition-specific biomass reactions were subsequently added to sMtb and constraints derived from the gene expression were removed (Figure 4.1E).

Calculating nutrient uptake for various objective functions
We compared five different objective functions for their ability to correctly predict nutrient uptake rates by Mtb in the phagosome. The following biomass reactions were used: CSI, CSM, the regular biomass reaction from model sMtb representing growth (REB) [10], the biomass reaction from model sMtb representing *in vitro* growth (IVB) [10] and a reaction representing Mtb in a non-replicative state (NRC) [25].
The bounds on uptake rates of all nutrients representing phagosomal conditions (Supplementary File 4.4) were unconstrained, with the sole exception of constraining the oxygen uptake rate to 0.01 mmol gDw^{-1} h^{-1}. Subsequently, each of the five objective functions was maximized while the sum of all other enzymatically catalyzed reactions was minimized (equations 4.1.1-4.1.3). The weight factor for the biomass reaction, f_b, was varied while keeping the weight factor for enzymatically catalyzed reactions, $f_{e,i}$, constant at 0.001 for all reactions i, hence effectively varying f_r, the ratio between f_b and the sum

of all $f_{e,i}$ values (equation 4.2). Subsequently, a flux variability analysis was performed, wherein the maximum objective function value, *w* (Figure 4.2, right panel), was set as a constraint and the nutrient uptake rates were minimized and maximized individually, resulting in maximal and minimal uptake rate boundaries for various nutrients for each objective function (Figures 4.3-4.5) [4, 49].

Comparison of gene essentialities for various objective functions
At $\log(f_r)$ > -0.5 flux is channeled through all (condition-specific) biomass reactions (Figure 4.2, right panel). At $\log(f_r)$ < 0.3 minimizing flux through the sum of all enzymatically catalyzed reactions is still relevant and the restrictions of the oxygen uptake threshold are not yet overcome. Beyond this point, the (condition-specific) biomass reaction weight factor, f_b, is so large as compared to the sum of all $f_{e,i}$ values, that an optimal objective function value, *w*, is obtained by solely minimizing enzyme usage, and ignoring maximizing flux through the (condition-specific) biomass reaction. An f_r value of 0.8 was chosen from figures 4.2-4.5 as asparagine, alanine and glutamate in addition to glycerol-3-phosphate and CO_2 are taken up from the host, which is likely to occur during infection [31] and this value $\log(0.8) \approx 0.1$ is centered between the boundaries of $\log(f_r)$ = 0.5 and $\log(f_r)$ = 0.3. The growth rates (i.e. the *in silico* calculated flux of CSI, CSM, IVB, REB and NRC) were maximized indirectly by maximizing the aforementioned bi-objective optimization problem. Each biomass reaction will always obtain its maximal value using this approach. Thereafter, using the COBRA toolbox [50], genes and their corresponding reactions where deleted one by one and the resulting specific growth rates were computed by maximizing the aforementioned bi-objective optimization problem. These growth rates were divided by the wild-type growth rate, resulting in a number between 0 and 1 for each knocked-out gene, representing the relative specific growth rate. We applied a 95% reduction in the relative growth rate as a threshold to indicate essential genes as described before [10].
Mendum and colleagues infected human dendritic cells with an Mtb transposon library to identify genes that are required for *in vivo* survival after 3 days and after 7 days [32]. These experimentally identified essential genes were compared to the predicted essential genes using the aforementioned five different objective functions. Subsequently, the accuracy, sensitivity and specificity of the predictions, were calculated for all five objective functions and for both experimental time points.

Availability of supporting data

Supplementary Files can be accessed via: Rienk A. Rienksma, Peter J. Schaap, Vitor A.P. Martins dos Santos, Maria Suarez-Diez. Modeling the metabolic state of *Mycobacterium tuberculosis* upon infection. Frontiers in Cellular and Infection Microbiology 8 (2018).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

References

4

1. Kauffman, K.J., P. Prakash, and J.S. Edwards, *Advances in flux balance analysis.* Current Opinion in Biotechnology, 2003. **14**(5): p. 491-496.

2. Bordbar, A., J.M. Monk, Z.A. King, and B.O. Palsson, *Constraint-based models predict metabolic and associated cellular functions.* Nat Rev Genet, 2014. **15**(2): p. 107-20.

3. Reed, J.L., *Shrinking the Metabolic Solution Space Using Experimental Datasets.* PLoS Comput Biol, 2012. **8**(8): p. e1002662.

4. Orth, J.D., I. Thiele, and B.O. Palsson, *What is flux balance analysis?* Nat Biotechnol, 2010. **28**(3): p. 245-8.

5. Raman, K. and N. Chandra, *Flux balance analysis of biological systems: applications and challenges.* Briefings in Bioinformatics, 2009. **10**(4): p. 435-449.

6. Schuetz, R., L. Kuepfer, and U. Sauer, *Systematic evaluation of objective functions for predicting intracellular fluxes in* Escherichia coli. Mol Syst Biol, 2007. **3**: p. 119.

7. Zondervan, A.N., C.J. van Dam, J.P. Schaap, A.V. Martins dos Santos, and M. Suarez-Diez, *Regulation of Three Virulence Strategies of* Mycobacterium tuberculosis: *A Success Story.* International Journal of Molecular Sciences, 2018. **19**(2).

8. Gengenbacher, M. and S.H.E. Kaufmann, Mycobacterium tuberculosis: *success through dormancy.* FEMS Microbiology Reviews, 2012. **36**(3): p. 514-532.

9. Jamshidi, N. and B. Palsson, *Investigating the metabolic capabilities of* Mycobacterium tuberculosis *H37Rv using the* in silico *strain iNJ661 and proposing alternative drug targets.* BMC Syst Biol, 2007. **1**: p. 26.

10. Rienksma, R.A., M. Suarez-Diez, L. Spina, P.J. Schaap, and V.A.P. Martins dos Santos, *Systems-level modeling of mycobacterial metabolism for the identification of new (multi-)drug targets.* Seminars in Immunology, 2014. **26**(6): p. 610-622.

11. Hoppe, A., *What mRNA Abundances Can Tell us about Metabolism.* Metabolites, 2012. **2**(3): p. 614-631.

12. Fels, U., K. Gevaert, and P. Van Damme, *Proteogenomics in Aid of Host–Pathogen Interaction Studies: A Bacterial Perspective.* Proteomes, 2017. **5**(4): p. 26.

13. Mangan, J.A., I.M. Monahan, and P.D. Butcher, *Gene expression during host—pathogen interactions: Approaches to bacterial mRNA extraction and labelling for microarray analysis*, in *Methods in Microbiology*, N.D. Brendan Wren, Editor. 2002, Academic Press. p. 137-151.

14. Rienksma, R.A., M. Suarez-Diez, H.-J. Mollenkopf, G.M. Dolganov, A. Dorhoi, G.K. Schoolnik, V.A.P. Martins dos Santos, S.H.E. Kaufmann, P.J. Schaap, and M. Gengenbacher, *Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing.* BMC Genomics, 2015. **16**(34).

15. Zimmermann, M., M. Kogadeeva, M. Gengenbacher, G. McEwen, H.-J. Mollenkopf, N. Zamboni, S.H.E. Kaufmann, and U. Sauer, *Integration of Metabolomics and Transcriptomics Reveals a Complex Diet of* Mycobacterium tuberculosis *during Early Macrophage Infection.* mSystems, 2017. **2**(4).

16. Shlomi, T., M.N. Cabili, M.J. Herrgard, B.O. Palsson, and E. Ruppin, *Network-based prediction of human tissue-specific metabolism.* Nat Biotech, 2008. **26**(9): p. 1003-1010.

17. Jensen, P.A. and J.A. Papin, *Functional integration of a metabolic network model and expression data without arbitrary thresholding.* Bioinformatics, 2011. **27**(4): p. 541-547.

18. Becker, S.A. and B.O. Palsson, *Context-Specific Metabolic Networks Are Consistent with Experiments.* PLoS Comput Biol, 2008. **4**(5): p. e1000082.

19. Colijn, C., A. Brandes, J. Zucker, D.S. Lun, B. Weiner, M.R. Farhat, T.Y. Cheng, D.B. Moody, M. Murray, and J.E. Galagan, *Interpreting expression data with metabolic flux models: predicting* Mycobacterium tuberculosis *mycolic acid production.* PLoS Comput Biol, 2009. **5**(8): p. e1000489.

20. Motamedian, E., M. Mohammadi, S.A. Shojaosadati, and M. Heydari, *TRFBA: an algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data.* Bioinformatics, 2017. **33**(7): p. 1057-1063.

21. Lewis, N.E., H. Nagarajan, and B.O. Palsson, *Constraining the metabolic genotype– phenotype relationship using a phylogeny of* in silico *methods.* Nat Rev Micro, 2012. **10**(4): p. 291-305.

22. Rossell, S., M.A. Huynen, and R.A. Notebaart, *Inferring Metabolic States in Uncharacterized Environments Using Gene-Expression Measurements.* PLoS Comput Biol, 2013. **9**(3): p. e1002988.

23. Machado, D. and M. Herrgard, *Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism.* PLoS Comput Biol, 2014. **10**(4): p. e1003580.

24. Bordbar, A., N.E. Lewis, J. Schellenberger, B.O. Palsson, and N. Jamshidi, *Insight into human alveolar macrophage and* M. tuberculosis *interactions via metabolic reconstructions.* Mol Syst Biol, 2010. **6**: p. 422.

25. Shi, L., C.D. Sohaskey, C. Pfeiffer, P. Datta, M. Parks, J. McFadden, R.J. North, and M.L. Gennaro, *Carbon flux rerouting during* Mycobacterium tuberculosis *growth arrest.* Mol Microbiol, 2010. **78**(5): p. 1199-215.

26. Fontan, P., V. Aris, S. Ghanny, P. Soteropoulos, and I. Smith, *Global transcriptional profile of* Mycobacterium tuberculosis *during THP-1 human macrophage infection.* Infect Immun, 2008. **76**(2): p. 717-25.

27. Papalambros, P.Y. and D.J. Wilde, *Principles of optimal design: modeling and computation*. 2nd ed. ed. 2000, United States of America: The Press Syndicate of the University of Cambridge.

28. Messac, A., A. Ismail-Yahaya, and C.A. Mattson, *The normalized normal constraint method for generating the Pareto frontier.* Structural and Multidisciplinary Optimization, 2003. **25**(2): p. 86-98.

29. Nagrath, D., M. Avila-Elchiver, F. Berthiaume, A.W. Tilles, A. Messac, and M.L. Yarmush, *Integrated Energy and Flux Balance Based Multiobjective Framework for Large-Scale Metabolic Networks.* Annals of Biomedical Engineering, 2007. **35**(6): p. 863-885.

4

30. Suárez, M., P. Tortosa, J. Carrera, and A. Jaramillo, *Pareto optimization in computational protein design with multiple objectives.* Journal of Computational Chemistry, 2008. **29**(16): p. 2704-2711.

31. Beste, Dany J.V., K. Nöh, S. Niedenführ, Tom A. Mendum, Nathaniel D. Hawkins, Jane L. Ward, Michael H. Beale, W. Wiechert, and J. McFadden, *¹³C-Flux Spectral Analysis of Host-Pathogen Metabolism Reveals a Mixed Diet for Intracellular* Mycobacterium tuberculosis. Chemistry & Biology, 2013. **20**(8): p. 1012-1021.

32. Mendum, T., H. Wu, A. Kierzek, and G. Stewart, *Lipid metabolism and Type VII secretion systems dominate the genome scale virulence profile of* Mycobacterium tuberculosis *in human dendritic cells.* BMC Genomics, 2015. **16**(1): p. 372.

33. Garnier, T., K. Eiglmeier, J.C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P.R. Wheeler, J. Parkhill, B.G. Barrell, S.T. Cole, S.V. Gordon, and R.G. Hewinson, *The complete genome sequence of* Mycobacterium bovis. Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7877-82.

34. Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan, and G.K. Schoolnik, *Transcriptional Adaptation of* Mycobacterium tuberculosis *within Macrophages: Insights into the Phagosomal Environment.* J Exp Med, 2003. **198**(5): p. 693-704.

35. Fontán, P., V. Aris, S. Ghanny, P. Soteropoulos, and I. Smith, *Global Transcriptional Profile of* Mycobacterium tuberculosis *during THP-1 Human Macrophage Infection.* Infection and Immunity, 2008. **76**(2): p. 717-725.

36. Wipperman, M.F., N.S. Sampson, and S.T. Thomas, *Pathogen roid rage: Cholesterol utilization by* Mycobacterium tuberculosis. Critical Reviews in Biochemistry and Molecular Biology, 2014: p. 1-25.

37. Beste, D., T. Hooper, G. Stewart, B. Bonde, C. Avignone-Rossa, M. Bushell, P. Wheeler, S. Klamt, A. Kierzek, and J. McFadden, *GSMN-TB: a web-based genome-scale network model of* Mycobacterium tuberculosis *metabolism.* Genome Biol, 2007. **8**: p. R89.

38. Bordel, S., R. Agren, and J. Nielsen, *Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes.* PLoS Comput Biol, 2010. **6**(7): p. e1000859.

39. Fullam, E., I. Prokes, K. Fütterer, and G.S. Besra, *Structural and functional analysis of the solute-binding protein UspC from* Mycobacterium tuberculosis *that is specific for amino sugars.* Open Biology, 2016. **6**(6): p. 160105.

40. Kalscheuer, R., B. Weinrick, U. Veeraraghavan, G.S. Besra, and W.R. Jacobs, *Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of* Mycobacterium tuberculosis. Proc Natl Acad Sci, 2010. **107**: p. 21761-21766.

41. Sartain, M.J., D.L. Dick, C.D. Rithner, D.C. Crick, and J.T. Belisle, *Lipidomic analyses of* Mycobacterium tuberculosis *based on accurate mass measurements and the novel "Mtb LipidDB".* Journal of Lipid Research, 2011. **52**(5): p. 861-872.

42. Qin, L., S. Gao, J. Wang, R. Zheng, J. Lu, and Z. Hu, *The Conservation and Application of Three Hypothetical Protein Coding Gene for Direct Detection of* Mycobacterium tuberculosis *in Sputum Specimens.* PLoS ONE, 2013. **8**(9): p. e73955.

43. Garton, Natalie J. and Helen M. O'Hare, *Tuberculosis: Feeding the Enemy.* Chemistry & Biology, 2013. **20**(8): p. 971-972.

44. Lofthouse, E.K., P.R. Wheeler, D.J.V. Beste, B.L. Khatri, H. Wu, Tom A. Mendum, A.M. Kierzek, and J. McFadden, *Systems-based approaches to probing metabolic variation within the* Mycobacterium tuberculosis *complex.* PLoS ONE, 2013. **8**(9): e75913.

45. Maier, T., M. Güell, and L. Serrano, *Correlation of mRNA and protein in complex biological samples.* FEBS Letters, 2009. **583**(24): p. 3966-3973.

46. Payne, S.H., *The utility of protein and mRNA correlation.* Trends in Biochemical Sciences, 2015. **40**(1): p. 1-3.

47. Edfors, F., F. Danielsson, B.M. Hallström, L. Käll, E. Lundberg, F. Pontén, B. Forsström, and M. Uhlén, *Gene-specific correlation of RNA and protein levels in human cells and tissues.* Molecular Systems Biology, 2016. **12**(10).

48. Sánchez Benjamín, J., C. Zhang, A. Nilsson, P.J. Lahtvee, J. Kerkhoven Eduard, and J. Nielsen, *Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints.* Molecular Systems Biology, 2017. **13**(8): p. 935.

49. Gudmundsson, S. and I. Thiele, *Computationally efficient flux variability analysis.* BMC Bioinformatics, 2010. **11**(1): p. 489.

50. Schellenberger, J., R. Que, R.M.T. Fleming, I. Thiele, J.D. Orth, A.M. Feist, D.C. Zielinski, A. Bordbar, N.E. Lewis, S. Rahmanian, J. Kang, D.R. Hyduke, and B.O. Palsson, *Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.* Nat. Protocols, 2011. **6**(9): p. 1290-1307.
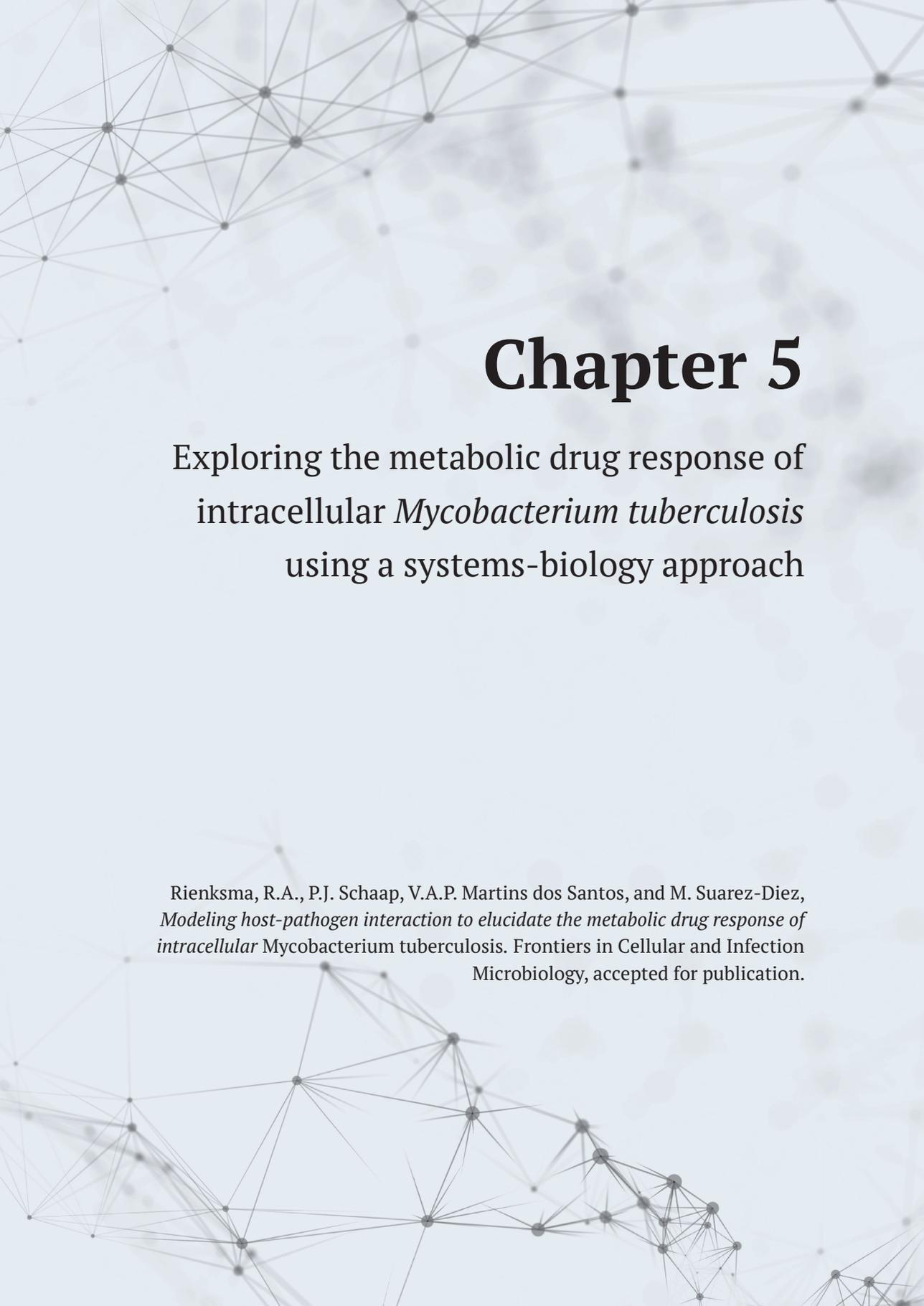
Chapter 5

Exploring the metabolic drug response of intracellular *Mycobacterium tuberculosis* using a systems-biology approach

Abstract

Little is known about the metabolic state of *Mycobacterium tuberculosis* (Mtb) inside the phagosome, a compartment inside phagocytes for killing pathogens and other foreign substances. We have developed a combined model of Mtb and human metabolism, sMtb-RECON and used this model to predict the metabolic state of Mtb during infection of the host. Amino acids are predicted to be used for energy production as well as biomass formation. Subsequently we assessed the effect of increasing dosages of drugs targeting metabolism on the metabolic state of the pathogen and predict resulting metabolic adaptations and flux rerouting through various pathways. In particular, the TCA cycle becomes more important upon drug application, as well as alanine, aspartate, glutamate, proline, arginine and porphyrin metabolism, while glycine, serine and threonine metabolism become less important. We modeled the effect of eleven metabolically active drugs. Notably, the effect of eight could be recreated and two major profiles of the metabolic state were predicted. The profiles of the metabolic states of Mtb affected by the drugs BTZ043, cycloserine and its derivative terizidone, ethambutol, ethionamide, propionamide, and isoniazid were very similar, while TMC207 is predicted to have quite a different effect on metabolism as it inhibits ATP synthase and therefore indirectly interferes with a multitude of metabolic pathways.

Background

Mycobacterium tuberculosis (Mtb), the etiological agent of tuberculosis, is an intracellular pathogen that thrives inside the phagosome of the host's macrophages [1, 2]. This environment prevents obtaining accurate *in vivo* measurements characterizing the metabolic state of the pathogen during infection. Genome-scale models (GSMs) of metabolism have been proposed as efficient tools to explore bacterial metabolism, even in conditions difficult to access experimentally. Flux balance analysis (FBA) is a widely used approach to study GSMs and relies on the definition of an objective function that characterizes the metabolic objective of the organism under study [3]. Predictions made using GSMs are highly dependent on the objective that is being used and the constraints placed on the uptake and excretion of nutrients and metabolites. To perform predictions on *in vitro* growth, most often a biomass reaction is selected as objective function for maximization. The biomass reaction details the biomass composition in terms of its constituents such as proteins, lipids and nucleic acids. This composition might vary in different growth conditions [4]. Biomass as an objective function has been used for the earliest genome-scale metabolic models of Mtb [5, 6] as well as more recent models [7-13]. Recently, a condition-specific biomass reaction has been formulated for Mtb inside the host by integrating model sMtb [9] and gene expression data during on-going infection [14].

In silico gene knockout analysis has been the method of choice to predict metabolic drug targets, in the form of genes and their associated enzyme products [6, 13, 15]. Such methods are based on analyzing the effect of completely blocking the flux through the corresponding reactions on growth predictions. This full blockage approach would fail to predict cases wherein drugs reach Mtb in relatively small amounts so that enzyme function is only partly lost, allowing Mtb to counteract the negative effects of such a drug by altering its metabolic state to overcome non-optimal fluxes due to the drug affected enzymes. Bhat and colleagues developed a method to study dose dependent effects of isoniazid on the metabolic state that relied on simulating the effects of partial loss of function of the affected enzymes [16].

To capture the interaction between Mtb and its host on a metabolic level, a model of macrophage metabolism is required. Nutrients for Mtb are obtained from the phagosome, a cellular compartment specific to macrophages, and from the cytosol after Mtb gains access [17]. The phagosome represents a nutrient-poor, hypoxic, and nitrosative environment wherein Mtb is able to survive [18]. Nevertheless, the nutrients available in the phagosome, arguably after cytosolic access, are predicted to be varied [19, 20], and as such, allow metabolic flexibility of Mtb, which is best captured using a combined host-pathogen model. Although a host-pathogen

metabolic model is more elaborate, and as such introduces more uncertainty and variability, several gene expression datasets have been introduced that cover both host and pathogen [19, 21], and as such are suited to constrain such a model to make it condition-specific.

For Mtb, a host-pathogen model was first created by Bordbar and colleagues [22], based on iNJ661, a well known Mtb model published in 2007 [6] and RECON 1, the first global human metabolic reconstruction [23]. The combined model allowed simulation of metabolic changes during infection and three distinct pathological states of Mtb were described.

Improved versions of the individual models describing the metabolism of host and pathogen are available. Model sMtb is a comprehensive model of Mtb metabolism with an increased scope of the underlying metabolic network, and increased predictive power regarding the metabolic state and gene essentiality [9]. RECON 2.2 [24, 25], almost doubles the size of the metabolic network of RECON 1. Here, we integrate sMtb and RECON 2.2 to create an Mtb-Macrophage model, sMtb-RECON. The combined model has condition-specific objective functions for both pathogen and host, based on dual RNA-sequencing data. By applying various known metabolic drugs *in silico*, we highlight pathways that are important for Mtb to escape eradication by drug and host. Drugs that specifically target these pathways could therefore prove to be a valuable addition to the existing drugs

Results

Host-pathogen model sMtb-RECON

Model sMtb-RECON has a total of 8987 reactions and 13.4% are from Mtb. Model sMtb-RECON contains 6373 metabolites and 2605 genes, of which 16.5% and 35.7% are from Mtb, respectively. RECON 2.2 has 9 compartments in total (number of metabolites indicated between brackets): cytoplasm (1918), extracellular space (770), golgi apparatus (312), lysosome (291), mitochondrion (756), nucleus (161), endoplasmic reticulum (675), peroxisome (440), and the mitochondrial intermembrane space (1). No phagosome compartment is available in RECON 2.2. There is a lysosome, but Mtb is known to block phagosome-lysosome fusion, therefore, the metabolites in this compartment are not likely to be available as nutrients for Mtb. Mtb is however assumed to acquire access to the cytosol. As this compartment contains the majority of the biomass precursors for the host, Mtb is assumed to have access to these. The cytosolic biomass precursors in model RECON 2.2 are thus set as metabolites that can be taken up by the Mtb part of model sMtb-RECON. A list of the metabolites and their maximal uptake rates is given in Table 5.1. Model sMtb-RECON can be accessed via https://doi.org/10.18174/466578.

Table 5.1 - Nutrient availability inside the host.

A list of metabolites predicted to be available for Mtb inside the host and their predicted maximal uptake rates.

	Metabolite		Maximal uptake rate
	RECON 2	sMtb	(mmol·gDW^{-1}·h^{-1})
1	ala_L	ALA	1
2	pa_hs		1
3	amp	AMP	1
4	arg_L	ARG	1
5	asn_L	ASN	1
6	asp_L	ASP	1
7	atp	ATP	1
8	chsterol	CHOLESTEROL	0
9	cmp	CMP	0.37
10	cys_L	CYS	1
11	dag_hs	DAG	1
12	damp	DAMP	0.30
13	dcmp	DCMP	1
14	dgmp	DGMP	0.28
15	dtmp	DTMP	0.08
16	gln_L	GLN	1
17	glu_L	GLU	1
18	gly	GLY	0.02
19	glygn2		0.21
20	gmp	GMP	1
21	hdca	HEXADECANOATE	0.32
22	hdcea	9HEXADECENOATE	1
23	his_L	HIS	1
24	ile_L	ILE	1
25	leu_L	LEU	1
26	met_L	MET	0.12
27	ocdca	OCTADECANOATE	1
28	ocdcea	9OCTADECENOATE	0.03
29	pail_hs	PITBA	0.23
30	pchol_hs	PHOSPHATIDYLCHOLINE	1
31	pe_hs	ETHA	1
32	pglyc_hs	PG	1
33	phe_L	PHE	1
34	pro_L	PRO	1

5

Table 5.1 - continued

	Metabolite		Maximal uptake rate
	RECON 2	sMtb	(mmol·gDW^{-1}·h^{-1})
36	ser_L	SER	0
37	sphmyln_hs		0.05
38	tag_hs	TAG	1
39	thr_L	THR	1
40	trp_L	TRP	0
41	ttdca	TETRADECANOATE	1
42	tyr_L	TYR	1
43	ump	UMP	1
44	val_L	VAL	1
45	adp	ADP	1
46	no	NO	1
47	co2	CO2	1
48	glyc3p	GL3P	1
49	o2	O2	1
50	lys_L	LYS	1

Modeling host-pathogen interaction

We extended the method presented in [14] to integrate model and gene expression data to arrive at a model describing the metabolic state of the system during infection. The approach is summarized in Figure 5.1. We used dual RNA sequencing data obtained 24 hours after exposing macrophage-like THP-1 cells to *Mycobacterium bovis* BCG, a close relative to Mtb and as such, this data reflects data derived from a real life infection with Mtb [21].

First, the combined sMtb-RECON model was modified so that all reversible reactions of the Mtb part of sMtb-RECON were split in a forward and backward reaction, to make the sMtb part of the model irreversible (thus bringing the total number of reactions to 9408). Then the combined model was constrained using the dual RNA seq data and condition-specific biomass reactions were obtained, for host and pathogen, by maximizing each human biomass precursor one-by-one for both the Mtb and the human part of sMtb-RECON. Afterwards, the constraints placed on sMtb-Recon were removed and the condition-specific biomass reaction of the human part of sMtb-RECON was used as a proxy for nutrient availability for the Mtb part of sMtb-RECON. The maximum allowable uptake rates of the Mtb part were thus limited to the maximum obtainable fluxes for each human biomass
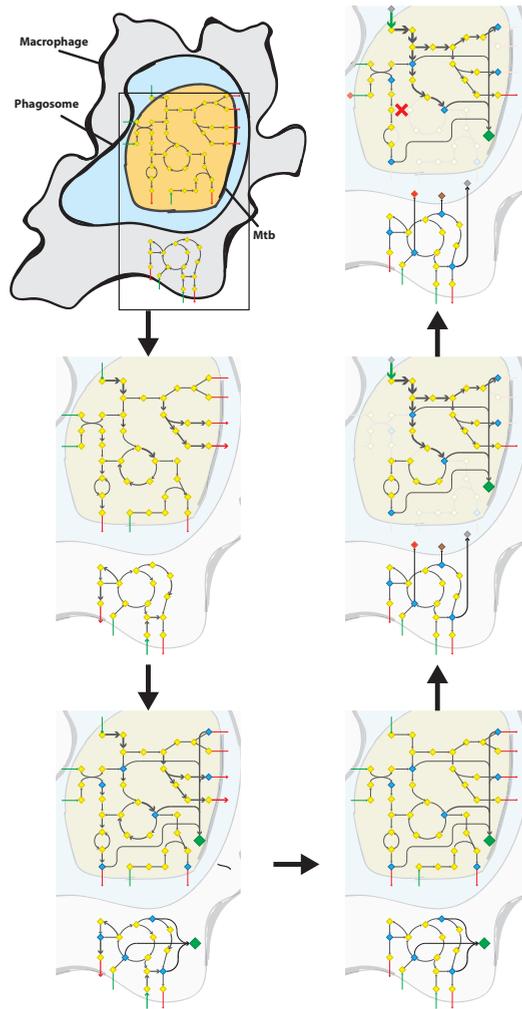
5

Figure 5.1 - Predicting the metabolic state of Mtb during infection and drug application.
Schematic overview of the steps required to calculate a metabolic state of Mtb during infection from the upper left panel following the arrows to the upper right panel. Mtb (yellow shape) is depicted inside the phagosome (light blue shape) of a macrophage (grey shape). Metabolisms of both Mtb and Macrophage are indicated with diamonds and arrows. Yellow diamonds represent metabolites, blue diamonds represent biomass precursors, grey/brown/red diamonds represent phagosomal nutrients, and the large green diamond represents the condition-specific biomass reaction. Grey arrows represent metabolic conversion rates, green arrows represent uptake rates, and red arrows represent secretion rates, wherein the thickness of the arrows is proportional to the rate. Red crosses represent blocked metabolic conversions.
Upper left panel: A constraint-based genome-scale model of Mtb metabolism is coupled to one of human metabolism and a combined model is obtained.
Middle left panel: Metabolic conversion rates (i.e. metabolic fluxes) are constrained proportional to mRNA transcript abundance in both Mtb and macrophage.
Lower left panel: For each potential biomass precursor, the flux through metabolism towards that precursor is maximized, both for Mtb and the macrophage, obtaining two condition-specific biomass reactions.
Lower right panel: The constraints based on mRNA transcript abundance are removed from the combined model.
Middle right panel: The nutrients available for Mtb in the phagosome, and their maximum uptake rate are set according to the condition-specific biomass reaction of the macrophage. Subsequently, the condition-specific biomass reaction of Mtb is maximized, while the total flux through enzymatically-catalyzed reactions is minimized.
Upper right panel: The rate of an enzymatically-catalyzed reaction is constricted by the effect of a metabolically active drug (red cross within the rounded square) and metabolic rerouting occurs towards a part of metabolism that contains a relatively higher number of enzymatically catalyzed reactions.

precursor. Thereafter, all reactions affected by a drug were gradually constrained as metabolism starts to reroute.

Using this method, uptake and secretion profiles can be predicted. This allows prediction of metabolites that are taken up or secreted even if in the starting model no additional constraints are imposed to limit uptake or secretion rates, apart from the oxygen uptake rate. Mtb encounters a hypoxic environment inside the host [18, 26, 27]. Therefore, we used the oxygen uptake rate to constrain the model to such an extent, that prediction of uptake and secretion rates becomes feasible without arbitrarily chosen bounds on other uptake and/or secretion rates.

Metabolic state during infection

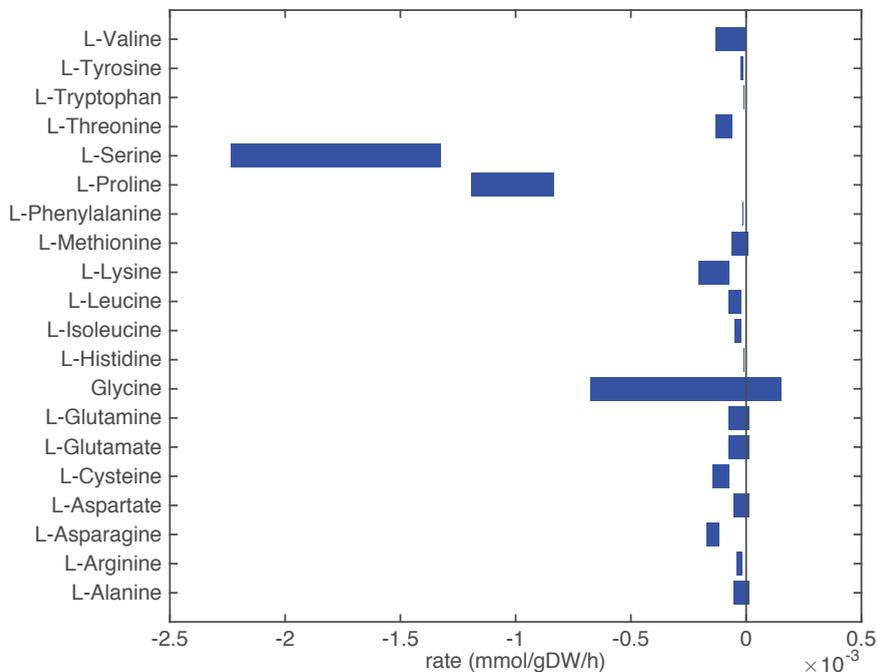Mtb is predicted to take up and secrete a plethora of different metabolites at varying rates (Figures 5.2-5.5).

Figure 5.2 - Predicted amino acids uptake and secretion rates by Mtb in the host.
Predicted ranges of uptake and secretion rates (mmol gDW^{-1} h^{-1}) of amino acids by Mtb inside the host are indicated by blue bars. Negative values denote uptake and positive values denote secretion.

Almost all amino acids are predicted to be taken up by Mtb (Figure 5.2). Most notably serine and proline are predicted to be taken up at relatively high rates. Glycine uptake and/or secretion rates remain largely underdetermined, in fact

using this approach the model is not able to predict whether it is produced or consumed. Notably, our approach also allows a small production of glutamine, glutamate, aspartate and alanine. Glutamate can be interconverted to glutamine by, for example, glutamine synthase at the expense of ATP [28]. In this way any additional uptake of glutamate can serve as a potential source of glutamine, or the other way around, although ATP expenditure limits this interconversion. Therefore, the ranges wherein glutamate and glutamine are predicted to be taken up are equal.

The pattern observed in Figure 5.2 is not a reflection of the coefficients for the amino acids in the condition specific biomass reaction of sMtb, as one would perhaps expect. This can be seen in figure 5.6, where one would expect that all amino acids that are taken up, would be incorporated into biomass, which is obviously not the case. By multiplying the flux through the condition specific biomass reaction with the respective column of the stoichiometric matrix corresponding to this biomass reaction, the fluxes required for synthesis of the individual biomass precursors can be obtained. When comparing the fluxes required for biomass synthesis with their respective predicted uptake rates, most notably alanine and aspartate are predicted to be synthesized by Mtb (Figure 5.6 upper panel). On the other hand, almost all serine, proline and glycine is used for purposes other than biomass synthesis, i.e. ATP and NADH production required for maintenance (Figure 5.6 lower panel). Such behavior has been described in cancer cells [29, 30], but not for Mtb.
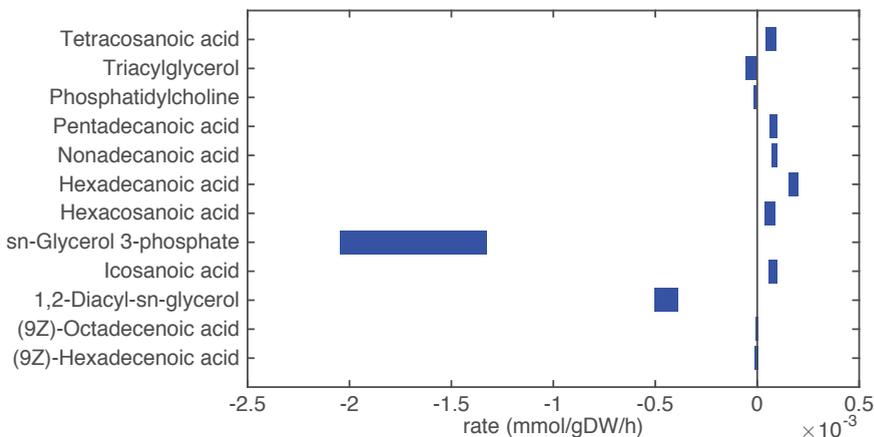

Figure 5.3 - Predicted lipid uptake and secretion rates by Mtb in the host.
Predicted ranges of uptake and secretion rates (mmol gDW⁻¹ h⁻¹) of lipids by Mtb inside the host are indicated by blue bars. Negative values denote uptake and positive values denote secretion.

Glycerol-3-phosphate is a lipid (precursor) that is taken up in a relatively high amount (Figure 5.3). It is known that glycerol-3-phosphate serves as a major carbon source for several intracellular pathogens [31] and it has been suggested that glycerol-3-phosphate might serve as an alternative carbon source for Mtb *in vivo* [32].

Glycerolipids such as diacylglycerol (DAG), triacylglycerol (TAG), and phosphatidylcholine are predicted to be taken up as well. Most notably, DAG is taken up in relatively large amounts (0.5 mmol gDW^{-1} h^{-1}). These three metabolites are closely related, as phosphatidylcholine can be converted to DAG and choline phosphate by phospholipase C [33]. DAG can be converted to TAG, which is subsequently stored in lipid droplets [34]. Unsaturated fatty acids (octadecenoic and hexadecenoic acids) are also predicted to be taken up. On the other hand, a range of saturated fatty acids is predicted to be secreted. These fatty acids are derived from TAG and DAG, indicating that there is a higher requirement for the glycerol backbone of TAG and DAG than for the attached fatty acids.
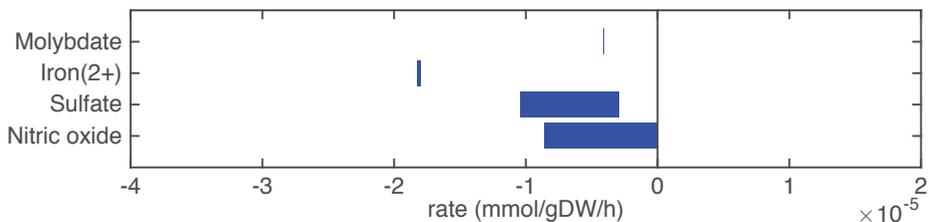

Figure 5.4 - Predicted uptake and secretion rates of cofactors and small molecules by Mtb in the host. Predicted ranges of uptake and secretion rates (mmol gDW^{-1} h^{-1}) of cofactors and small molecules by Mtb inside the host are indicated by blue bars. Negative values denote uptake and positive values denote secretion.

Small metal cofactors, such as molybdate and iron are predicted to be taken up (Figure 5.4). Mtb is known to chelate iron using siderophores, called mycobactins, via a specialized ESX-3 system [35]. This ESX-3 system is essential for *in vitro* growth [36]. In sMtb, iron as an ion or element, without being integrated in a larger molecule, is not incorporated in the condition-specific biomass reaction, as is the case in most biomass reactions of models of Mtb metabolism [6, 9]. However, iron incorporated in larger molecules, such as heme groups, is present in the condition-specific biomass reaction(s) of sMtb.

The excretion of orthophosphate and nitrite (Figure 5.5) are probably artifacts from the model, where phosphate might be derived from the phosphate group of glycerol-3-phosphate and nitrite could be related to the nitrogen groups of the variety of amino acids that are taken up. The model predictions show that free
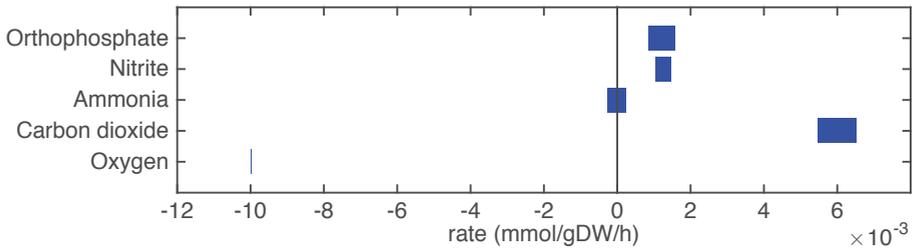
Figure 5.5 - Predicted uptake and secretion rates of molecular oxygen, carbon dioxide and other small molecules by Mtb in the host.
Predicted ranges of uptake and secretion rates (mmol gDW^{-1} h^{-1}) of oxygen, carbon dioxide, and other small molecules by Mtb inside the host are indicated by blue bars. Negative values denote uptake and positive values denote secretion.

5

ammonia can be taken up as well as secreted at about equal rates, so the fate of ammonia uptake or secretion remains inconclusive from these predictions.

The oxygen uptake rate equals 0.01 mmol gDW^{-1} h^{-1} which equals the imposed maximum uptake rate (Figure 5.5). Decreasing the lower bound on oxygen exchange, i.e. allowing a higher uptake rate of oxygen, results in a higher specific growth rate. As such, the system is limited by oxygen and it is obvious that oxygen is taken up at its maximum rate. Carbon dioxide is secreted mainly due to respiration.
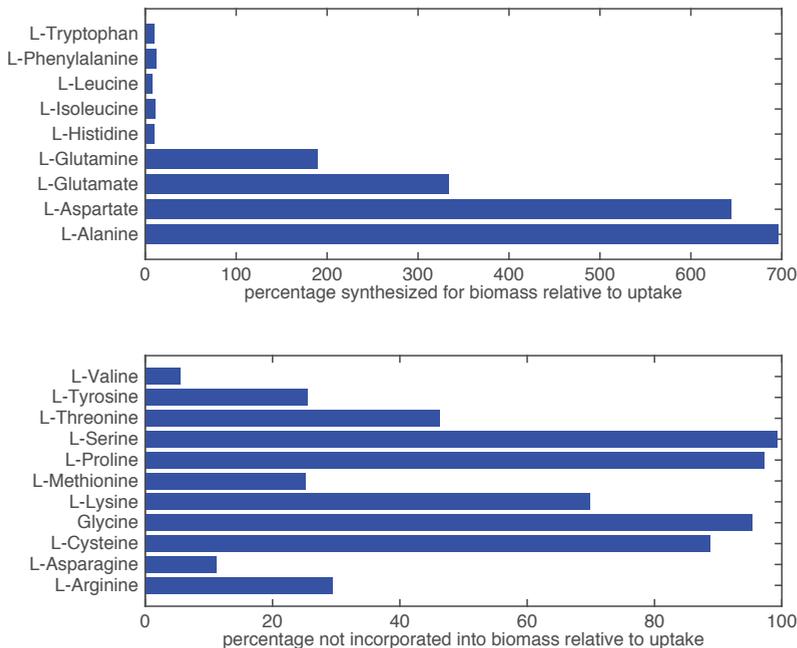

Figure 5.6 - Usage of amino acids derived from the host.
Top: Rate of amino acids synthesis by Mtb for biomass incorporation relative to their respective uptake rates.
Bottom: Relative rate of amino acid uptake that is used in processes other than biomass synthesis.

Metabolic states of drugged Mtb: Rerouting metabolism

Table 5.2 illustrates the mode of action of twelve anti-TB drugs known to interfere with metabolic enzyme activity with known targets. To simulate increasing dosages of these drugs, we gradually decreased the flux through the reactions catalyzed by the affected enzymes.

Table 5.2 - Drugs acting on metabolic enzymes.

Drug	Mode of action	Target
Isoniazid	Inhibits mycolic acid synthesis and folate synthesis	Activated by KatG, targets InhA, KasA and DfrA
Ethambutol	Inhibits arabinogalactan synthesis	Possibly EmbB
Ethionamide/ Prothionamide*	Inhibits mycolic acid synthesis	Activated by EthA, targets InhA
Cycloserine/ Terizidone*	Inhibits peptidoglycan synthesis by blocking the synthesis and use of D-alanine	Targets Alr and ddlA.
Para-amino salicylic acid	Inhibits folate metabolism	DfrA
TMC207	Inhibits ATP synthase	AtpE
BTZ043	Inhibits essential cell-wall arabinan synthesis	DprE1
V-13-011503/ V-13-012725	Inhibits cholesterol catabolism	HsaAB
V-13-009920	Inhibits the methylcitrate cycle	PrpC

* Drugs with common targets have been grouped.

Of the twelve drugs in Table 5.2, some are grouped as they have the same enzyme target, resulting in nine drugs or groups of drugs with different targets. No effect on Mtb metabolism could be predicted for three of these: V-13-011503/V-13-012725, V-13-009920, and para-amino salicylic acid. The drugs BTZ043, cycloserine/ terizidone, ethambutol, ethionamide/propionamide, and isoniazid are predicted to have a very similar effect on metabolism and their effect is therefore grouped (Figures 5.7-5.9, left panels), even though their enzymatic targets are very different (Table 5.2). Notably, TMC207 has a very different effect on metabolism (Figures 5.7-5.9, right panels).
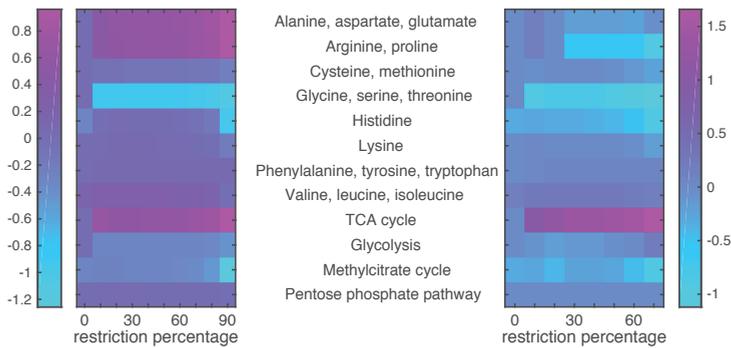
Figure 5.7 - Flux rerouting through amino acid metabolism and major pathways upon application of drugs.
Heat maps indicate a predicted relative increase (pink) or decrease (light blue) through various pathways upon the application of a variety of drugs. The logarithm of the sum of all absolute fluxes is given per pathway (values are indicated in the color bars on either side), so information on directionality is not comprised. The x-axis indicates the percentage of restriction of the drug-affected reaction(s). **Left**: average of BTZ043, cycloserine/terizidone, ethambutol, ethionamide/propionamide, and isoniazid; **Right**: TMC207.
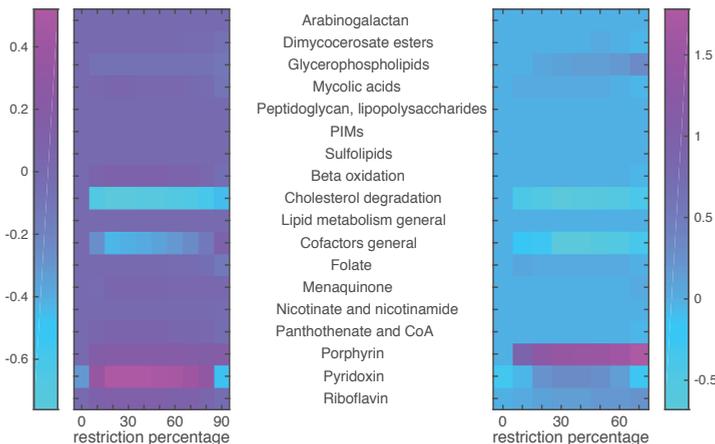
Figure 5.8 - Flux rerouting through cell wall component metabolism, lipid metabolism, and cofactor metabolism upon application of drugs.
Heat maps indicate a predicted relative increase (pink) or decrease (light blue) through various pathways upon the application of a variety of drugs. The logarithm of the sum of all absolute fluxes is given per pathway (values are indicated in the color bars on either side), so information on directionality is not comprised. The x-axis indicates the percentage of restriction of the drug-affected reaction(s). **Left**: average of BTZ043, cycloserine/terizidone, ethambutol, ethionamide/propionamide, and isoniazid; **Right**: TMC207.
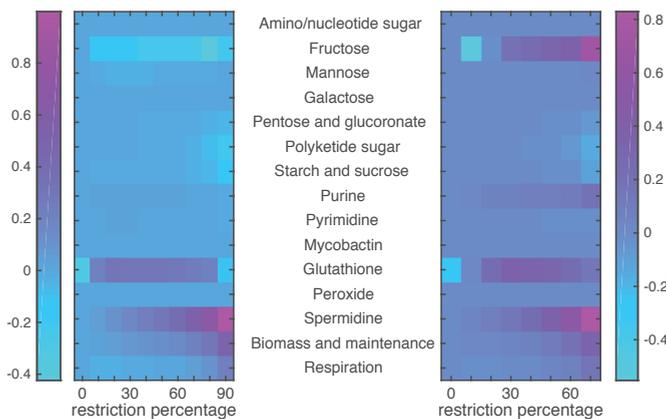
Figure 5.9 - Flux rerouting through sugar metabolism, nucleotide metabolism, and various other metabolic pathways upon application of drugs.
Heat maps indicate a predicted relative increase (pink) or decrease (light blue) through various pathways upon the application of a variety of drugs. The logarithm of the sum of all absolute fluxes is given per pathway (values are indicated in the color bars on either side), so information on directionality is not comprised. The x-axis indicates the percentage of restriction of the drug-affected reaction(s). **Left**: average of BTZ043, cycloserine/terizidone, ethambutol, ethionamide/propionamide, and isoniazid; **Right**: TMC207.

For TMC207, alanine, aspartate, glutamate, arginine, and proline metabolism is relatively unimportant (Figure 5.7), while porphyrin metabolism and fructose metabolism become more important, as the percentage of constriction of the reactions affected by TMC207 increases (Figures 5.8-5.9). Porphyrins are heterocyclic compounds able to form metal complexes, such as heme, the latter attenuating growth of Mtb if absent from the growth medium [37]. In addition, with moderate constriction (around 40% in Figures 5.8-5.9) flux is lowered through pathways such as cholesterol degradation and cofactor metabolism in general while flux trough these pathways is relatively large when the drug-affected reactions are constrained mildly (<10%) or heavily (>70%). Metabolism of glutathione, an antioxidant, shows behavior opposite to that of cholesterol degradation and cofactor metabolism in general (Figure 5.9).

The specific growth rates gradually approach zero upon the application of the drugs. The application of TMC207 is predicted to result in a relatively faster drop to zero growth rate (at 80% restriction of the flux through the affected reactions instead of 100%) (figure 5.10).
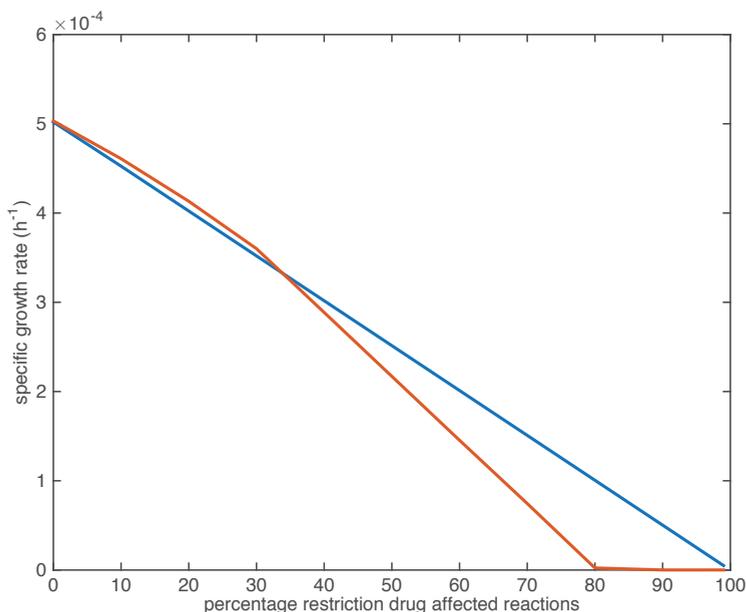
Figure 5.10 - Growth rate decline upon application of drugs.
Blue line: predicted average specific growth rate of Mtb upon application of BTZ043, cycloserine/terizidone, ethambutol, ethionamide/propionamide, or isoniazid.
Red line: predicted average specific growth rate of Mtb upon application of TMC207.

Discussion

Mtb is under considerable stress from the host during infection. This situation can worsen with the application of drugs. As the dosage of a metabolic drug(s) increases, so does the pressure on Mtb to circumvent the effects of the(se) drug(s) by rerouting metabolism. The percentage of restriction, as shown in Figures 5.7-5.10 can be viewed as a proxy for the drug dosage, as the effectiveness of a drug is dependent on its ability to interact with as many target enzymes as possible. As metabolism is an interconnected network, alternative metabolic states can exist to bypass the part of Mtb metabolism that is malfunctioning due to the effect of the drug(s).

For the simulations of metabolic states using sMtb-RECON, three different assumptions have been made. The first being that the non-growth associated maintenance flux is 0.1 mmol gDW^{-1} h^{-1} or higher, as this maintenance flux was shown to have the best fit to experimental data [9], the second being that the condition-specific biomass reaction is maximized and the third being that overall enzyme usage is minimized. The first requirement is set as a constraint, while

requirements two and three are captured as a bi-objective function in a bi-objective optimization problem. The weight factor ratio, f_r, between the condition-specific biomass reaction weight, f_b, and the total enzymatically catalyzed reaction weight, f_e, equals 0.8. By lowering this factor, more emphasis is put on the minimization of enzyme usage, while increasing this factor puts more emphasis on attaining a higher value for the condition-specific biomass reaction. Lowering this factor would result in a larger part of nutrients being used by Mtb to generate energy, while increasing this factor would result in the uptake profile looking more similar to the condition-specific biomass reaction itself. Lowering and raising f_r is however limited to a range wherein a single objective of the bi-objective optimization problem is not dominant over the other [14].

Even though the bi-objective optimization problem maximizes the specific biomass reaction and simultaneously minimizes the overall enzyme usage, this does not result in the direct uptake of all amino acids in a ratio that is proportional to the ratio of the corresponding coefficients in the condition-specific biomass reaction (Figure 5.6). As the direct uptake of amino acids represents a much shorter, and thus less enzyme intensive route, such a route would be preferred if the only function of the amino acids were direct incorporation into biomass. This is however not the case, because the amino acids are also used to generate energy in the form of ATP. Whether amino acids are required to synthesize biomass or are required to generate energy or both, it has been shown that Mtb is reliant on amino acids to thrive within the host [38-40].

Some nutrients are closely related and their interconversion involves the usage of one or only a few enzymes. These interlinked metabolites, such as glutamate/glutamine and TAG/DAG/phosphatidylcholine can be relatively easily substituted and predictions on their individual uptake and/or secretion rates can only be derived from the predictions of their combined uptake and/or secretion rate. This is especially visible in Figure 5.2, where the ranges of uptake and secretion rates are equal for both glutamate and glutamine.

Perhaps surprisingly, the metabolic states predicted with sMtb-Recon after perturbation with BTZ043, cycloserine/terizidone, ethambutol, ethionamide/propionamide, or isoniazid are all highly similar. For some drugs, such as isoniazid and ethionamide/propionamide, this can be explained by an overlapping enzyme target InhA. Inhibition of this enoyl acyl carrier protein (ACP) reductase is one of the most effective ways to eradicate Mtb. This enzyme catalyses 2-trans-enoyl ACP reduction and catalyzes the final step in fatty acid synthesis and is involved in mycolic acid synthesis. Inactivation of InhA results in cell wall alterations and eventually lysis of the cell [41]. However, the mycolic acid synthesis inhibition caused by isoniazid and ethionamide/propionamide is not directly related to the cell wall synthesis inhibiting effects of ethambutol and cycloserine/ terizidone.

Even though there is evidence that arabinogalactan and mycolic acids are physically attached to each other, this is not reflected in model sMtb-RECON [42]. Nevertheless, both mycolic acids and arabinogalactan are part of the condition-specific biomass reaction of Mtb. The condition-specific biomass reaction of Mtb is based on RNA sequencing data derived from *M. bovis* BCG 24 hours post infection [21]. The ratio between the metabolic precursors in this biomass reaction is constant within model sMtb-Recon. Therefore, a decrease in the ability to synthesize mycolic acids, which are part of the biomass reaction, by constraining the reactions catalyzed by InhA, results in a decreased maximally achievable value of the condition-specific biomass reaction. This in turn will result in a decreased arabinogalactan need to achieve this value. The minimization of enzyme usage ensures that the overall flux through the arabinogalactan synthesis pathway is minimized. This process works the other way around as well. A limitation of the flux through the arabinosyltransferase EmbB, that is required for the synthesis of arabinogalactan [43], will result in a lower value of the condition-specific biomass reaction, in turn leading to a lower need of mycolic acids. The mycolic acid synthesis pathway is a highly linear pathway, and completely unidirectional in model sMtb-RECON. No ATP for maintenance can be generated by mycolic synthesis in the model. The lower need of mycolic acids will result in the mycolic acid synthesis pathway being minimized as it serves no other purpose other than synthesizing mycolic acids for Mtb biomass as energy in the form of ATP cannot be generated from this pathway.

The reason the predicted metabolic state of TMC207-affected Mtb differs from the other drugs is due to the function of AtpE as an ATP generating enzyme involved in respiration. ATP is on the one hand a direct biomass precursor in the condition-specific biomass reaction, but it is also required to synthesize almost all other biomass precursors. In addition, ATP is required to satisfy the non-growth associated maintenance constraint. As such, the effect of constraining AtpE is not nearly as straightforward as the effect of constraining InhA or EmbB. This effect can be seen in Figure 5.10, wherein the effect of limiting AtpE, due to the application of TMC207, on the maximum condition-specific biomass reaction value is visible (red line). The line contains multiple bends, the most notable at 30%, but at 10% and 20% as well. These bends represent metabolic rerouting that can vary with the severity of constraining the respective reaction(s). An example of such variance can be seen in Figure 5.7, in the methyl citrate cycle. The flux through this cycle does not linearly increase or decrease with the constraining percentage at all, which can be more clearly seen when comparing the methyl citrate cycle in Figure 5.7 with a linear increase in biomass and maintenance, as seen in Figure 5.9. As TMC207 is predicted to have an effect that substantially differs from the effect of BTZ043, cycloserine/terizidone, ethambutol, ethionamide/propionamide, and

isoniazid, a combination of TMC207 and the latter drugs would probably provide a more effective strategy to combat TB than combinations of drugs without TMC207. Understanding the metabolic rerouting upon drug administration can lead to the identification of new metabolic bottlenecks, the identification of new targets and in the long run the development of new therapies based on combination of drugs. Moreover, detailed analysis of the mechanisms deployed by Mtb to counteract the impact of drugs might offer insights on the role of genetic modifications related to the development of drug resistances.

Model RECON 2.2 is a general model of human metabolism. A macrophage is however, a very specialized human phagocytic cell, which engulfs and digests pathogens in a specialized compartment, the phagosome. An important mechanism of pathogen killing by phagocytes involves generating the superoxide anion, which reacts with iron sulfur clusters in the pathogen, releasing iron and subsequently damaging DNA [44, 45]. RECON 2.2 does not have a phagosomal compartment and the applied metabolic state simulation strategy (Figure 5.1) assumes that all cytoplasmic biomass precursors for the macrophage are available for Mtb inside the phagosome, while the effect of or presence of other compounds is overlooked. The effect of oxygen radicals and resulting hydrogen peroxide is not captured by the approach applied in this study, which can be seen in Figure 5.9 where no change is visible in peroxide (degradation). The overall flux through peroxide degradation processes should increase relative to the flux through the condition-specific biomass reaction, assuming a more or less constant supply of superoxide anions by the macrophage.

Divalent metal cofactors such as iron, manganese, and zinc are essential for Mtb virulence [2]. Currently, only the iron requirement is reflected in model sMtb-Recon in the form of heme being an essential precursor for Mtb biomass. The metal availability in cells in general is limited and proteins compete for these metals [46]. Therefore, a better strategy would be to identify Mtb enzymes that require a certain metal cofactor and to simulate low availability of such a cofactor by constraining the total flux of all reactions associated with these enzymes [47]. This could provide a more accurate representation of the metabolic state of Mtb during infection, especially as macrophages are known to use high affinity iron binding proteins to limit the availability of iron [48], making this a promising modeling strategy.

The question remains whether these predictions are accurate enough to warrant pinpointing specific genes and their corresponding enzymes as drug targets. Previous modeling efforts have shown a poor predictive power of essential genes, using a bi-objective optimization strategy [14]. Continuous step-by-step improvements of Mtb models to reach one functional standardized model of Mtb metabolism is a solid step in this direction [13].

Materials and Methods

Mtb and human models of metabolism

We used the genome-scale metabolic model of *Mycobacterium tuberculosis* called sMtb, *in silico Mycobacterium tuberculosis* [14] which is represents a modification of the model presented in [9]. The GSM reconstruction of human metabolism RECON 2.2 [25] was used as a model representing the host.

Creating a combined sMtb-RECON model

From the biomass precursors of the biomass reaction of RECON 2.2, all precursors were selected that could be present in the cytoplasm. As Mtb is known to be able to escape from the phagosome to the cytosol [49] and no phagosomal compartment was present in RECON 2.2 we took all metabolic precursors from the cytoplasm as biomass precursors for the macrophage condition-specific biomass reaction. A macrophage condition-specific biomass reaction was created using this list of biomass precursors and the gene expression profile of the macrophage-like THP-1 cells [21]. The same method was applied to create an Mtb condition-specific biomass reaction using the gene expression profile of the Mtb-like *Mycobacterium bovis* BCG gene expression profile (Figure 5.1, up to the lower right panel). This condition-specific biomass reaction is used as a proxy for the number of available nutrients and their corresponding maximum uptake rates for Mtb in the combined sMtb-RECON model (Figure 5.1, middle right panel).

Constraining sMtb and RECON 2.2 with gene expression data

Model sMtb was constrained as described in [14] using raw sequence read data available in the EMBL-EBI European Nucleotide Archive under the Accession No. PRJEB6552, http://www.ebi.ac.uk/ena/data/view/PRJEB6552 for *M. bovis* BCG cells infecting THP-1 cells. RECON 2.2 was constrained in a similar manner.

Obtaining a condition-specific biomass reaction for sMtb and a phagosomal environment

After assigning counts to the reactions in both sMtb and RECON 2.2 a condition-specific biomass reaction was obtained for sMtb as described in [14]. For RECON 2.2 a similar approach was taken. A list of cytoplasmic biomass precursors, obtained from the biomass reaction present in RECON 2.2 was used for maximization of said precursors one-by-one, while keeping all uptake rates for model RECON 2.2 unconstrained. The resulting condition-specific biomass reaction, which is in essence a ratio between different cytoplasmic biomass precursors, was used as a proxy for the phagosomal composition. These biomass precursor values were subsequently linearly scaled such that all values range between 0 and 1 mmol/h by dividing each value by the largest value obtained.

Applying constraints and calculating a reference metabolic state

A reference metabolic state, representing non-drugged growth of Mtb in the phagosome, was calculated in a similar manner as previous efforts [14], using a bi-objective optimization method, with the exception of using the calculated phagosomal composition and its corresponding values as maximal allowable nutrient uptake rates. An f_r value of 0.8 was used for these calculations [14], wherein

$$f_r = \frac{f_b}{\sum_{i=1}^{n} f_{e,i}} \tag{5.1}$$

and $f_{e,i}$ represents the weight factor for enzymatically catalyzed reaction i and f_b represents the weight factor for the biomass reaction.

Creating a weights vector for each metabolic drug

A list of available metabolic drugs was created and the genes encoding the enzymes that are known or expected to be affected by these drugs were listed (Table 5.2). For each drug, i, a vector, $\mathbf{c}_{d,i}$, containing weights ranging between 0 and 1 was created. A 0 represents a non-affected reaction and a 1 represents a fully affected reaction. A reaction that, for example, is catalyzed by three isozymes, and from among these isozymes only one is affected by the respective drug, would receive a value of 0.33. Likewise, if two out of the three isozymes are affected by the drug, a value of 0.67 is be attributed to that reaction. On the other hand, if the reaction would be catalyzed by a complex of three enzymes, a value of 1 would always be attributed to the respective reaction if at least one enzyme in the complex would be affected by the respective drug.

Calculating drugged metabolic states for sMtb

After obtaining drug weight vectors, $\mathbf{c}_{d,i}$ for each drug, first, the bi-objective optimization problem as described in [14], was solved:

$$w = \max\left\{\left(\sum_{i=1}^{n} -f_{e,i} \cdot |v_{e,i}|\right) + f_b \cdot v_b\right\} \tag{5.2.1}$$

subject to:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b} \tag{5.2.2}$$
$$\mathbf{l} \leq \mathbf{v} \leq \mathbf{u} \tag{5.2.3}$$

Wherein w is the objective function value, $v_{e,i}$ represents the flux or rate of a reaction catalyzed by at least one enzyme; $f_{e,i}$ represents the weight factor for each

of those reactions; v_b represents the specific growth rate, i.e. the flux through one of the aforementioned (condition-specific) biomass reactions; f_b represents the weight factor for the biomass reaction; n is the total number of reactions catalyzed by at least one enzyme; $\mathbf{S}$ represents the stoichiometric matrix; $\mathbf{v}$ represents a vector with all fluxes (comprising $v_{e,i}$ and v_b); $\mathbf{b}$ represents a vector with zeros; $\mathbf{l}$ represents a vector with lower bounds for all fluxes and $\mathbf{u}$ represents a vector with upper bounds for all fluxes.

Afterwards, the bi-objective optimization problem is altered such that the objective function value that was obtained is set as a constraint. A new optimization problem is formulated to calculate the minimal flux through the reaction(s) affected by the respective drug:

5

$$w' = \min\left\{\mathbf{c}_{d,i} \cdot \mathbf{v}\right\} \tag{5.3.1}$$

subject to:

$$w = \left(\sum_{i=1}^{n} -f_{e,i} \cdot \left|v_{e,i}\right|\right) + f_b \cdot v_b \tag{5.3.2}$$

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b} \tag{5.3.3}$$
$$\mathbf{l} \leq \mathbf{v} \leq \mathbf{u} \tag{5.3.4}$$

Wherein w' represents the new objective function value and $\mathbf{c}_{d,i}$ represents the drug weight vector for the current drug, i. Finally, a third optimization problem is formulated:

$$w = \max\left\{\left(\sum_{i=1}^{n} -f_{e,i} \cdot \left|v_{e,i}\right|\right) + f_b \cdot v_b\right\} \tag{5.4.1}$$

subject to:

$$f_d \cdot w' = \mathbf{c}_{d,i} \cdot \mathbf{v} \tag{5.4.2}$$
$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b} \tag{5.4.3}$$
$$\mathbf{l} \leq \mathbf{v} \leq \mathbf{u} \tag{5.4.4}$$

Wherein f_d is a value that is gradually lowered from 1 to 0 to represent increasing drug dosages, wherein a value of 1 represents no drug is applied or total ineffectiveness of the drug, and a value of 0 represents total effectiveness of the drug.

References

1. Gengenbacher, M. and S.H.E. Kaufmann, Mycobacterium tuberculosis*: success through dormancy.* FEMS Microbiology Reviews, 2012. **36**(3): p. 514-532.
2. Zondervan, A.N., C.J. van Dam, J.P. Schaap, A.V. Martins dos Santos, and M. Suarez-Diez, *Regulation of Three Virulence Strategies of* Mycobacterium tuberculosis*: A Success Story.* International Journal of Molecular Sciences, 2018. **19**(2).
3. Orth, J.D., I. Thiele, and B.O. Palsson, *What is flux balance analysis?* Nat Biotechnol, 2010. **28**(3): p. 245-8.
4. Hanegraaf, P.P.F. and E.B. Muller, *The Dynamics of the Macromolecular Composition of Biomass.* Journal of Theoretical Biology, 2001. **212**(2): p. 237-251.
5. Beste, D.J., T. Hooper, G. Stewart, B. Bonde, C. Avignone-Rossa, M.E. Bushell, P. Wheeler, S. Klamt, A.M. Kierzek, and J. McFadden, *GSMN-TB: a web-based genome-scale network model of* Mycobacterium tuberculosis *metabolism.* Genome Biol, 2007. **8**(5): p. R89.
6. Jamshidi, N. and B.O. Palsson, *Investigating the metabolic capabilities of* Mycobacterium tuberculosis *H37Rv using the* in silico *strain iNJ661 and proposing alternative drug targets.* BMC Syst Biol, 2007. **1**: p. 26.
7. Lofthouse, E.K., P.R. Wheeler, D.J.V. Beste, B.L. Khatri, H. Wu, Tom A. Mendum, A.M. Kierzek, and J. McFadden, *Systems-based approaches to probing metabolic variation within the* Mycobacterium tuberculosis *complex.* PLoS ONE, 2013. **8**(9): p. e75913.
8. Vashisht, R., A. Bhat, S. Kushwaha, A. Bhardwaj, O. Consortium, and S. Brahmachari, *Systems level mapping of metabolic complexity in* Mycobacterium tuberculosis *to identify high-value drug targets.* Journal of Translational Medicine, 2014. **12**(1): p. 263.
9. Rienksma, R.A., M. Suarez-Diez, L. Spina, P.J. Schaap, and V.A.P. Martins dos Santos, *Systems-level modeling of mycobacterial metabolism for the identification of new (multi-) drug targets.* Seminars in Immunology, 2014. **26**(6): p. 610-622.
10. Ma, S., K.J. Minch, T.R. Rustad, S. Hobbs, S.-L. Zhou, D.R. Sherman, and N.D. Price, *Integrated Modeling of Gene Regulatory and Metabolic Networks in* Mycobacterium tuberculosis. PLOS Computational Biology, 2015. **11**(11): p. e1004543.
11. Garay, C.D., J.M. Dreyfuss, and J.E. Galagan, *Metabolic modeling predicts metabolite changes in* Mycobacterium tuberculosis. BMC Systems Biology, 2015. **9**: p. 57.
12. Puniya, B.L., D. Kulshreshtha, I. Mittal, A. Mobeen, and S. Ramachandran, *Corrigendum: Integration of Metabolic Modeling with Gene Co-expression Reveals Transcriptionally Programmed Reactions Explaining Robustness in* Mycobacterium tuberculosis. Scientific Reports, 2016. **6**: p. 24916.
13. Kavvas, E.S., Y. Seif, J.T. Yurkovich, C. Norsigian, S. Poudel, W.W. Greenwald, S. Ghatak, B.O. Palsson, and J.M. Monk, *Updated and standardized genome-scale reconstruction of* Mycobacterium tuberculosis *H37Rv, iEK1011, simulates flux states indicative of physiological conditions.* BMC Systems Biology, 2018. **12**: p. 25.

14. Rienksma, R.A., P.J. Schaap, V.A.P. Martins dos Santos, and M. Suarez-Diez, *Modeling the Metabolic State of Mycobacterium tuberculosis Upon Infection.* Frontiers in Cellular and Infection Microbiology, 2018. **8**: p. 264.

15. Beste, D., T. Hooper, G. Stewart, B. Bonde, C. Avignone-Rossa, M. Bushell, P. Wheeler, S. Klamt, A. Kierzek, and J. McFadden, *GSMN-TB: a web-based genome-scale network model of* Mycobacterium tuberculosis *metabolism.* Genome Biol, 2007. **8**: p. R89.

16. Bhat, A., R. Vashisht, and N. Chandra, *Modeling metabolic adjustment in* Mycobacterium tuberculosis *upon treatment with isoniazid.* Systems and Synthetic Biology, 2010. **4**(4): p. 299-309.

17. Lerner, T.R., C.J. Queval, A. Fearns, U. Repnik, G. Griffiths, and M.G. Gutierrez, *Phthiocerol dimycocerosates promote access to the cytosol and intracellular burden of* Mycobacterium tuberculosis *in lymphatic endothelial cells.* BMC Biology, 2018. **16**(1): p. 1.

18. Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan, and G.K. Schoolnik, *Transcriptional Adaptation of* Mycobacterium tuberculosis *within Macrophages: Insights into the Phagosomal Environment.* J Exp Med, 2003. **198**(5): p. 693-704.

19. Zimmermann, M., M. Kogadeeva, M. Gengenbacher, G. McEwen, H.-J. Mollenkopf, N. Zamboni, S.H.E. Kaufmann, and U. Sauer, *Integration of Metabolomics and Transcriptomics Reveals a Complex Diet of* Mycobacterium tuberculosis *during Early Macrophage Infection.* mSystems, 2017. **2**(4).

20. Beste, Dany J.V., K. Nöh, S. Niedenführ, Tom A. Mendum, Nathaniel D. Hawkins, Jane L. Ward, Michael H. Beale, W. Wiechert, and J. McFadden, *^{13}C-Flux Spectral Analysis of Host-Pathogen Metabolism Reveals a Mixed Diet for Intracellular* Mycobacterium tuberculosis. Chemistry & Biology, 2013. **20**(8): p. 1012-1021.

21. Rienksma, R.A., M. Suarez-Diez, H.-J. Mollenkopf, G.M. Dolganov, A. Dorhoi, G.K. Schoolnik, V.A.P. Martins dos Santos, S.H.E. Kaufmann, P.J. Schaap, and M. Gengenbacher, *Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing.* BMC Genomics, 2015. **16**(34).

22. Bordbar, A., N.E. Lewis, J. Schellenberger, B.O. Palsson, and N. Jamshidi, *Insight into human alveolar macrophage and* M. tuberculosis *interactions via metabolic reconstructions.* Mol Syst Biol, 2010. **6**: p. 422.

23. Duarte, N.C., S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, and B.Ø. Palsson, *Global reconstruction of the human metabolic network based on genomic and bibliomic data.* Proceedings of the National Academy of Sciences, 2007. **104**(6): p. 1777-1782.

24. Thiele, I., N. Swainston, R.M.T. Fleming, A. Hoppe, S. Sahoo, M.K. Aurich, H. Haraldsdottir, M.L. Mo, O. Rolfsson, M.D. Stobbe, S.G. Thorleifsson, R. Agren, C. Bolling, S. Bordel, A.K. Chavali, P. Dobson, W.B. Dunn, L. Endler, D. Hala, M. Hucka, D. Hull, D. Jameson, N. Jamshidi, J.J. Jonsson, N. Juty, S. Keating, I. Nookaew, N. Le Novere, N.

Malys, A. Mazein, J.A. Papin, N.D. Price, E. Selkov Sr, M.I. Sigurdsson, E. Simeonidis, N. Sonnenschein, K. Smallbone, A. Sorokin, J.H.G.M. van Beek, D. Weichart, I. Goryanin, J. Nielsen, H.V. Westerhoff, D.B. Kell, P. Mendes, and B.O. Palsson, *A community-driven global reconstruction of human metabolism.* Nat Biotech, 2013. **31**(5): p. 419-425.

25. Swainston, N., K. Smallbone, H. Hefzi, P.D. Dobson, J. Brewer, M. Hanscho, D.C. Zielinski, K.S. Ang, N.J. Gardiner, J.M. Gutierrez, S. Kyriakopoulos, M. Lakshmanan, S. Li, J.K. Liu, V.S. Martínez, C.A. Orellana, L.-E. Quek, A. Thomas, J. Zanghellini, N. Borth, D.-Y. Lee, L.K. Nielsen, D.B. Kell, N.E. Lewis, and P. Mendes, *Recon 2.2: from reconstruction to model of human metabolism.* Metabolomics, 2016. **12**: p. 109.

26. Matta, S.K. and D. Kumar, *Hypoxia and classical activation limits* Mycobacterium tuberculosis *survival by Akt-dependent glycolytic shift in macrophages.* Cell Death Discovery, 2016. **2**: p. 16022.

27. Prosser, G., J. Brandenburg, N. Reiling, C.E. Barry, 3rd, R.J. Wilkinson, and K.A. Wilkinson, *The bacillary and macrophage response to hypoxia in tuberculosis and the consequences for T cell antigen recognition.* Microbes and infection, 2017. **19**(3): p. 177-192.

28. Tullius, M.V., G. Harth, and M.A. Horwitz, *Glutamine Synthetase GlnA1 Is Essential for Growth of* Mycobacterium tuberculosis *in Human THP-1 Macrophages and Guinea Pigs.* Infection and Immunity, 2003. **71**(7): p. 3927-3936.

29. Tedeschi, P.M., E.K. Markert, M. Gounder, H. Lin, D. Dvorzhinski, S.C. Dolfi, L.L.Y. Chan, J. Qiu, R.S. DiPaola, K.M. Hirshfield, L.G. Boros, J.R. Bertino, Z.N. Oltvai, and A. Vazquez, *Contribution of serine, folate and glycine metabolism to the ATP, NADPH and purine requirements of cancer cells.* Cell death & disease, 2013. **4**(10): p. e877-e877.

30. Amelio, I., F. Cutruzzolá, A. Antonov, M. Agostini, and G. Melino, *Serine and glycine metabolism in cancer.* Trends in Biochemical Sciences, 2014. **39**(4): p. 191-198.

31. Eisenreich, W., T. Dandekar, J. Heesemann, and W. Goebel, *Carbon metabolism of intracellular bacterial pathogens and possible links to virulence.* Nature Reviews Microbiology, 2010. **8**: p. 401.

32. Pieters, J. and J.D. McKinney, *Pathogenesis of* Mycobacterium tuberculosis *and its Interaction with the Host Organism*, ed. F. Melchers. 2013, Berlin: Springer. 243.

33. Bopape, M.C., H.C. Steel, R. Cockeran, N.M. Matlola, P.B. Fourie, and R. Anderson, *Antimicrobial activity of clofazimine is not dependent on mycobacterial C-type phospholipases.* Journal of Antimicrobial Chemotherapy, 2004. **53**(6): p. 971-974.

34. Daniel, J., H. Maamar, C. Deb, T.D. Sirakova, and P.E. Kolattukudy, Mycobacterium tuberculosis *uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy-like phenotype in lipid-loaded macrophages.* PLoS Pathog, 2011. **7**(6): p. e1002093.

35. Siegrist, M.S., M. Unnikrishnan, M.J. McConnell, M. Borowsky, T.Y. Cheng, N. Siddiqi, S.M. Fortune, D.B. Moody, and E.J. Rubin, *Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition.* Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18792-7.

36. Griffin, J.E., J.D. Gawronski, M.A. Dejesus, T.R. Ioerger, B.J. Akerley, and C.M. Sassetti, *High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism.* PLoS Pathog, 2011. **7**(9): p. e1002251.

37. Owens, C.P., N. Chim, and C.W. Goulding, *Insights on how the* Mycobacterium tuberculosis *heme uptake pathway can be used as a drug target.* Future medicinal chemistry, 2013. **5**(12): p. 10.4155/fmc.13.109.

38. Gouzy, A., G. Larrouy-Maumus, D. Bottai, F. Levillain, A. Dumas, J.B. Wallach, I. Caire-Brandli, C. de Chastellier, T.D. Wu, R. Poincloux, R. Brosch, J.L. Guerquin-Kern, D. Schnappinger, L.P. Sorio de Carvalho, Y. Poquet, and O. Neyrolles, Mycobacterium tuberculosis *exploits asparagine to assimilate nitrogen and resist acid stress during infection.* PLoS Pathog, 2014. **10**(2): p. e1003928.

39. Gouzy, A., G. Larrouy-Maumus, T.-D. Wu, A. Peixoto, F. Levillain, G. Lugo-Villarino, J.-L. Gerquin-Kern, L.P.S. de Carvalho, Y. Poquet, and O. Neyrolles, Mycobacterium tuberculosis *nitrogen assimilation and host colonization require aspartate.* Nat Chem Biol, 2013. **9**(11): p. 674-676.

40. Gouzy, A., Y. Poquet, and O. Neyrolles, *Amino acid capture and utilization within the* Mycobacterium tuberculosis *phagosome.* Future Microbiology, 2014. **9**(5): p. 631-637.

41. Duan, X., X. Xiang, and J. Xie, *Crucial components of mycobacterium type II fatty acid biosynthesis (Fas-II) and their inhibitors.* FEMS Microbiology Letters, 2014. **360**(2): p. 87-99.

42. Birch, H.L., L.J. Alderwick, A. Bhatt, D. Rittmann, K. Krumbach, A. Singh, Y. Bai, T.L. Lowary, L. Eggeling, and G.S. Besra, *Biosynthesis of mycobacterial arabinogalactan: identification of a novel* α*(1→3) arabinofuranosyltransferase.* Molecular Microbiology, 2008. **69**(5): p. 1191-1206.

43. Goude, R., A.G. Amin, D. Chatterjee, and T. Parish, *The Arabinosyltransferase EmbC Is Inhibited by Ethambutol in* Mycobacterium tuberculosis. Antimicrobial Agents and Chemotherapy, 2009. **53**(10): p. 4138.

44. Hurst, J.K., *What really happens in the neutrophil phagosome?* Free radical biology & medicine, 2012. **53**(3): p. 508-520.

45. Winterbourn, C.C. and A.J. Kettle, *Redox Reactions and Microbial Killing in the Neutrophil Phagosome.* Antioxidants & Redox Signaling, 2012. **18**(6): p. 642-660.

46. Foster, A.W., D. Osman, and N.J. Robinson, *Metal preferences and metallation.* The Journal of biological chemistry, 2014. **289**(41): p. 28095-28103.

47. Wegrzyn, A.B., S. Stolle, R.A. Rienksma, V.A.P. Martins dos Santos, B.M. Bakker, and M. Suarez-Diez, *Cofactors revisited – Predicting the impact of flavoprotein-related diseases on a genome scale.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2019. **1865**(2): p. 360-370.

48. Kurthkoti, K., H. Amin, M.J. Marakalala, S. Ghanny, S. Subbian, A. Sakatos, J. Livny, S.M. Fortune, M. Berney, and G.M. Rodriguez, *The Capacity of* Mycobacterium tuberculosis *To Survive Iron Starvation Might Enable It To Persist in Iron-Deprived Microenvironments of Human Granulomas.* mBio, 2017. **8**(4).

5

49. van der Wel, N., D. Hava, D. Houben, D. Fluitsma, M. van Zon, J. Pierson, M. Brenner, and P.J. Peters, M. tuberculosis *and* M. leprae *Translocate from the Phagolysosome to the Cytosol in Myeloid Cells.* Cell, 2007. **129**(7): p. 1287-1298.
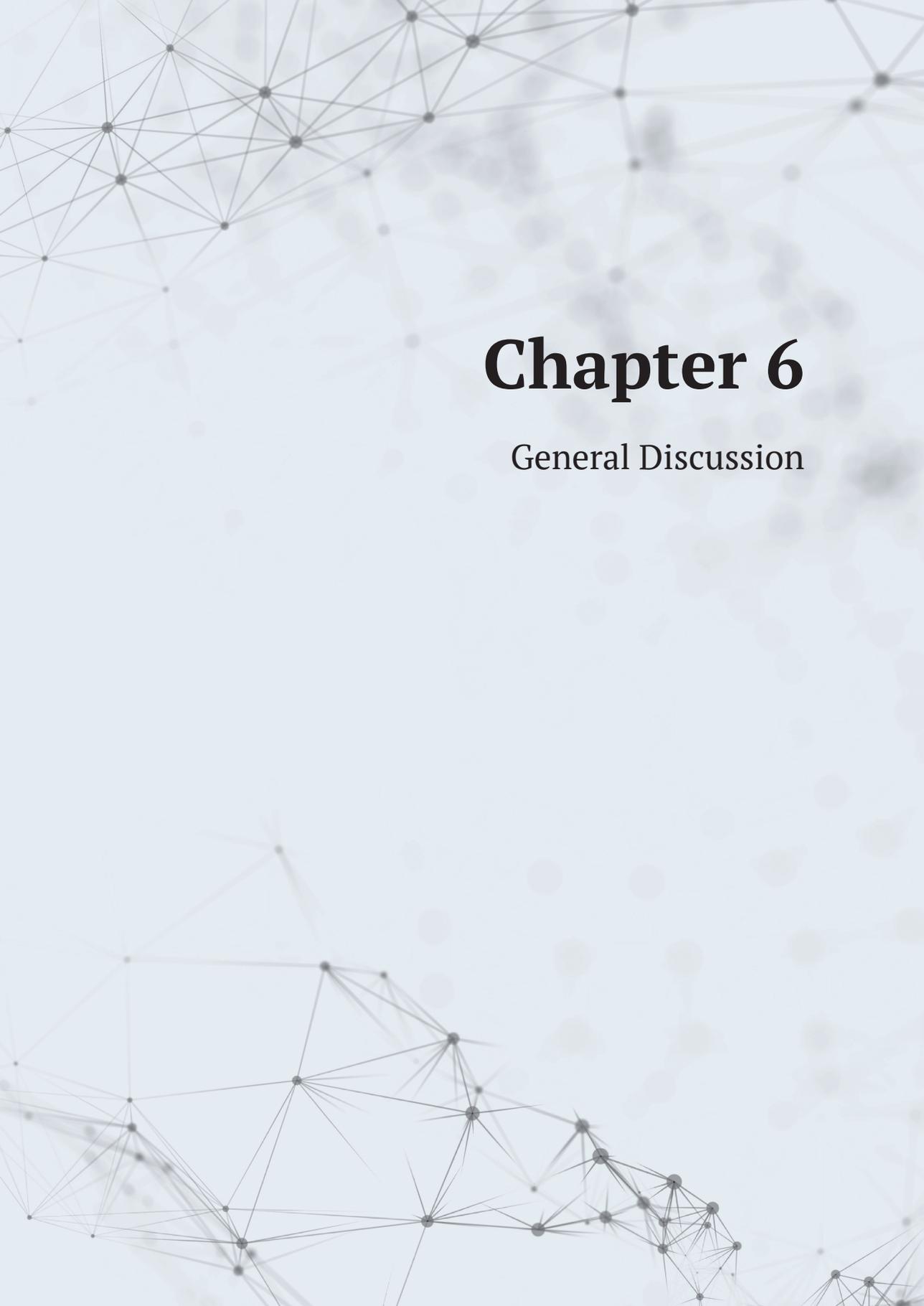
Chapter 6

General Discussion

Systems medicine

Systems medicine is a novel systems approach in medicine, which aims to improve diagnosis, targeted therapy, better prognosis and prevention. A central aspect of systems medicine is the integration of different sources of data [1]. A metabolic network is an example of a system that can be used to integrate these different sources of data. About 1200 genes out of 4000 Mtb genes are estimated to be involved in metabolism [2]. The reactions catalyzed by the enzyme products of this part of the genome can be modeled in a single GSM as shown in **previous chapters**. In this way, a large reservoir of potential drug targets is captured in a single model. The general principle is to utilize the predictive power of such a GSM to point out bottleneck reactions in metabolism, i.e. edges in the metabolic network underlying the GSM that carry a relatively large flux, while optimizing the objective function value. The enzymes catalyzing these bottleneck reactions have various ligands of which at least one natural substrate is known; otherwise the enzyme would not be incorporated in the GSM. These ligand(s) can subsequently be used as a template for structure-based drug design for drugs that competitively inhibit the enzyme(s) in question [3]. This systems medicine use is, besides the use for metabolic engineering purposes, one of the main uses of GSM's in general [4].

It is currently not practically feasible to experimentally obtain a profile of fluxes of the entire metabolism of Mtb, also called the metabolic state of Mtb, which is required to pinpoint the aforementioned bottleneck reactions. During infection, experimentally acquiring a profile of nutrients that are taken up by Mtb is not even possible, let alone experimentally acquiring a metabolic state [5]. The use of GSM's to predict these bottleneck reactions represents a first step in the total drug development process. One important requirement to predict the metabolic state accurately is the establishment of a suitable objective function.

Metabolic state of Mtb during infection

A major focus point of this thesis is the formulation of a condition-specific biomass reaction that can be used in an objective function to predict the metabolic state of Mtb during infection. The objective function, together with the constraints applied, has been shown to widely predefine the degree of freedom in terms of specific pathway usage [6]. This implies that the solution space is greatly altered by the objective function(s) and the constraints. Hence, the prediction of bottleneck reactions is also likely to be affected by the objective function(s) and the constraints. Under growth conditions outside the human body, on defined media, it makes sense that Mtb spends it resources on growth only, as these conditions are stable and there is no need to spend resources on anything else. Inside the host, a hostile environment is encountered, shifting the focus from growth to 'survival', as pointed out in **chapters 4 and 5**.

Pointing out vulnerable pathways

The first step in pointing out bottleneck reactions, or more in general, bottleneck pathways, is to provide an up to date and fully functional model of Mtb metabolism.

Providing an up to date model: sMtb

Following a publication on a metabolic network of mycolic acid synthesis, MAP [7], two GSMs of Mtb metabolism, GSMN-TB [8] and iNJ661 [9], were independently created. iNJ661 was created by a group that specializes in constraint-based modeling, while GSMN-TB was created by a group that specializes in Mtb itself. While GSMN-TB was, in my view, the most comprehensive of both models, protons and water molecules were omitted and chemical formulas were absent. Basic references to metabolite databases were lacking, as well as any other information that could be used to identify metabolites. It's predecessor, GSMN-TB 1.1 [10] was an extended and curated version of model GSMN-TB, but it unfortunately did not contain references to metabolite databases as well. Nevertheless, I set out to combine MAP, GSMN-TB 1.1, and iNJ661, which involved a substantial amount of manual curation. This process greatly increased the number of genes from the 661 genes in model iNJ661 and the 759 genes in model GSMN-TB 1.1 to the 915 genes in model sMtb. Model sMtb has a relatively high percentage of reactions that are associated to genes (83% as compared to 75% for GSMN-TB 1.1 and 77% for iNJ661). The abundance of these gene-associated reactions provides for a high quality model and sMtb outperforms model GSMN-TB 1.1 and iNJ661 both in terms of metabolic state predictions as well as gene essentiality predictions. Inferred and predicted fluxes correlate quite well with a Pearson's correlation coefficient of 0.94 and gene essentiality predictions have a Matthews correlation coefficient (MCC) of 0.58 with gene essentiality data [11] as compared to an MCC of 0.49 for GSMN-TB 1.1 and an MCC of 0.28 for iNJ661. Note that these values are based on the numbers in Table 2.2 of chapter 2, and that the MCC calculated for sMtb is slightly higher than the MCC of 0.53 predicted by Kavvas and colleagues [12], the same holds true for the MCC values of iNJ661. Nevertheless, even under such well-defined conditions, sMtb cannot capture the fluxes through e.g. the pentose phosphate pathway, as shown in **chapter 2**. This pathway is interconnected with glycolysis, and highly conserved in many organisms [13], as such one expects it to carry flux under a wide range of conditions. The observation that sMtb is unable to recreate the fluxes of the pentose phosphate pathway inferred with ^{13}C metabolic flux analysis, gives an indication that the prediction of fluxes through pathways that are not directly interconnected with central carbon metabolism might be off as well. Another possibility could be a measurement error or an error in the model used to infer the ^{13}C metabolic fluxes.

Creating an in-host objective using gene expression data

Mtb is capable of halting maturation of the phagosome inside immune cells and thus providing a niche for the bacterium to thrive [14]. Measuring the uptake and/or secretion rates of Mtb within this niche is practically not feasible. Even getting a profile of the metabolites or nutrients that Mtb acquires within this niche is still a hurdle that has not been overcome [5].

Proteomic measurements are in theory a good representative of flux, as the regulatory processes between enzymes and fluxes are fewer than those between transcripts and fluxes and thus their 'distance' to flux is shorter. In addition, while enzymes overlap in function, they most often differ in primary structure between Mtb and host, allowing them to be pinpointed to the respective organism. Recent developments in purification techniques and proteomics allows the capture and analysis of Mycobacteria-containing vacuoles and the proteomic analysis of these vacuoles, thus generating an Mtb-specific phagosomal proteome [15]. The study on Mtb proteins within the host is however still hampered by the skewed ratio of host versus Mtb proteins [16].

Transcript abundances can be pinpointed to both host and pathogen, as shown in **chapter 3**. Frustratingly, the only true genome-wide dataset of Mtb during infection is still a transcriptomics dataset. The reason is that guanidinium isothiocyanate is used to directly penetrate the mycobacterial cell wall, stop the activity of intracellular ribonucleases, and stabilize the RNA. In addition, this does not lyse Mtb cells, but it does lyse the host cells [17]. The stabilization and simultaneous enrichment effect is the key factor to obtain a genome-wide transcript profile of Mtb residing inside the host macrophage. Pending a stabilization method for proteins, the proteome of Mtb within its host remains elusive, and the genome-wide data source that best represents fluxes will be transcript abundances.

These two data sources: nutrient availability and transcript abundance represent the sole sources of data on Mtb metabolism within the host to this date, and this data was thus used to create an objective function of Mtb within the host to be able to accurately infer a metabolic state of Mtb during infection.

Host models

A GSM of host metabolism could yield interesting insights once such a model is tailored to represent an infected host macrophage. I started with the idea to create a condition-specific biomass reaction for a host GSM, just as I had done for Mtb in **chapter 4**. Once created, the condition-specific biomass reaction could be coupled to another objective, such as minimizing enzyme usage, to obtain a bi-ojective optimization problem specifically for the host. Thereafter, a weighed sum of the host bi-objective function and the Mtb bi-objective function could be optimized, wherein different weights would represent different scenarios wherein either the

host or Mtb has the upper hand. For such an approach it is essential that both objectives of each bi-objective function counteract each other, as explained in **chapter 4** for Mtb, and that the quality of the GSM's employed is sufficient. The lack of any constraints on the uptake rates (apart from constraining the oxygen uptake rate, as in **chapter 4**) implies in particular that there should not be any loops present in the network that could somehow result in a net energy production. Varying the oxygen uptake rate in host model RECON 2.2 had no effect on the obtained objective function value when the host bi-objective optimization problem was optimized. Remedying errors in models that cause such unwanted behavior is a tedious procedure. The main problem is that these loops are not caused by an imbalance in atoms on the left and right hand side of any given reaction, but rather by having different options to obtain the same result, such as ATP driven transport and passive diffusion of the same metabolite, paving the way for unlimited ATP production. The loops that cause similar effects as the one mentioned here can be large and complex, and solving these errors involves an iterative process wherein the number of iterations is unknown beforehand.

6

Cholesterol catabolism

The link between cholesterol and tuberculosis has been known for several decades [18]. However, the notion that Mtb might catabolize cholesterol is far more recent [19]. The complete pathway of cholesterol catabolism in Mtb up to the degradation of the side chain and rings A and B is well known. Recently, a mechanism for the catabolism of steroid rings C and D has been proposed and partly validated [20]. Completion of this pathway in sMtb would result in an additional gain of 2 acetyl-CoA molecules, 1 propionyl-CoA molecule and 1 succinyl-CoA molecule, instead of requiring a sink for the compound consisting of steroid rings C and D and not gaining anything in terms of useful compounds out of these steroid rings.

The extension and curation of Mtb GSMs is an ongoing process, as is shown by Kavvas and colleagues [12]. The latest Mtb GSM is named iEK1011[1], of which sMtb is the primary base model. The independent selection of sMtb as a primary base model by another research group, verifies its quality. Model iEK1011 is shown to outperform its predecessors, including sMtb, and is therefore the state of the art GSM of Mtb metabolism. But even this model has not incorporated the full cholesterol catabolic pathway yet, reflecting the need of the whole Mtb modeling community to strive for the most comprehensive GSM of Mtb metabolism.

---

[1] There exists a peculiar habit in the modeling community to name GSM's according to a specific format, starting with the letter 'i', for *in silico*, followed by the author's first initial, the first letter of his last name, and the number of genes in the model. This habit, in my view, is very odd. First of all, the letter 'i' might just as well stand for *in vitro* or *in vivo*, hence, the letter 's' for *in silico*, would be a much better choice. Second, there exists not the slightest hint about which organism's metabolism has been modeled, which is remarkable, to say the least. A single author, or different authors with the same initials for that matter, who model two different organisms having the same number of metabolic genes, is a disaster just waiting to happen...

Optimal model size

Model iEK1011 currently, as its name suggests, encompasses 1011 genes. I estimated that about 1200 genes of the Mtb genome would be involved in metabolism. But what is exactly metabolism? Arguably, the synthesis of proteins by the action of ribosomes falls under metabolism and the transport of cell wall lipids towards the outer cellular surface. These two examples, and many more can be thought of, highlight that the line between metabolic and non-metabolic is a vague line. Therefore, the number of genes on its own in a metabolic reconstruction is not so relevant. The average number of genes per reaction, and the way these genes are distributed over the reactions however, is relevant.

Lumping reactions
A first approach to show that a metabolic reconstruction, and its resulting GSM is an improvement over previous reconstructions, *de facto* boils down to showing that the number of genes in the current reconstruction has increased relative to the previous reconstructions. If one counts the number of reactions associated with genes and divides this by the total number of reactions, the resulting fraction is an indication of the reactions that are actually known to happen inside the modeled microorganism (neglecting spontaneously occurring reactions), and as such this fraction gives a clearer indication of 'improvement' than the sheer number of genes or reactions on its own.

Another quality issue is that some pathways, such as fatty acid synthesis and elongation, involve repetitive steps carried out by the same enzyme. The choice whether or not to lump these reactions has an effect on the solution space. If the reactions are not lumped, a random sampling process would assign more sampling points to the relatively linear pathway, putting more weight on this enzyme as compared to others, resulting in a skewed solution space. The modeler thus influences the solution space based on his own choices, irrespective of the metabolism that is being modeled.

Solution space of sMtb
But what does this solution space actually represent in the case of Mtb? Mtb cells are notorious for their heterogeneity, as mycobacterial populations often contain large numbers of viable, but non-culturable cells (Mukamolova et al. 2003; Oliver 2010; Trevors 2011). In addition, mycobacterial cultures are heterogeneous due to their asymmetric cell division [21]. This leads to distinct cell populations with varying antibiotic susceptibility, which in turn might be related to persistent cells. It is not unimaginable that these subpopulations, and also the cells within each subpopulation, are in different metabolic states. So, if gene expression data from

in-host mycobacteria are used to model the metabolic state, like in **chapters 4 and 5**, this state likely reflects an average metabolic state composed of many different metabolic states. By using a continuous flow reactor to culture Mtb, a population may be obtained wherein the individuals are likely to be in a similar metabolic state [22], but this hardly possible for in-host mycobacteria. Here, the subpopulation problem is likely to persist. Moreover, when treating in-host Mtb with drugs, Mtb responds in two stages. First, the majority of Mtb cells are killed within a few days, thereafter, the killing rate drops dramatically, and a drug-tolerant, persistent population remains [23]. From the perspective of trying to find bottleneck reactions, to develop novel therapeutics, the metabolism of this persistent subpopulation would be much more interesting. The disregard of the existence of Mtb subpopulations by GSM's of Mtb metabolism could imply that the metabolic weak points of the drug tolerant population are overshadowed by a large population that would be readily eradicated by drugs.

When predicting a metabolic state of Mtb of a persistent subpopulation, it would thus make sense to hold back on constraining the model too much based on experimental data.

6

Obtaining experimental data

Various model experiments exist that represent infection and can be used to obtain experimental data. In general there is a balance between control over the experimental conditions and the degree of representation a real Mtb infection. Several different experimental models exist, each having their own strengths and weaknesses.

THP-1 cells and *M. bovis* BCG
The model experiment employed in **chapter 3** relies on putting *M. bovis* BCG into contact with THP-1 cells. After 24 hours of interaction and phagocytosis of *M. bovis* BCG by the macrophage-like THP-1 cells, guadinium thiocyanate is introduced and the transcriptional profile of BCG is stabilized, while the THP-1 cells are lysed. In this model, host-pathogen interaction is present, but dormancy does not occur because the *M. bovis* BCG cells are ultimately eradicated by the THP-1 cells.

Wayne model
A well known model is the Wayne model [24]. This model relies on slowly removing oxygen available to Mtb. Given enough time, the Mtb bacilli enter an anaerobic stage that is thought represent persistent Mtb. In addition, this model allows synchronous replication of the Mtb bacilli when oxygen is reintroduced, offering

opportunities to study cell cycle related mechanisms of Mtb. The model however does not capture host-pathogen interaction.

Latently infected humans

The Wayne model thus represents dormancy, but host cells are absent, and the model experiment in **chapter 3** represents Mtb-host interaction, but dormancy is absent. In an ideal model experiment, both Mtb-host interaction and dormancy would occur. An option would be to take a longue tissue sample of a latently infected human and immediately stabilize the transcript profile by introducing guanidinium isothiocyanate. It will however be difficult to obtain Mtb from such tissue samples, as latently infected humans do not show the symptoms associated with tuberculosis. Moreover, the number of Mtb bacilli would probably be very low, as they are already relatively low during active tuberculosis. Such an approach would be like looking for a needle in a haystack.

Nevertheless, dormant Mtb is able to infect THP-1 cells, postponing Mtb growth for 6 days post infection [25]. Such an *in vitro* model experiment combines both dormancy and infection and would perhaps be a good substitute for the model experiment employed in **chapter 3**.

Integrating gene expression data

A plethora of methods to integrate various -omics datasets and GSMs have been developed since genome scale metabolic networks have become available. Here, I will focus on the integration of transcriptomics datasets, as it is the only type of genome-wide data of Mtb that is available under infection conditions.

Some well-known methods to integrate transcriptomics data and GSMs are mentioned in **chapter 4**. All these methods differ slightly from each other, but their main focus is the same: shrinking the solution space based on transcript abundances of the various enzyme-encoding transcripts. The end result is, obviously, always a solution space that 'looks like' the transcriptomics data. The resulting GSM thus reflects a range of metabolic states (the solution space) that is smaller than the range of metabolic states obtained without the data. One of these metabolic states should be the actual, real metabolic state of the given organism at the moment and under the conditions of obtaining the transcriptomics data. So, actually a subset of metabolic states is selected based on the transcriptomics data, regardless of the chosen integration method. When predictions of metabolic states in similar (but not the same) conditions are subsequently made, it is not guaranteed that the real metabolic state falls within the selected subset of metabolic states. The condition-specific GSM with the selected subset of metabolic states most likely approaches

the real metabolic state, and as such gives a better description of metabolism as a GSM with a full size solution space would. However, this condition-specific GSM is unsuited to make predictions on metabolic states wherein the conditions wherein the given organism resides, differ from the conditions wherein the transcriptomics data were obtained. This nullifies the use of GSMs as predictive tools and merely makes transcript data look like fluxes. Therefore, I set out to develop an approach in **chapter 4** that captures both the information from the transcriptomics data, while retaining the full solution space.

Validation of in-host sMtb
Use of the method described in **chapter 4** results in a prediction of an uptake and secretion profile that fits the experimental data quite well. However, for the *in vivo* prediction of gene essentiality the story is completely different, as the gene essentiality predictions dramatically worsen. The MCC values obtained vary around 0 (ranging from -0.14 to 0.01), and are thus no better than a random prediction. There are several reasons that could cause gene essentiality predictions and experimental data [26] to disagree:
Mtb and its host continuously interact during infection, resulting in an absence of steady state [27]. Therefore, the fundamental principles on which GSMs rely are not valid.
The data is acquired 3 and 7 days after infection with Mtb, while the condition-specific biomass reaction is derived from RNA sequencing data of *M. bovis* BCG 24 hours after infection.
There is little correlation between fluxes and gene expression and/or the list of biomass precursors is too short, causing the condition-specific biomass reaction to not reflect the true biomass reaction.
The network topology of model sMtb is incomplete and important parts during infection are not captured.
The bi-objective function approach does not represent the real objective(s) of Mtb during infection.
Important constraints are neglected.
The method of predicting the essential genes is incorrect, as the bi-objective function approach necessitates a method that differs from one using a single objective function.
This list is by no means intended to be comprehensive; there could be other causes to the poor *in vivo* gene essentiality predictions.

Absence of steady state
The lack of a steady state is a very fundamental problem. Firstly, because the equations on which sMtb is based are not valid, and secondly, because the dynamic

interaction between Mtb and the host cause variation in gene essentiality over time, as out of the number of genes essential during infection only 78-80% overlap between 3 and 7 days after infection [26]. It is not unimaginable that a similar percentage of overlapping essential genes exists between 1 day after infection and 3 and 7 days after infection as well.

It is difficult to quantify the impact of the equations of sMtb being invalid in a dynamic situation on the predictions made using sMtb. Arguably, this impact is relatively unimportant for predicting essential genes, as this is a qualitative prediction and a relatively higher or lower flux through essential genes in general does not change the essentiality predictions. If data and predictions are obtained in a sufficiently small time frame, the dynamic situation inside the host can be viewed as a steady state situation. This, however, does render sMtb unsuited for predictions on changes in metabolic state during infection over larger time periods.

Are *M. bovis* BCG and Mtb the same?

Mtb and *M. bovis* BCG have a genome identity of 99.95%, so it would be tempting to suggest that both bacteria are virtually the same. However, the most obvious difference between *M. bovis* BCG and Mtb is the fact that Mtb is a human pathogen and *M. bovis* BCG is a vaccine strain [28]. When tuberculosis as a human disease is studied, it is obvious that *M. bovis* BCG is not suited as a study object. However, I set out to model metabolism of Mtb with the aim of using the model to ultimately develop new drugs. When the entire metabolism of Mtb is compared to the metabolism of *M. bovis* BCG, there are only minor differences [10]. So, if a metabolic drug is found that can eradicate *M. bovis* BCG, it is likely that it can eradicate Mtb as well. There are however obvious differences between Mtb and *M. bovis* BCG and some having an impact on metabolism are discussed below.

Pyrazinamidase
An example of a single basepair difference can found in the *pncA* gene encoding pyrazinamidase. *M. bovis* BCG has a specific point mutation of a C to G in position 169 of this gene [29]. The mutation of this basepair causes a histidine residue to be replaced by an asparagine at position 57 in the PncA enzyme. This trait is typical to *M. bovis* BCG and results in a loss of activity of the pyrazinamidase. The effective first line drug pyrazinamide (PZA) is a prodrug that is converted to its active form by this pyrazinamidase. The loss of activity thus has severe implications, as mycobacteria having this point mutation are resistant to PZA [30].

Cytochrome P450

Another difference between *M. bovis* BCG and Mtb lies within their regulation of cholesterol degradation. Cytochrome P450s Cyp125 and Cyp142 are encoded in the genomes of both *M. bovis* BCG and Mtb. These cytochrome P450s catalyze the first steps of the degradation of the side chain of cholest-4-en-3-one [19]. In *M. bovis* BCG *cyp142* contains a specific point mutation, resulting in a premature stop codon and a non-functional product. There is thus redundancy and thus backup of cholesterol oxidation capacity in Mtb as compared to *M. bovis* BCG [31].

Type VII ESX secretion systems

Would the key to unraveling the pathogenicity of Mtb lie within the differing 0.05% of its genome? While 0.05% of approximately 4 million base pairs merely leaves 2000 base pairs, the effect on the phenotype can vary wildly per base pair, as discussed above. Besides specific point mutations of these 2000 base pairs having a large effect on similar enzyme products of *M. bovis* BCG and Mtb, there are larger differences as well, such as whole genes being present in the genome of Mtb, while being absent in the genome of *M. bovis* BCG, and *vice versa*.

One such example is found in the type VII ESX secretion systems. There are five types of ESX secretion systems, ESX-1 to ESX-5. The genetic clusters encoding these ESX systems share a common set of features and these systems are used to transport molecular cargo across the mycobacterial envelope [32].

M. bovis BCG lacks the ESX-1 secretion system genes *espA*, *espC*, and *espD* [33-35]. There is increasing evidence that this system is used to access the cytosol already during the early stages of infection [36, 37]. The absence of these genes perhaps causes a difference in nutrient availability for Mtb and *M. bovis* BCG. This difference in nutrient availability may be an underlying cause of the poor gene essentiality predictions in **chapter 4**, as differences in nutrient availability set during simulations of the metabolic state have a large impact thereon.

Effects of genomic differences between *M. bovis* BCG and Mtb

The above-discussed genomic differences between *M. bovis* BCG and Mtb have varying effects on the transcriptomic differences between both bacteria. Arguably, when cholesterol is present in conditions wherein *M. bovis* BCG grows, *cyp142* is not expressed, as it encodes a non-functional product, whereas Mtb in under the same conditions likely would express *cyp142*, as in this case, the product would be functional. As the genomes of both bacteria encode the *cyp125* gene, they are both likely to express *cyp125* under conditions wherein cholesterol is present, and both would have the ability to degrade cholesterol. If any, the effect of this difference on the rest of the expression profile would probably by minor. The difference in the ESX-1 secretion system between *M. bovis* BCG and Mtb arguably has little

6

implications on the difference in expression profile under conditions wherein the host is absent, but could have major implications on the difference in expression profiles under intra-host growth conditions. The ESX-1 secretion system of Mtb could provide access to a whole range of nutrients that *M. bovis* BCG under the same conditions does not have access to. This would imply that the metabolic states of both bacteria could vary significantly under intra-host conditions. The condition-specific biomass reaction in **chapter 4** was formulated based on expression data of *M. bovis* BCG in THP-1 cells and thus might not be applicable to Mtb if the diet of Mtb is arguable much richer, as it could have access to cytosolic nutrients. Dual RNA sequencing performed on Mtb [5] might therefore result in an improved MCC regarding gene essentiality predictions.

Combining host and pathogen model

An effort was made in **chapter 5** to combine RECON 2.2 and sMtb to create a host-pathogen model. This was only partly successful as RECON 2.2 was of insufficient quality to apply the method described in **chapter 4** to the host part of this combined model, as mentioned before.

RECON represents an average metabolism of many different cells

Another point to consider is that the human body consists of many specialized cells, which are all markedly different. One of these cells is the macrophage, whose lifestyle and expression pattern is undoubtedly different from other types of human cells. Bordbar and colleagues [38] addressed this issue by referring to gene expression data from inactive macrophages, exchange reactions from primary literature, enzyme databases, immunohistological staining databases, and transport databases and using this information to tailor RECON 1, the first global human reconstruction [39]. The resulting macrophage model was essentially a subset of RECON 1.

A problem with such an approach is that it maintains parts of metabolism that are expressed under conditions measured, and it disposes parts of metabolism that could be expressed, but were not expressed under those conditions. In my view, this strategy is not suited to make metabolic state predictions in conditions that are not within the scope of the data used to create the macrophage model (which are likely the most interesting conditions). In principle, as each somatic cell within the human body contains the same genome, each cell could express every gene contained within this genome. Limiting the scope of possible metabolic states beforehand, by discarding part of this genome, will limit predicting unknown metabolic states. Therefore we opted for the method described in **chapter 5**.

Capturing effects of ESX secretion systems and siderophores

Another valid reason to create a host-pathogen model is to capture the metabolic interaction between the two. In general, one looks at the pathogen for the development of novel therapeutic interventions. One could however also look at the metabolism of the host and try to block or support certain metabolic fluxes in order to tip the scales in favor of the host.

Outlook: Improving Mtb GSM's

As can be seen in Figure 1.1 of the introduction, the GSM's of Mtb metabolism have gone through several iterations over the years. Currently, model iEK1011 is the latest GSM of Mtb metabolism, which is based on sMtb [12]. One could get the impression that there is nothing left to improve after this many iterations. This is, however, far from the truth, as there is ample opportunity to improve the quality the latest GSM of Mtb metabolism starting from this latest iteration, by for example taking cofactors and translation into account.

Two types of cofactors

A cofactor can be an inorganic metal ion or a more complex organic or metallo-organic compound [40]. There is a distinct difference between the two, as the first has to be acquired from the environment while the latter can in some cases be synthesized. Synthesizable cofactors, such as flavin adenine dinucleotide (FAD), are fully incorporated into model sMtb, as the compound is a metabolite in the metabolic network underlying sMtb. The same holds true for many other cofactors, such as various vitamins. For metal ion cofactors, it is a different story, as these are not included as metabolites in the reactions wherein they do function as a cofactor. An important metal ion cofactor is iron, which is known to be withdrawn by the host as a defense mechanism [41]. In addition to iron, the macrophage continuously exports other metal cofactors, such as manganese and zinc, out of the phagosome [42]. As these divalent metal ions are required for several metabolic enzymes to function, a lack thereof would likely impair this functioning. Such a lack cannot be captured using sMtb by merely lowering the uptake rate of the given divalent metal ion, as its effect on the activity of metal cofactor requiring enzymes is not incorporated in the reaction(s) catalyzed by the involved enzyme. An approach can be taken wherein the sum of fluxes catalyzed by metal cofactor requiring enzymes is constrained, thereby enabling the model to uphold the fluxes deemed most important for obtaining the objective, while simulating a metal cofactor shortage [40]. However, such an approach does not take into account the different affinities for their respective

cofactors that the enzymes might have. In a broader sense, the kinetics and turnover number of all enzymes in the model can be taken into account by using recently developed methods such as GECKO [43].

Taking translation into account

In **chapter 4 and 5** a bi-objective approach was taken to tackle the problem of unknown uptake rates. The bi-objective function contains two objectives that act in different directions. On the one hand, a specific growth rate, or (condition-specific) biomass reaction is maximized, while on the other hand, enzyme utilization is minimized. There are thus two opposing forces that balance one another.

ME-models couple gene expression to metabolism [44]. RNA and protein are removed from the biomass reaction and these are set as variables. Coupling constraints relate these variables to their catalytic functions in the cell. By setting an upper limit to the catalytic rates of RNA and protein producing machinery, there is an implicit upper limit to their associated metabolic fluxes, and thus an upper limit to growth. This approach circumvents the need for setting arbitrary bounds on the uptake and secretion rates. A major disadvantage is however, that catalytic rates of an entire range of enzymes need to be known or estimated. Nevertheless, this approach could prove valuable for modeling the effect of drugs on Mtb, as some well-known anti-TB drugs do not have a direct effect on an enzyme involved in metabolism, but rather have an effect on RNA polymerase (rifampicin) or ribosomal proteins (streptomycin). An ME-model of Mtb might be able to capture these effects and provide additional combinatorial drug targets.

Quantitative validation of GSMs

Although metabolic state predictions have a quantitative nature, i.e. flux values and directions are predicted, the validation of these models is often performed in a qualitative manner. While examples of quantitative validation of a handful of predicted fluxes by isotopic labeling experiments exist [6, 45, 46], qualitative validation remains the default. Validation of gene essentiality predictions is a preferred qualitative validation method, and many GSMs of Mtb metabolism have been validated in this way [8, 9, 12, 47, 48]. Limitations to the scope of metabolism covered by ^{13}C metabolic flux analysis imply that quantitative validation is limited as well, resulting in a tiny fraction of all predicted fluxes that can be validated using ^{13}C metabolic flux analysis (an estimated 1-5%).

Final remarks

In general, models are constructed based on experimental data. Later on, models are used to make predictions, which are validated by other experimental data. Once the validation has been performed, models can take over the effort associated with obtaining further experimental data by predicting the outcome of experiments.

For metabolic engineering purposes, a well-validated GSM can predict the effect of adding or removing genes associated with metabolic reactions via their enzyme products. If optimal production of a given compound is sought, such a GSM could provide the answer without the need for experimentation.

In systems medicine, there is but one goal, which is the eradication of the pathogen. There is no need to make the pathogen produce a certain compound, let alone optimize such production. Could one just take a single experimental dataset and come up with a strategy to eradicate the pathogen without constructing a GSM? In fact, genome-wide transcript abundances and gene essentiality data is available [5, 26] for Mtb residing in the host. The gene essentiality data is even used in **chapter 4**, to validate sMtb. It would not make sense to use sMtb to retroactively predict essential genes again. It would be more convenient to simply pick an essential gene or a highly expressed gene from the gene essentiality data or the genome-wide transcript abundance data and design an inhibitory compound for the corresponding enzyme. A GSM would not be required for such a purpose.

Systems medicine is a bit of an odd duck in that sense, as the data that you are interested in is likely to be used for validation. Its true value does not lie in the prediction of essential metabolic enzymes, but rather lies in organizing our knowledge of Mtb in a systematic way so we can make sense of the wealth of genome-scale data available. A GSM may function like a compendium of metabolic reactions and metabolites as well. This compendium is continuously updated by the scientific community and thus provides a valuable source of information. In general, a GSM can only point out things hidden in experimental data. It can serve as a tool to generate hypotheses from experimental data. As experimental datasets become increasingly elaborate and bulky, GSMs such as sMtb will be an essential part of many scientific projects to organize acquired data and generate new hypotheses.

References

1. Apweiler, R., T. Beissbarth, M.R. Berthold, N. Blüthgen, Y. Burmeister, O. Dammann, A. Deutsch, F. Feuerhake, A. Franke, J. Hasenauer, S. Hoffmann, T. Höfer, P.L.M. Jansen, L. Kaderali, U. Klingmüller, I. Koch, O. Kohlbacher, L. Kuepfer, F. Lammert, D. Maier, N. Pfeifer, N. Radde, M. Rehm, I. Roeder, J. Saez-Rodriguez, U. Sax, B. Schmeck, A. Schuppert, B. Seilheimer, F.J. Theis, J. Vera, and O. Wolkenhauer, *Whither systems medicine?* Experimental &Amp; Molecular Medicine, 2018. **50**: p. e453.

2. Rienksma, R.A., P.J. Schaap, V.A.P. Martins dos Santos, and M. Suarez-Diez, *Modeling the Metabolic State of* Mycobacterium tuberculosis *Upon Infection.* Frontiers in Cellular and Infection Microbiology, 2018. **8**: p. 264.

3. Raškevičius, V., V. Mikalayeva, I. Antanavičiūtė, I. Ceslevičienė, V.A. Skeberdis, V. Kairys, and S. Bordel, *Genome scale metabolic models as tools for drug design and personalized medicine.* PLOS ONE, 2018. **13**(1): p. e0190636.

4. Zhang, C. and Q. Hua, *Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine.* Frontiers in Physiology, 2016. **6**(413).

5. Zimmermann, M., M. Kogadeeva, M. Gengenbacher, G. McEwen, H.-J. Mollenkopf, N. Zamboni, S.H.E. Kaufmann, and U. Sauer, *Integration of Metabolomics and Transcriptomics Reveals a Complex Diet of* Mycobacterium tuberculosis *during Early Macrophage Infection.* mSystems, 2017. **2**(4).

6. Schuetz, R., L. Kuepfer, and U. Sauer, *Systematic evaluation of objective functions for predicting intracellular fluxes in* Escherichia coli. Mol Syst Biol, 2007. **3**: p. 119.

7. Raman, K., P. Rajagopalan, and N. Chandra, *Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs.* PLoS Comput Biol, 2005. **1**(5): p. e46.

8. Beste, D., T. Hooper, G. Stewart, B. Bonde, C. Avignone-Rossa, M. Bushell, P. Wheeler, S. Klamt, A. Kierzek, and J. McFadden, *GSMN-TB: a web-based genome-scale network model of* Mycobacterium tuberculosis *metabolism.* Genome Biol, 2007. **8**: p. R89.

9. Jamshidi, N. and B. Palsson, *Investigating the metabolic capabilities of* Mycobacterium tuberculosis *H37Rv using the* in silico *strain iNJ661 and proposing alternative drug targets.* BMC Syst Biol, 2007. **1**: p. 26.

10. Lofthouse, E.K., P.R. Wheeler, D.J.V. Beste, B.L. Khatri, H. Wu, Tom A. Mendum, A.M. Kierzek, and J. McFadden, *Systems-based approaches to probing metabolic variation within the* Mycobacterium tuberculosis *complex.* PLoS ONE, 2013. **8**(9): p. e75913.

11. Griffin, J.E., J.D. Gawronski, M.A. Dejesus, T.R. Ioerger, B.J. Akerley, and C.M. Sassetti, *High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism.* PLoS Pathog, 2011. **7**(9): p. e1002251.

12. Kavvas, E.S., Y. Seif, J.T. Yurkovich, C. Norsigian, S. Poudel, W.W. Greenwald, S. Ghatak, B.O. Palsson, and J.M. Monk, *Updated and standardized genome-scale reconstruction of* Mycobacterium tuberculosis *H37Rv, iEK1011, simulates flux states indicative of physiological conditions.* BMC Systems Biology, 2018. **12**: p. 25.

13. Stincone, A., A. Prigione, T. Cramer, M.M.C. Wamelink, K. Campbell, E. Cheung, V. Olin-Sandoval, N.-M. Grüning, A. Krüger, M.T. Alam, M.A. Keller, M. Breitenbach, K.M. Brindle, J.D. Rabinowitz, and M. Ralser, *The return of metabolism: biochemistry and physiology of the pentose phosphate pathway.* Biological reviews of the Cambridge Philosophical Society, 2015. **90**(3): p. 927-963.

14. Gengenbacher, M. and S.H.E. Kaufmann, Mycobacterium tuberculosis: *success through dormancy.* FEMS Microbiology Reviews, 2012. **36**(3): p. 514-532.

15. Hoffmann, E., A. Machelart, O.-R. Song, and P. Brodin, *Proteomics of* Mycobacterium *Infection: Moving towards a Better Understanding of Pathogen-Driven Immunomodulation.* Frontiers in Immunology, 2018. **9**(86).

16. Chande, A.G., Z. Siddiqui, M.K. Midha, V. Sirohi, S. Ravichandran, and K.V.S. Rao, *Selective enrichment of mycobacterial proteins from infected host macrophages.* Scientific Reports, 2015. **5**: p. 13430.

17. Mangan, J.A., I.M. Monahan, and P.D. Butcher, *Gene expression during host—pathogen interactions: Approaches to bacterial mRNA extraction and labelling for microarray analysis*, in *Methods in Microbiology*, N.D. Brendan Wren, Editor. 2002, Academic Press. p. 137-151.

18. Kondo, E., T. Murohashi, K. Kanai, and S. Kubota, *Esterification of tissue cholesterol with fatty acids in the lungs of tuberculous mice.* Japanese Journal of Medical Science and Biology, 1971. **24**(6): p. 345-356.

19. Ouellet, H., J.B. Johnston, and P.R. de Montellano, *Cholesterol catabolism as a therapeutic target in* Mycobacterium tuberculosis. Trends Microbiol, 2011. **19**(11): p. 530-9.

20. Crowe, A.M., I. Casabon, K.L. Brown, J. Liu, J. Lian, J.C. Rogalski, T.E. Hurst, V. Snieckus, L.J. Foster, and L.D. Eltis, *Catabolism of the Last Two Steroid Rings in* Mycobacterium tuberculosis *and Other Bacteria.* mBio, 2017. **8**(2).

21. Aldridge, B.B., M. Fernandez-Suarez, D. Heller, V. Ambravaneswaran, D. Irimia, M. Toner, and S.M. Fortune, *Asymmetry and Aging of Mycobacterial Cells Lead to Variable Growth and Antibiotic Susceptibility.* Science, 2012. **335**(6064): p. 100.

22. Beste, D.J.V., M. Espasa, B. Bonde, A.M. Kierzek, G.R. Stewart, and J. McFadden, *The Genetic Requirements for Fast and Slow Growth in Mycobacteria.* PLoS ONE, 2009. **4**(4): p. e5349.

23. Jindani, A., V.R. Aber, E.A. Edwards, and D.A. Mitchison, *The Early Bactericidal Activity of Drugs in Patients with Pulmonary Tuberculosis.* American Review of Respiratory Disease, 1980. **121**(6): p. 939-949.

24. Wayne, L.G. and L.G. Hayes, *An in vitro model for sequential study of shiftdown of* Mycobacterium tuberculosis *through two stages of nonreplicating persistence.* Infection and Immunity, 1996. **64**(6): p. 2062-2069.

25. Iona, E., M. Pardini, M.C. Gagliardi, M. Colone, A.R. Stringaro, R. Teloni, L. Brunori, R. Nisini, L. Fattorini, and F. Giannoni, *Infection of human THP-1 cells with dormant* Mycobacterium tuberculosis. Microbes and Infection, 2012. **14**(11): p. 959-967.

6

26. Mendum, T., H. Wu, A. Kierzek, and G. Stewart, *Lipid metabolism and Type VII secretion systems dominate the genome scale virulence profile of* Mycobacterium tuberculosis *in human dendritic cells.* BMC Genomics, 2015. **16**(1): p. 372.

27. Rohde, K., M. Yates Robin, E. Purdy Georgiana, and G. Russell David, Mycobacterium tuberculosis *and the environment within the phagosome.* Immunological Reviews, 2007. **219**(1): p. 37-54.

28. O'Reilly, L.M. and C.J. Daborn, *The epidemiology of* Mycobacterium bovis *infections in animals and man: A review.* Tubercle and Lung Disease, 1995. **76**: p. 1-46.

29. Scorpio, A., D. Collins, D. Whipple, D. Cave, J. Bates, and Y. Zhang, *Rapid differentiation of bovine and human tubercle bacilli based on a characteristic mutation in the bovine pyrazinamidase gene.* Journal of Clinical Microbiology, 1997. **35**(1): p. 106.

30. Juréen, P., J. Werngren, J.-C. Toro, and S. Hoffner, *Pyrazinamide resistance and* pncA *gene mutations in* Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy, 2008. **52**(5): p. 1852-1854.

31. Driscoll, M.D., K.J. McLean, C. Levy, N. Mast, I.A. Pikuleva, P. Lafite, S.E. Rigby, D. Leys, and A.W. Munro, *Structural and biochemical characterization of* Mycobacterium tuberculosis *CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen.* J Biol Chem, 2010. **285**(49): p. 38270-82.

32. Chen, J.M., *Mycosins of the Mycobacterial Type VII ESX Secretion System: the Glue That Holds the Party Together.* mBio, 2016. **7**(6).

33. Mahairas, G.G., P.J. Sabo, M.J. Hickey, D.C. Singh, and C.K. Stover, *Molecular analysis of genetic differences between* Mycobacterium bovis *BCG and virulent* M. bovis. Journal of Bacteriology, 1996. **178**(5): p. 1274-1282.

34. Gey van Pittius, N.C., J. Gamieldien, W. Hide, G.D. Brown, R.J. Siezen, and A.D. Beyers, *The ESAT-6 gene cluster of* Mycobacterium tuberculosis *and other high G+C Gram-positive bacteria.* Genome Biology, 2001. **2**(10): p. research0044.1-research0044.18.

35. Gröschel, M.I., F. Sayes, R. Simeone, L. Majlessi, and R. Brosch, *ESX secretion systems: mycobacterial evolution to counter host immunity.* Nature Reviews Microbiology, 2016. **14**: p. 677.

36. Manzanillo, Paolo S., Michael U. Shiloh, Daniel A. Portnoy, and Jeffery S. Cox, Mycobacterium Tuberculosis *Activates the DNA-Dependent Cytosolic Surveillance Pathway within Macrophages.* Cell Host & Microbe, 2012. **11**(5): p. 469-480.

37. Simeone, R., A. Bobard, J. Lippmann, W. Bitter, L. Majlessi, R. Brosch, and J. Enninga, *Phagosomal Rupture by* Mycobacterium tuberculosis *Results in Toxicity and Host Cell Death.* PLOS Pathogens, 2012. **8**(2): p. e1002507.

38. Bordbar, A., N.E. Lewis, J. Schellenberger, B.O. Palsson, and N. Jamshidi, *Insight into human alveolar macrophage and* M. tuberculosis *interactions via metabolic reconstructions.* Mol Syst Biol, 2010. **6**: p. 422.

39. Duarte, N.C., S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, and B.Ø. Palsson, *Global reconstruction of the human metabolic network based on genomic and bibliomic data.* Proceedings of the National Academy of Sciences, 2007. **104**(6): p. 1777-1782.

40. Wegrzyn, A.B., S. Stolle, R.A. Rienksma, V.A.P. Martins dos Santos, B.M. Bakker, and M. Suarez-Diez, *Cofactors revisited – Predicting the impact of flavoprotein-related diseases on a genome scale.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2019. **1865**(2): p. 360-370.

41. Weinberg, E.D., *Iron withholding: a defense against infection and neoplasia.* Physiological Reviews, 1984. **64**(1): p. 65-102.

42. Zondervan, A.N., C.J. van Dam, J.P. Schaap, A.V. Martins dos Santos, and M. Suarez-Diez, *Regulation of Three Virulence Strategies of* Mycobacterium tuberculosis*: A Success Story.* International Journal of Molecular Sciences, 2018. **19**(2).

43. Sánchez Benjamín, J., C. Zhang, A. Nilsson, P.J. Lahtvee, J. Kerkhoven Eduard, and J. Nielsen, *Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints.* Molecular Systems Biology, 2017. **13**(8): p. 935.

44. O'Brien, E.J., J.A. Lerman, R.L. Chang, D.R. Hyduke, and B.Ø. Palsson, *Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction.* Molecular Systems Biology, 2013. **9**(1).

45. Puchałka, J., M.A. Oberhardt, M. Godinho, A. Bielecka, D. Regenhardt, K.N. Timmis, J.A. Papin, and V.A.P. Martins dos Santos, *Genome-Scale Reconstruction and Analysis of the* Pseudomonas putida *KT2440 Metabolic Network Facilitates Applications in Biotechnology.* PLoS Computational Biology, 2008. **4**(10): p. e1000210.

46. Rienksma, R.A., M. Suarez-Diez, L. Spina, P.J. Schaap, and V.A.P. Martins dos Santos, *Systems-level modeling of mycobacterial metabolism for the identification of new (multi-) drug targets.* Seminars in Immunology, 2014. **26**(6): p. 610-622.

47. Ma, S., K.J. Minch, T.R. Rustad, S. Hobbs, S.-L. Zhou, D.R. Sherman, and N.D. Price, *Integrated Modeling of Gene Regulatory and Metabolic Networks in* Mycobacterium tuberculosis. PLOS Computational Biology, 2015. **11**(11): p. e1004543.

48. Vashisht, R., A. Bhat, S. Kushwaha, A. Bhardwaj, O. Consortium, and S. Brahmachari, *Systems level mapping of metabolic complexity in* Mycobacterium tuberculosis *to identify high-value drug targets.* Journal of Translational Medicine, 2014. **12**(1): p. 263.

6

Appendices

Summary

Approximately one-fourth of the *Mycobacterium tuberculosis* (Mtb) genome contains genes that encode enzymes directly involved in its metabolism. These enzymes represent potential drug targets that can be systematically probed with constraint based (CB) models through the prediction of genes essential (or the combination thereof) for the pathogen to grow. However, gene essentiality depends on the growth conditions and, so far, no *in vitro* model precisely mimics the host at the different stages of mycobacterial infection, limiting model predictions. A first step in creating such a model is a thoroughly curated and extended genome-scale CB metabolic model of Mtb metabolism. The history of genome-scale CB models of Mtb metabolism up to model sMtb are discussed and sMtb is quantitatively validated using ^{13}C measurements.

The human pathogen Mtb has the capacity to escape eradication by professional phagocytes. During infection, Mtb resists the harsh environment of phagosomes and actively manipulates macrophages and dendritic cells to ensure prolonged intracellular survival. In contrast to many other intracellular pathogens, it has remained difficult to capture the transcriptome of mycobacteria during infection due to an unfavorable host-to-pathogen ratio.

The human macrophage-like cell line THP-1 was infected with the attenuated Mtb surrogate *Mycobacterium bovis* Bacillus Calmette–Guérin (*M. bovis* BCG). Mycobacterial RNA was up to 1000-fold underrepresented in total RNA preparations of infected host cells. By combining microbial enrichment with specific ribosomal RNA depletion the transcriptional responses of host and pathogen during infection were simultaneously analyzed using dual RNA sequencing. Mycobacterial pathways for cholesterol degradation and iron acquisition are upregulated during infection. In addition, genes involved in the methylcitrate cycle, aspartate metabolism and recycling of mycolic acids are induced. In response to *M. bovis* BCG infection, host cells upregulate *de novo* cholesterol biosynthesis presumably to compensate for the loss of this metabolite by bacterial catabolism.

By systematically probing the metabolic network underpinning sMtb, the reactions that are essential for Mtb are identified. A majority of these reactions are catalyzed by enzymes and thus represent candidate drug targets to fight an Mtb infection. Modeling the behavior of the bacteria during infection requires knowledge of the so-called biomass reaction that represents bacterial biomass composition. This composition varies in different environments or bacterial growth phases. Accurate modeling of all fluxes through metabolism under a given condition at a moment

in time, the so called metabolic state, requires a precise description of the biomass reaction for the described condition.

The transcript abundance data obtained by dual RNA sequencing was used to develop a straightforward and systematic method to obtain a condition-specific biomass reaction for Mtb during *in vitro* growth and during infection of its host. The method described herein is virtually free of any pre-set assumptions on uptake rates of nutrients, making it suitable for exploring environments with limited accessibility.

The condition-specific biomass reaction represents the 'metabolic objective' of Mtb in a given environment (in-host growth and growth on defined medium) at a specific time point, and as such allows modeling the bacterial metabolic state in these environments. Five different biomass reactions were used predict nutrient uptake rates and gene essentiality. Predictions were subsequently compared to available experimental data. Nutrient uptake can accurately be predicted, but accurate gene essentiality predictions remain difficult to obtain.

By combining sMtb and a model of human metabolism, model sMtb-RECON was developed and used to predict the metabolic state of Mtb during infection of the host. Amino acids are predicted to be used for energy production as well as biomass formation. Subsequently the effect of increasing dosages of drugs, targeting metabolism, on the metabolic state of the pathogen was assessed and resulting metabolic adaptations and flux rerouting through various pathways is predicted.

In particular, the TCA cycle becomes more important upon drug application, as well as alanine, aspartate, glutamate, proline, arginine and porphyrin metabolism, while glycine, serine and threonine metabolism become less important for survival. Notably, an effect of eight out of eleven metabolically active drugs could be recreated and two major profiles of the metabolic state were predicted. The profiles of the metabolic states of Mtb affected by the drugs BTZ043, cycloserine and its derivative terizidone, ethambutol, ethionamide, propionamide, and isoniazid were very similar, while TMC207 is predicted to have quite a different effect on metabolism as it inhibits ATP synthase and therefore indirectly interferes with a multitude of metabolic pathways.

Samenvatting

Ongeveer een kwart van het genoom van *Mycobacterium tuberculosis* (Mtb) bevat genen die voor enzymen coderen die rechtstreeks betrokken zijn bij het metabolisme. Deze enzymen vertegenwoordigen mogelijke doelen voor geneesmiddelen die met constraint-based (CB) modellen systematisch onderzocht kunnen worden door het voorspellen van genen (of combinaties daarvan) die essentieel zijn voor de groei van de pathogeen. Echter hangt de essentialiteit van genen af van de groeiomstandigheden en tot nu toe bestaat er geen *in vitro* model dat de gastheer in de verschillende stadia van infectie nauwkeurig nabootst, wat modelvoorspellingen beperkt. Een eerste stap in het creëren van een dergelijk model is een grondig georganiseerd en uitvoerig CB-model van Mtb metabolisme op genoom-schaal. De geschiedenis van CB-modellen van het Mtb metabolisme tot aan het model sMtb worden besproken en sMtb wordt kwantitatief gevalideerd door gebruik te maken van ^{13}C-metingen.

De humane ziekteverwekker Mtb is in staat om aan uitroeiing door specialistische fagocyten te ontsnappen. Tijdens infectie weerstaat Mtb de zware omstandigheden in fagosomen en manipuleert op actieve wijze macrofagen en dendritische cellen om op die manier er voor te zorgen dat het langdurig intracellulair kan overleven. In tegenstelling tot vele andere intracellulaire pathogenen is het lastig gebleken om het transcriptoom van mycobacterium tijdens infectie vast te leggen vanwege de ongunstige verhouding tussen gastheercel en pathogeen.

De humane macrofaag-achtige cellijn THP-1 werd geïnfecteerd met de verzwakte Mtb-surrogaat *Mycobacterium bovis* Bacillus Calmette–Guérin (*M. bovis* BCG). Mycobacterieel RNA was tot een factor 1000 ondervertegenwoordigd in complete RNA-preparaten van geïnfecteerde gastheercellen. Door microbiële verrijking te combineren met specifieke ribosomale RNA depletie kon de transcriptionele respons van de gastheer en de pathogeen tijdens infectie simultaan worden geanalyseerd met behulp van duaal RNA-sequencen. Mycobacteriële metabole routes voor de afbraak van cholesterol en de verkrijging van ijzer zijn gestimuleerd tijdens infectie. Daarnaast worden genen die betrokken zijn bij de methylcitraatcyclus, het aspartaatmetabolisme en het hergebruik van mycolzuren geïnduceerd. Als reactie op *M. bovis* BCG infectie, stimuleren gastheercellen de *de novo* cholesterolbiosynthese, waarschijnlijk om het verlies van dit metaboliet door het bacteriële katabolisme te compenseren.

Door systematisch het onderliggende metabole netwerk van sMtb te onderzoeken zijn de reacties die essentieel zijn voor Mtb geïdentificeerd. Een meerderheid

van deze reacties worden gekatalyseerd door enzymen en zijn dus kandidaat om als geneesmiddelendoelwit te dienen om een Mtb-infectie te bestrijden. Het modelleren van het gedrag van de bacterie tijdens de infectie vereist kennis van de zogenaamde biomassareactie die de bacteriële biomassasamenstelling vertegenwoordigt. Deze samenstelling varieert in verschillende omgevingen of bacteriële groeifasen. Nauwkeurige modellering van alle fluxen door het metabolisme onder een bepaalde conditie op een bepaald moment, de zogenaamd metabole toestand, vereist een nauwkeurige beschrijving van de biomassareactie voor de beschreven conditie.

De transcripthoeveelheid-data verkregen door het duale RNA sequencen werd gebruikt om een eenvoudige en systematische methode te ontwikkelen voor het verkrijgen van een conditie-specifieke biomassareactie voor Mtb tijdens *in vitro* groei en tijdens infectie van de gastheer. De hierin beschreven methode vereist bijna geen vooraf genomen aannames over de opname van voedingsstoffen, waardoor het geschikt is voor het verkennen van omgevingen met beperkte toegankelijkheid.

De conditie-specifieke biomassareactie vertegenwoordigt de 'metabole doelstelling' van Mtb in een bepaalde omgeving (groei in de gastheer en groei op een bepaald medium) op een specifiek tijdstip, en maakt als zodanig modellering van de bacteriële metabole toestand in deze omgevingen mogelijk. Vijf verschillende biomassareacties werden gebruikt om de opname van voedingsstoffen en essentialiteit van genen te voorspellen. De voorspellingen werden vervolgens vergeleken met de beschikbare experimentele gegevens. De opname van voedingsstoffen kan nauwkeurig worden voorspeld, maar nauwkeurige voorspellingen van de essentialiteit van genen blijven moeilijk te verkrijgen.

Door het combineren van sMtb en een model van het humane metabolisme werd model sMtb-RECON ontwikkeld en gebruikt om de metabole toestand van Mtb tijdens de infectie van de gastheer te voorspellen. Er wordt voorspeld dat aminozuren gebruikt zullen worden voor zowel energieproductie als voor de vorming van biomassa. Vervolgens werd het effect van toenemende doseringen van geneesmiddelen, gericht op het metabolisme, op de metabole toestand van de pathogeen beoordeeld en de daaruit volgende metabole aanpassingen en fluxomleidingen via verschillende routes wordt voorspeld.

Met name de TCA-cyclus wordt belangrijker bij medicijngebruik, evenals alanine-, aspartaat-, glutamaat-, proline-, arginine- en porfyrine-metabolisme, terwijl glycine-, serine- en threonine-metabolisme minder belangrijk worden

voor overleving. Opmerkelijk was dat het effect van acht van de elf metabolisch actieve geneesmiddelen kon worden nagebootst en twee belangrijke profielen van de metabole toestand konden worden voorspeld. De profielen van de metabole toestanden van Mtb beïnvloed door de geneesmiddelen BTZ043, cycloserine en zijn derivaat terizidon, ethambutol, ethionamide, propionamide en isoniazide leken erg op elkaar, terwijl van TMC207 wordt voorspeld dat het een heel ander effect heeft op het metabolisme omdat het ATP synthase remt en daardoor indirect invloed heeft op vele metabole routes.

Acknowledgements

In 2011 begon ik aan een PhD project binnen het Europese SysteMTb consortium. Dit proefschrift vormt het resultaat van ruim vier jaar werk binnen dit consortium waar ik erg trots op ben. Doordat ik tussen het begin van mijn PhD project in 2011 en nu nog even had besloten om een opleiding tot Nederlands Octrooigemachtigde succesvol af te ronden, ligt dit proefschrift hier nu op tafel met enige vertraging.

Finishing a PhD project and having a full time job at the same time is quite a challenge. I would not have made it without Maria's help. Maria, we have first met during a congress in Prague in 2011 and I have come to know you as a women of her word. Many hours you have spent improving my thesis chapters and all the work that underlies these chapters. I can't thank you enough for that.

Begin 2011 wist ik nog niets van bioinformatica. Datzelfde jaar heb ik de cursus bioinformatica gevolgd onder leiding van Peter. Het volgende jaar heb ik geassisteerd bij diezelfde cusus en in 2015 heb ik een paper gepubliceerd over dual RNA sequencing, een bioinformatica onderwerp ten voeten uit. Peter, bedankt voor al je hulp bij dit onderwerp en het bijspijkeren van mijn kennis.

Ook wil ik Vitor bedanken voor zijn opportunisme en hulp bij het leggen van contacten en het regelen van deelprojecten binnen het SysteMTb project.

Further, I would like to thank my colleagues at the Max Planck Instititute for Infection Biology. Especially Martin and Hans, for trusting me to analyse results of their carefully conducted experiments, and for the productive and very nice cooperation.

Uiteraard wil ik ook mijn paranimfen Mark en Ruben bedanken. Mark, bedankt voor al je creatieve ideeen, je hulp met programmeren en de gezellige tijd in ons kantoor. Ruben, we kennen elkaar al jaren en ik ben blij dat jij mijn paranimf bent.

Verder wil ik alle collega's en vrienden van zowel SSB, MIB en het SysteMTb project bedanken voor de prettige tijd die ik tijdens mijn PhD project heb gehad.

Ten slotte wil ik nog mijn familie, vrienden en bovenal Laura bedanken voor alle steun die tot dit mooie proefschrift heeft geleid.

Bedankt!

Rienk

List of publications

Rienksma, R.A., M. Suarez-Diez, L. Spina, P.J. Schaap, and V.A.P. Martins dos Santos, *Systems-level modeling of mycobacterial metabolism for the identification of new (multi-)drug targets.* Seminars in Immunology, 2014. **26**(6): p. 610-622.

Rienksma, R.A., M. Suarez-Diez, H.-J. Mollenkopf, G.M. Dolganov, A. Dorhoi, G.K. Schoolnik, V.A.P. Martins dos Santos, S.H.E. Kaufmann, P.J. Schaap, and M. Gengenbacher, *Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing.* BMC Genomics, 2015. **16**(34).

Rienksma, R.A., P.J. Schaap, V.A.P. Martins dos Santos, and M. Suarez-Diez, *Modeling the Metabolic State of* Mycobacterium tuberculosis *Upon Infection.* Frontiers in Cellular and Infection Microbiology, 2018. **8**: p. 264.

Wegrzyn, A.B., S. Stolle, **R.A. Rienksma**, V.A.P. Martins dos Santos, B.M. Bakker, and M. Suarez-Diez, *Cofactors revisited – Predicting the impact of flavoprotein-related diseases on a genome scale.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2019. **1865**(2): p. 360-370.

Rienksma, R.A., P.J. Schaap, V.A.P. Martins dos Santos, and M. Suarez-Diez, *Modeling host-pathogen interaction to elucidate the metabolic drug response of intracellular* Mycobacterium tuberculosis. Frontiers in Cellular and Infection Microbiology, Accepted for publication.

Overview of completed training activities

Discipline specific activities

1st Annual SysteMTb meeting	SysteMTb, Prague, Czech Republic, 2011
2nd Annual SysteMTb meeting[1]	SysteMTb, Stresa, Italy, 2012
Opening WCSB	WUR, Wageningen, NL, 2012
3rd Annual SysteMTb meeting[1,2]	SysteMTb, Berlin-Potsdam, Germany, 2013
4th Annual SysteMTb meeting[1,2]	SysteMTb, Mallorca, Spain, 2014
SB@NL[1,2]	NCSB, Maastricht, NL, 2015
YISB meeting[1]	NL consortium for SB, Arnhem, NL, 2011
NBC14[1]	NBV, Ede, NL, 2012
NCSB[1]	NCSB, Soesterberg, NL, 2012

[1] oral presentation
[2] poster presentation

General courses

VLAG introduction week	VLAG, Baarlo, NL, 2011
Competence assessment	WGS, Wageningen, NL, 2011
iTuby course	EMBL, Heidelberg, Germany, 2011
Techniques for writing and presenting a scientific paper	WGS, Wageningen, NL, 2011
Career assessment	WGS, Wageningen, NL, 2014
Communicative skills	CPO, Nijmegen, NL, 2016

Optional courses and activities

Preparing PhD research proposal	WUR, Wageningen, NL, 2011
PhD trip	MIB/SSB, Various, USA/Canada, 2014
PhD/PostDoc seminars	SSB, Wageningen, NL, 2011-2015
Bioinformation technology	WUR, Wageningen, NL, 2011