Systems biology of metabolism of *Mycobacterium tuberculosis*

Dany J V Besta and Johnjoe McFaddenb

aFaculty of Health and Medical Sciences, University of Surrey, Guildford GU2 7XH, United Kingdom. Phone: +44 (0)1483-686428. Fax: +44 (0)1483-300374. E-mail: d.beste@surrey.ac.uk

bFaculty of Health and Medical Sciences, Guildford GU2 7XH, United Kingdom. Phone: +44 13 1483-696494. Fax: +44 (0)1483-300374. E-mail: j.mcfadden@surrey.ac.uk

Abstract

Despite decades of research many aspects of the biology of *Mycobacterium tuberculosis* remain unclear and this is reflected in the antiquated tools available to treat and prevent tuberculosis and consequently this disease remains a serious public health problem. Important discoveries linking *M. tuberculosis*’s metabolism and pathogenesis has renewed interest in this area of research. Previous experimental studies were limited to the analysis of individual genes or enzymes whereas recent advances in computational systems biology and high throughput experimental technologies now allows metabolism to be studied on a genome scale. Here we discuss the progress being made in applying system level approaches to studying the metabolism of this important pathogen.

Introduction

*Mycobacterium tuberculosis* is an unusual bacterial pathogen which has the remarkable ability to cause both acute life threatening disease and also clinically latent infections which can persist for the lifetime of the human host. Unlike many pathogens *M. tuberculosis* does not rely on the production of specific toxins to cause disease but rather the secret of this bacterium’s great success seems to be the ability to adapt and survive within the changing and adverse environment provided by the human host during the course of an infection. It is becoming apparent that key to this adaptation is the metabolic reprogramming of *M. tuberculosis* during both the acute and chronic phase of TB disease.
and therefore a more complete understanding of mycobacterial metabolism remains a major goal of TB research.

Whilst recent increases in research funding has progressed our understanding of the basic biology of *M. tuberculosis* this has not yet impacted on the global TB trends which remain at staggering levels. A possible reason why it has been difficult to translate basic research into effective strategies for combating tuberculosis is that TB research has until recently, focused on studying individual parameters in isolation which can consequently result in an overestimation of the importance of these factors. This effect may be particularly profound for a persistent pathogen such as *M. tuberculosis* which lacks classical virulence factors. The systems biology framework which investigates the dynamic interactions of many components provides an alternative and complementary strategy to the more traditional reductionist approaches to TB research. This methodology has started to be applied to the metabolism of *M. tuberculosis* on a genome scale and promises to drive biological discovery in the TB research field by providing scaffolding for the interpretation of “omic” scale datasets, directing hypothesis driven discovery and also assisting in the identification of novel drug targets.

**Metabolic model building**

The ultimate goal of system biology approaches to studying TB is to construct a complete model of infection incorporating both the pathogen and host, but this is currently infeasible as the information about the different components to be included in the model is lacking. Studies with other organisms have demonstrated that metabolism is, by far, the best understood cellular network and is thereby an excellent starting point for a systems-based approach [6;11;14].

Ideally metabolic models would describe the dynamic behavior of *M. tuberculosis* metabolism using rate equations however this would require detailed kinetic parameters (*e.g.* $K_m$, $V_{max}$ and substrate/product concentrations) of all the participating reactions. This approach has the advantage of being capable of modeling dynamic systems but its major disadvantage is that the values of the parameters required are difficult to measure.
and are unavailable on a genome scale even for the extensively studied *Escherichia coli*. Kinetic models have therefore only been applied to examine the dynamics of small well-defined systems, such as glycolysis in *Escherichia coli* [4]. The alternative is to use constraints-based modeling approaches, such as flux balance analysis (FBA). Instead of modeling dynamic states, FBA models steady state systems where concentrations of metabolites are held constant. In this situation, metabolic fluxes obey simple stoichiometric relationships. For instance, the stoichiometric equation, $1 \text{A} + 2\text{B} \rightarrow 1\text{C}$, implies that one mole of A plus 2 moles of B will be converted to 1 mole of C. In FBA, the predicted rate (flux) at which this conversion takes place is constrained to a set of values that are consistent with measured parameters, such as substrate utilization or biomass production rates. For instance, if A is a substrate then its rate of utilization determines the rate of consumption of B and the rate of production of C. If C is involved in a single further reaction, for instance, $\text{C} \rightarrow \text{D}$, then the flux towards D will also be constrained to certain value by measurement of the uptake rate of A. Note that this approach only works if the rate of production and consumption of each internal metabolite is equal and therefore its concentration is unchanging. If the concentration of any internal metabolite is allowed to vary then the system can no longer be constrained by external fluxes and will undetermined by the data. It is for this reason that constraints-based methods, such as FBA, can only be applied to systems in steady-state, such as chemostat-grown cells.

The first step is constraint-based approaches is to construct a metabolic network. This is often inferred directly from an organism’s genome annotation, which predicts genes encoding enzymes with defined roles in well-described biochemical conversions. These enzyme steps can then be incorporated into pathways and integrated into a network to construct a genome-scale metabolic network. Constraint based modeling approaches can then be applied to the model to predict the metabolic capabilities or internal fluxes [23] [1] [10;15;22;25], given measured constraints. There are of course limitations to these approaches such as the requirement for steady state conditions. Also, since no consideration is made of either transcriptional, translational, metabolic regulation or enzyme kinetics the predictive capabilities of constraint based models are limited to
situations when these factors are not significantly influencing reaction rates [11]. Nevertheless, the approach has been successfully applied to predict the metabolic capabilities of many different cellular systems [23] [1] [10;15;22;25] and has also been used in metabolic engineering to optimize production [19;24].

Application of metabolic modeling approaches in tuberculosis is aided by the fact that metabolism is a reasonably well studied system even in mycobacteria. Indeed, metabolism has been shown to be linked to the virulence of *M. tuberculosis* and plays a key role in the development and maintenance of both acute and persistent TB infections [16;20;21]. It is perhaps not surprising therefore that initial modeling efforts in tuberculosis have focused on metabolism.

The first *M. tuberculosis* constraint based model was constructed by Raman *et al* (2005) and consisted of all the reactions in mycolic acid synthesis [26]. This sub-model of metabolism was composed of 219 reactions which involved 197 metabolites, catalysed by 28 enzymes. FBA was used to simulate mycolic acid metabolism and to identify seven potential drug targets in these pathways. Although targeting a small sub-system such as mycolic acid synthesis can yield valuable information on specific pathways, it has limited value in elucidating the metabolic capability of *M. tuberculosis*. This latter objective is best approached by constructing a genome-scale networks of metabolism [2] [17]. The first published genome scale network was built by our group using *Streptomyces coelicolor* as a starting model [2]. The final functional genome scale metabolic network of *M. tuberculosis* (GSMN-TB) consisted of 739 metabolites participating in 849 reactions and involves 726 genes. The model is freely available as both an excel file or in sbml format; and is accessible via a user friendly web tool for constraint-based simulations (http://sysbio.sbs.surrey.ac.uk.tb/). FBA-based predictions of *in vitro* gene essentiality using growth rate as the objective function correlated well with predictions of gene essentiality obtained by global transposon mutagenesis [29], with an overall predictive accuracy of 78% [2].

Quantitative validation of the model was performed using growth rate data from continuous culture chemostat experiments [2]. Flux variability analysis (FVA) was
performed to assess the *in silico* metabolic response of *M. tuberculosis* to slow growth. FVA is a variant of FBA which computes the range of fluxes in each reaction that are compatible with optimization of the objective function. The FVA predicted that whereas the relative fluxes through most reactions would be similar between slow and fast growth rate, a significant increase in flux through the isocitrate lyase reaction was expected at slow growth rate. This generated the hypothesis that isocitrate lyase was involved in mycobacterial survival at slow growth rates. Isocitrate lyase is the gating enzyme into the glyoxylate shunt, an anapleurotic pathway important during catabolism of substrates, such as fatty acids, that are incorporated as 2 carbon compounds. However, the glyoxylate shunt is not thought to be involved in catabolism of a 3 carbon compound, such as glycerol, so the prediction that isocitrate lyase was raised at slow growth rate was surprising. If flux through the glyoxylate shunt was indeed raised then we would expect to see a parallel rise in isocitrate lyase enzyme activity. This was confirmed in chemostat-grown *M. bovis* BCG, supporting the hypothesis that the glyoxylate shunt is involved in maintenance of slow growth of the tubercle bacillus, at least in the chemostat. The findings may also have relevance to the demonstration that isocitrate lyase is essential for survival of *M. tuberculosis in vivo* [20;21], prompting investigation of the enzyme as a possible drug target [5;30].

A second genome scale reconstruction of *M. tuberculosis* was published Jamshidi and Palsson (2007). Silicon *iNj66I* weighs in at 939 reactions and 661 genes and was constructed using several databases combined with manual curation [17]. This model was validated by comparing *in silico* predictions of growth rate and gene essentiality with previously published experimental data. The *in silico* predictions of the growth rate of *M. tuberculosis* on three different media agreed with experimental data but the model had a poor predictive accuracy for experimental essentiality data (55%).

Building upon the work described by Ramen *et al* (2005) hard-coupled reaction (HCR) sets were calculated for the *iNj66I* model in order to predict novel drug targets on a genome scale [17]. Hard coupling reaction (HCR) sets are generated by a method which does not require an objective function and therefore overcomes any bias introduced by using a specific objective function to constrain the solution space. HCR’s are groups of
reactions which, due to mass conservation and connectivity constraints, must operate in unison. This strategy identified known and also potential new drug targets which require further analysis.

The mycolic acid synthesis sub-model and the two genome scale network reconstructions available for *M. tuberculosis* illustrate the different approaches which can be applied to reconstructing, validating and applying metabolic models. They also provide a reference for future metabolic reconstructions. The next challenge is to combine these 3 models and build upon them by integrating any new experimental data in order to expand and refine the reconstructions in an iterative cycle. In this way the model can serve as an up-to-date representation of the cumulative knowledge of *M. tuberculosis*’s metabolic capabilities.

For comparison, the *E. coli* genome scale network has undergone six different successive reconstructions over the last 18 years, each one contributing positively to a large number of different studies [15]. A well curated reconstruction is a perquisite for all systems biology approaches to studying *M. tuberculosis*. We and others are currently reconstructing the next generation of genome scale TB models with the ultimate goal of building an integrated host-pathogen metabolic network model. The currently available and future models will lead to many advances and are likely to provide a crucial component of TB research.

**Applications of the models**

The reconstruction process itself demonstrated the utility of *in silico* models of *M. tuberculosis* metabolism in computing phenotypic behaviour, predicting drug targets and in the generation of novel hypotheses, which can then be used to guide experimentation [2;17;26]. In addition, several groups have already applied these models with significant initial success to identify drug targets, model drug inhibition and also as a context for interpretation of experimental data [3;8;13;18;27;28].

Metabolic models have also been used for drug discovery [7]. Raman *et al* (2008) have developed a very useful framework for drug discovery which incorporates the three published *M. tuberculosis* metabolic models as part of a comprehensive and integrated
scheme for drug identification. TargetTB imposes sequential filters on an initial list of essential genes in order to generate several shortlists of putative drug targets [28]. The first list was generated by combining essentiality data from a protein-protein interactome and FBA studies using both the available genome scale models (iNj661, GSMN-TB) and also the mycolic acid pathway (MAP) sub-model. This list was then filtered using sequence analysis to remove genes with homologues in humans; structural analysis to remove targets with similar binding sites to human proteins and this process identified 622 putative drug targets. Further filtering generated a short list of broad spectrum antibacterial, TB specific targets and also those candidates which could be important for treating persistent TB (216 targets). Several known and proposed candidate drug targets were identified using TargetTB, validating this approach and indicating that the other targets on this list could also be promising candidates for subsequent experimental investigations.

As an alternative to the above approaches which predict reactions as drug targets by evaluating their effects on metabolism, Kim et al (2009) used a metabolite-centric approach to identify potential candidate drug targets in a number of pathogens including M. tuberculosis. Their strategy was to use metabolite essentiality and chokepoint analysis as a starting point for prioritizing drug targets [18]. This approach has the advantage of identifying multiple target mechanisms which would be very computationally expensive by a reaction centric route. Using constraint based flux analysis of the iNj661 model, essential metabolites were identified as metabolites which resulted in zero growth when removed from the model using growth rate as the objective function. Chokepoint analysis identifies enzymes that uniquely produce and/or consume a particular metabolite. Metabolites which were both essential and chokepoints were filtered in order to remove metabolites with human homologues that could produce unwanted side effects. This analysis identified 413 essential metabolites, 554 chokepoints and a final list of 364 putative drug targets in M. tuberculosis. Chorisomate was identified by this analysis as a broad range target. As a product of the shikimate pathway which is essential in higher plants, fungi, bacteria and algae and is absent in mammals chorisomate production has been validated as an attractive drug target by others [9;12]. Mycolate was also identified
as a prioritized drug target and its metabolism is already targeted by current anti-
tubercular drugs.

In another study drug target combinations were identified using an iNj661 based reaction
influence network in order to analyze protein-protein dependencies [27]. Proteins were
identified which are highly influential in the network and their interaction with other
influential proteins analysed in order to ascertain influential pairs, triplets and
quadruplets. This analysis was complemented with FBA which demonstrated that most of
the identified combinations of proteins are also synthetic lethals in silico and may
therefore be very good drug targets.

**Future challenges**

Whilst the application of systems biology approaches to studying *M. tuberculosis* is still
in the early stages, several novel approaches has delivered new insight into the
metabolism of this pathogen. This review has illustrated that the metabolic model
reconstruction process can in itself reveal useful information about the biology of *M.
tuberculosis* [2] [17]. Ultimately these metabolic models will be complemented by the
integration of regulatory and signaling circuits. However as this process is only just
beginning in *E. coli* this is a very long term aim for TB systems biologists. While fully
integrated models of tuberculosis are still a very distant goal, systems biology efforts can
(and are beginning to) have an enormous impact on both our understanding of the
metabolic potential of *M. tuberculosis* and the success of the drug discovery process
which could lead to significant advances in the treatment of tuberculosis.
REFERENCES


