Johnjoe McFadden · Dany J.V. Beste
Andrzej M. Kierzek *Editors*

Systems Biology of Tuberculosis

Springer

Johnjoe McFadden • Dany J.V. Beste
Andrzej M. Kierzek
Editors

Systems Biology of Tuberculosis

Springer

Editors
Johnjoe McFadden
Faculty of Health and Medical Sciences
University of Surrey
Guildford, Surrey, UK

Dany J.V. Beste
Faculty of Health and Medical Sciences
University of Surrey
Guildford, Surrey, UK

Andrzej M. Kierzek
Faculty of Health and Medical Sciences
University of Surrey
Guildford, Surrey, UK

ISBN 978-1-4614-4965-2 ISBN 978-1-4614-4966-9 (eBook)
DOI 10.1007/978-1-4614-4966-9
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012951628

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Introduction

Douglas Young

Reconstruction of evolutionary history by Bayesian analysis of extant genome sequences suggests that the *Mycobacterium tuberculosis* complex emerged as an infection of anatomically modern humans carrying the L3 mitochondrial haplotype around one hundred thousand years ago. Since then, *M. tuberculosis* has demonstrated a remarkable ability to persist amongst small, highly vulnerable populations of early humans migrating out of Africa and to thrive in response to changing demography and recent massive population expansion. The Global Plan to Stop TB proposes its elimination—defined as fewer than one case per million individuals—by 2050 [1]. Given our history of intimate companionship, how can we envisage a strategy to drive this microbe towards extinction?

While we can readily collect data on tuberculosis epidemiology, attempts to explain patterns of disease—and hence to design rational control strategies—reveal contributions from multiple variables. It is clear that disease results from some combination of host, microbial, and environmental factors, but it is hard to generate mechanistic models that tease these apart. This is an example of an "inverse problem", in which we have to try and infer the parts by examination of the whole. This is anathema to reductionist biology, and it is the territory to which systems biology aspires.

Sydney Brenner highlights the attempt to address ill-posed inverse questions as a lethal flaw in the systems approach, predicting a spiral into low-input, high-throughput, no-output biology [2]. Systems biologists counter by proposing an iterative process of modelling and forward testing of derived hypotheses. This involves an interesting mix of logical reasoning. Wikipedia pithily describes deduction (in which a conclusion is determined by a precondition) as the province of mathematicians, induction (in which a conclusion is a probable outcome of a

D. Young
MRC National Institute for Medical Research, Mill Hill,
London NW7 1AA, UK
e-mail: d.young@imperial.ac.uk

precondition) as the province of scientists, and abduction (in which a precondition is inferred from a conclusion) as the province of detectives. While Sherlock Holmes may dispute elements of this categorisation, the systems biology agenda strives towards some clear-sighted integration of all three methods.

A systems approach to tuberculosis can be envisaged as a series of nested problems (often inverse problems) ranging in scale from the level of populations, to individuals, to the cellular and molecular that can conveniently be branded as "systems epidemiology" [3], "systems vaccinology" [4], and so on. The present volume comprises a series of contributions from groups who are at the forefront of applying systems biology approaches to understand tuberculosis, working primarily at the cell biology end of this spectrum.

Beste and McFadden [5] and Jamshidi et al. [6] review advances in metabolic modelling. Metabolomics is particularly appropriate for network modelling. Connectivity between components in the network is direct—one metabolite changes into another—in contrast to the indirect spatial and temporal interactions used to build networks of genes and proteins. Jamshidi et al. describe fundamental processes of network modelling, while Beste and McFadden stress the importance of a robust experimental system to generate data for modelling, outlining the advantages of growth in a chemostat as a means to optimise relative homogeneity of the bacterial populations under study. Both groups stress the dependence of metabolic modelling on the availability of an accurately annotated genome. Contributions from a series of outstanding researchers established the foundations of mycobacterial biochemistry and metabolism during the first half of the twentieth century, but a relative neglect of tuberculosis research in the 1960s and 1970s left multiple gaps in our knowledge. In the genome era, there is a tendency to fill these gaps using sequence homologies with other organisms. Both papers stress the importance of caution in this, favouring an iterative process in which metabolic models are used both to formulate and test forward hypotheses as well as to track back and correct misannotations. Even core textbook metabolic pathways are found to differ between *M. tuberculosis* and "canonical" *E. coli*.

Lack of information about the physiological state of *M. tuberculosis* within infected humans presents a major challenge for modelling tuberculosis pathogenesis, and models derived from microbial culture systems have been usefully extended to metabolomics of intracellular *M. tuberculosis* in macrophages. With the exception of a limited range of microbe-specific metabolites, it is difficult to derive direct experimental data on mycobacterial metabolism in infected tissues; in fact it probably makes sense to view host and microbe as a single, integrated metabolic system. It is likely that the bacteria sample a highly diverse range of intracellular and extracellular microenvironments during infection, with availability of nutrients and oxygen varying widely over space and time [7]. A systems biology challenge will be to infer microbial physiological states on the basis of measurements of host metabolism.

Both of the metabolic modelling papers stress the importance of integrating metabolomics with transcriptional profiling and functional genomics. Waddell et al. [8] and Rao et al. [9] take up the story from a transcriptome perspective. Traditional transcriptional profiling is also contingent on accurate genome annotation to select

genes that are interrogated in microarray platforms. More recent high-density tiling arrays and RNA sequencing approaches remove this limitation and are starting to uncover a considerable repertoire of non-coding RNA outside of annotated open reading frames [10]. This includes small intergenic RNAs, antisense transcripts, and *cis*-encoded untranslated regions that are likely to regulate the stability of mRNA transcripts and the efficiency of their interaction with ribosomes. Characterisation of this layer of post-transcriptional regulation will be an important element in integrating transcriptome and proteome data, and hence for inferring physiology from transcription profile. Both papers highlight a central role for the ability of transcriptional profiling to uncover crosstalk between host and microbe in macrophage infection models. As discussed for metabolomics, there is a trend in transcriptomic modelling to move from consideration of the isolated microbe to the infected cell or lesion as the system under study. Given the specificity and technical ease of nucleic acid amplification, transcriptome data represent the most accessible source of information about *M. tuberculosis* in lesions. Both papers stress the importance of linking transcriptional profiling to functional genomics, using transposon mutagenesis or formal gene deletion in the microbe side, and exploitation of RNAi screens for host manipulation.

Steyn et al. [11] contribute a review of mycobacterial proteomics, focusing in particular on different strategies for generating protein–protein interaction networks. They describe specific contributions to exploration of mechanisms that play a central role in host–pathogen interactions—two-component signalling, protein secretion, and DNA repair—and highlight the importance of studying post-translational modification events. Together with the chapters on metabolism and transcription, this builds an encouraging picture of a strong integrated 'omics platform that informs a systems biology understanding of *M. tuberculosis* and its interaction with host cells. Given our knowledge of the mycobacterial cell, it is obviously important that this platform includes the parallel advances occurring in the area of lipidomics.

Chandra [12] takes forward the application of a systems approach to identification of potential drug targets, integrating genetic essentiality with metabolic modelling to identify key choke-points. Target-driven approaches based on genetic essentiality have been profoundly disappointing in the field of antimicrobial drug discovery [13]. In part this reflects a series of technical limitations: compound libraries used in high-throughput screens may have inadequate representation of relevant chemical space, evidence from gene deletions does not stratify high- versus low-vulnerability targets, and potent enzyme inhibitors may not penetrate bacterial cell walls. There may also be a limitation in the general concept that simple inhibition of an essential enzyme is sufficient for bactericidal activity [14]. Events downstream of the initial drug–target interaction are probably crucial to the effectiveness of successful antibiotics, with accumulation of toxic effector molecules providing the actual trigger for cell death. Systems biology models that can predict lethal consequences of target of inhibition would provide an important advance. Given the large number of moderately potent hits arising from high-throughput screens against whole mycobacteria, a systems biology approach capable of identifying cell death parameters that are more experimentally tractable than measurement of

colony forming units would also be of considerable use for prioritisation. Chandra outlines a concept of "polypharmacology", involving analysis of the interaction of a single drug with multiple targets and the effect of drug combinations. Drug combinations are central to tuberculosis treatment regimens, and there is a need for systems biology approaches to rationalise and predict positive and negative drug–drug interactions.

Ghosh et al. [15] describe an exciting international collaborative effort to exploit systems biology for TB drug discovery in the context of community engagement in a "big science" initiative. This is a joint programme between two Japanese systems biology institutes and the Indian Open Source Drug Discovery project. The emphasis is on novel communication systems, generating a virtual collaborative space that accommodates input from a wide community of researchers. This illustrates a key aspect of the systems biology agenda: stimulation of multidisciplinary interactions across a wide range of biology, engineering and mathematics. Other multidisciplinary consortia exploring the systems biology of tuberculosis have been established in the USA [16] and Europe [17]. There are formal similarities between social interaction networks and protein–protein interaction networks, and it is clear that social factors will be at least as important as molecular factors in the success of future strategies for tuberculosis control. Perhaps community-based approaches to enhance communication amongst systems biology researchers could be extended to enhance communication between scientists and the wider public?

Two papers address host–pathogen interactions from the perspective of the immune response. Pine et al. [18] present a comprehensive overview of the highly interconnected host immune network, picking out molecular and cellular biomarkers that could be used in combination to stratify the position of individuals with the tuberculosis infection spectrum. Fallahi-Sichani et al. [19] describe various techniques to model the development and heterogeneity of tuberculous granulomas, including a powerful and innovative agent-based modelling approach. Focusing on the role of TNFα concentration as an example, they illustrate a very important aspect of systems biology modelling as an aid to feature selection. This technique, commonly used in machine learning, distinguishes parameters whose variation has a more or less critical effect on overall behaviour of the system and therefore warrants higher or lower prioritisation for further experimental definition. This provides a framework for attractive synergy between modellers and experimentalists.

Rocco et al. [20] return to the issue of microbial modelling from a novel perspective, reviewing stochastic influences on gene expression and their relevance to population heterogeneity and persistence. Conventional 'omics approaches use high-throughput data generated from bulk populations that are assumed to display a homogeneous phenotype. This is certainly not the reality. Noise is an important element in bacterial physiology, and there is extensive evidence of heterogeneity in gene expression amongst bacterial cells grown in an identical strictly controlled environment. It can be anticipated that such effects are amplified multi-fold in the complex microenvironments encountered during infection. Rocco et al. highlight the potential links between population heterogeneity and the acutely practical

issue of phenotypic tolerance to drugs. Addressing these issues requires alternative experimental and computational approaches, and it is crucial that strategies are developed to integrate single cell information with "mainstream" bulk population 'omics.

References

1. http://www.stoptb.org/global/plan/
2. Brenner S (2009) Interview with Sydney Brenner by Soraya de Chadarevian. Stud Hist Philos Biol Biomed Sci 40:65–71
3. Comas I, Gagneux S (2011) A role for systems epidemiology in tuberculosis research. Trends Microbiol 19:492–500
4. Pulendran B, Li S, Nakaya HI (2010) Systems vaccinology. Immunity 33:516–529
5. Beste DJV, McFadden J (2012) Metabolism of *Mycobacterium tuberculosis*. In: McFadden J (ed) Systems biology of mycobacteria. Springer
6. Jamshidi N, Bordbar A, Palsson B (2012) Modeling *Mycobacterium tuberculosis* H37Rv *in silico*. In: McFadden J (ed) Systems biology of tuberculosis. Springer
7. Barry CE 3rd, Boshoff HI, Dartois V, Dick T, Ehrt S et al (2009) The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol 7:845–855
8. Waddell SJ, von Kamp A, Klamt S, Neyrolles O (2012) Host-pathogen interactions. In: McFadden J (ed) Systems biology of tuberculosis. Springer
9. Rao KVS, Kumar D, Mande SC (2012) Probing gene regulatory networks to decipher host-pathogen interactions. In: McFadden J (ed) Systems biology of tuberculosis. Springer
10. Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, et al (2011) Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. PLoS Pathog 7:e1002342
11. Steyn AJC, Mai D, Saini V, Farhana A (2012) Protein-protein interaction in the –omics era: Understanding *Mycobacterium tuberculosis* function. In: McFadden J (ed) V. Springer
12. Chandra N (2012) Drug discovery. In: McFadden J (ed) Systems biology of tuberculosis. Springer
13. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat Rev Drug Discov 6:29–40
14. Dick T, Young D (2011) How antibacterials really work: impact on drug discovery. Future Microbiol 6:603–604
15. Ghosh S, Matsuoka Y, Asai Y, Kitano H, Bhardwaj A, et al (2012) Software platform for metabolic network reconstruction of *Mycobacterium tuberculosis*. In: McFadden J (ed) Systems biology of tuberculosis. Springer
16. https://tbsystemsbiology.org/
17. http://www.systemtb.org/
18. Pine R, Bushkin Y, Gennaro ML (2012) Immunological biomarkers for tuberculosis: potential for a combinatorial approach. In: McFadden J (ed) Systems biology of tuberculosis. Springer
19. Fallahi-Sichani M, Marino S, Flynn JL, Linderman JJ, Kirschner DE (2012) A systems biology approach for understanding granuloma formation and function in tuberculosis. In: McFadden J (ed) Systems biology of tuberculosis. Springer
20. Rocco A, McFadden J, Kierzek AM (2012) Stochastic gene expression in bacterial pathogens: a model for persistence? In: McFadden J (ed) Systems biology of tuberculosis. Springer

Contributors

Yoshiyuki Asai Okinawa Institute of Science and Technology, Okinawa, Japan

Anshu Bhardwaj Council of Scientific and Industrial Research, Delhi, India

Dany J.V. Beste Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey, UK

Kausik Bhattacharyya CSIR-Institute of Genomics and Integrative Biology, Delhi, India

Aarash Bordbar Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA

Samir Brahmachari Council of Scientific and Industrial Research, Delhi, India
CSIR-Institute of Genomics and Integrative Biology, Delhi, India

Yuri Bushkin Public Health Research Institute, New Jersey Medical School, Newark, NJ, USA

Nagasuma Chandra Department of Biochemistry, Indian Institute of Science, Bangalore, India

Mohammad Fallahi-Sichani Department of Systems Biology, Harvard Medical School, Boston, MA, USA

A. Farhana Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL, USA

JoAnne L. Flynn Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Maria Laura Gennaro Public Health Research Institute, New Jersey Medical School, Newark, NJ, USA

Samik Ghosh The Systems Biology Institute, Tokyo, Japan

Akanksha Jain Council of Scientific and Industrial Research, Delhi, India

Neema Jamshidi Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA

Andrzej Kierzek Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey, UK

Denise E. Kirschner Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA

Hiroaki Kitano The Systems Biology Institute, Tokyo, Japan

Okinawa Institute of Science and Technology, Okinawa, Japan

Sony Computer Science Laboratories, Inc., Tokyo, Japan

Steffen Klamt Max Planck Institute for Dynamics of Complex Technical Systems Magdeburg, Germany

Dhiraj Kumar International Centre for Genetic Engineering and Biotechnology, New Delhi, India

Vikas Kumar CSIR-Institute of Genomics and Integrative Biology, Delhi, India

Jennifer J. Linderman Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA

D. Mai Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL, USA

Shekhar C. Mande National Centre for Cell Science, Pune, India

Simeone Marino Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA

Yukiko Matsuoka The Systems Biology Institute, Tokyo, Japan

ERATO Kawaoka Infection-Induced Host Response Project, Japan Science and Technology Agency, Tokyo, Japan

Johnjoe McFadden Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey, UK

Anupam Kumar Mondal CSIR-Institute of Genomics and Integrative Biology, Delhi, India

Olivier Neyrolles Centre National de la Recherche Scientifique, Institut de Pharmacologie et de Biologie Structurale, Toulouse, France

Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, Université Paul Sabatier, Toulouse, France

Bernhard Palsson Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA

Anurag Passi Council of Scientific and Industrial Research, Delhi, India

Richard Pine Public Health Research Institute, New Jersey Medical School, Newark, NJ, USA

Priyanka Priyadarshini Council of Scientific and Industrial Research, Delhi, India

Kanury V.S. Rao International Centre for Genetic Engineering and Biotechnology, New Delhi, India

Andrea Rocco Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey, UK

V. Saini Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL, USA

Vinod Scaria CSIR-Institute of Genomics and Integrative Biology, Delhi, India

Anup Shah CSIR-Institute of Genomics and Integrative Biology, Delhi, India

Pratibha Sharma CSIR-Institute of Genomics and Integrative Biology, Delhi, India

Kumari Sonal CSIR-Institute of Genomics and Integrative Biology, Delhi, India

Adrie J.C. Steyn Department of Microbiology and Centers for AIDS Research, University of Alabama at Birmingham, Birmingham, AL, USA

KwaZulu-Natal Research Institute for Tuberculosis and HIV, Congella, Durban, South Africa

Rohit Vashisht Council of Scientific and Industrial Research, Delhi, India

Priti Vishnoi CSIR-Institute of Genomics and Integrative Biology, Delhi, India

Axel von Kamp Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

Simon J. Waddell Brighton and Sussex Medical School, University of Sussex, Brighton, UK

Douglas Young MRC National Institute for Medical Research, Mill Hill, London, UK

Contents

Chapter 1
Modeling *Mycobacterium tuberculosis* H37Rv In Silico

Neema Jamshidi, Aarash Bordbar, and Bernhard Palsson

Abstract Network reconstructions and constraint-based modeling have been shown to be effective methods for understanding complex processes, such as metabolism. These reconstructions are in fact biologically structured knowledge-bases that can be queried through computations, and thus have become valuable tools for Systems Biology. Strengths of this approach include flexibility in incorporating "incomplete" data measurements, the ability to incorporate different types of data (high-throughput as well as physiological), simultaneously, as well as the ability to make predictions with minimal reliance on parameter and curve fitting. Thus, this approach aims to move away from fitting data to describe experimental results using the current understanding of metabolism in order to interpret the data, make predictions, and to identify the gaps and bridges in knowledge.

The critical components for creating genome-scale reconstructions of metabolism include a sequenced and annotated genome, reaction stoichiometry for the annotated enzymes, and a bibliome for the organism (combined primary and secondary literature sources). Network reconstructions of the devastating pathogen *Mycobacterium tuberculosis* have been developed and have enabled the ability to query functional capabilities using constraint-based modeling approaches. Since these networks are then structured in terms of "gene–protein–reaction" associations, these knowledge-bases can serve as biologically structured databases onto which various high-throughput data types can be directly mapped on.

This chapter will focus on the model reconstruction process, methods that have been employed for analysis, and predictive applications of modeling the pathogen H37Rv strain of tuberculosis. Employing the existing analysis methods and available datasets there have already been a large number of applications for modeling constraint-based modeling of H37Rv. The reconstruction process is a time and resource intensive procedure and to date high quality reconstructions have not been

N. Jamshidi (✉) • A. Bordbar • B. Palsson
Department of Bioengineering, University of California, San Diego,
9500 Gilman Drive, La Jolla, CA 92093-0412, USA
e-mail: neema@ucsd.edu

J. McFadden et al. (eds.), *Systems Biology of Tuberculosis*,
DOI 10.1007/978-1-4614-4966-9_1,

possible without manual curation. The benefit of having a detailed and quality controlled reconstruction procedure is to help determine a high quality model that will provide more meaningful predictions from simulations. Applications of *M. tuberculosis* models have included the prediction of growth rates, assessment of different growth media, prediction of gene knockouts, identification of new drug targets, identification of alternative drug targets for existing drugs, and modeling the interaction macrophages during different infectious states.Historically, technological advancements have driven biological discovery and have thus been a limiting factor in the development of methods to modify and alter biology, e.g., antibiotics. However, in the past decade with various high-throughput technologies (e.g., transcriptomics, proteomics, metabolomics, etc.) are being employed more frequently, thus there is a growing burden and need for means to integrate, interpret, and ideally make predictions for these datasets. Given the successes to date, with further development of new methods in conjunction with deeper experimental probing of tuberculosis in vitro and in vivo, constraint-based modeling will likely become even more important in the finding new targets and treatments for tuberculosis.

1 Introduction

The principle behind constraint-based modeling is to use physico-chemical or biological constraints to provide insights that are biologically insightful. The rich, detailed history of biochemistry during the past 60 years has resulted in the ability to describe metabolic networks in terms of elementally balanced biochemical reactions. Furthermore, thermodynamic and kinetic characterization of many of the enzymes in an effort to characterize the physical properties of these biological catalysts has enabled additional levels of characterization.

Flux Balance Analysis (FBA) has become the bread and butter of constraint-based modeling [1, 2]. Employing this approach involves the quasi-steady state assumption and knowledge of the identity and composition of the interacting components, as well as the set of the biochemical reactions that occur in the system of interest. If quantitative thermodynamic data is not available, qualitative thermodynamic data can also be incorporated, simply by specification of reversibility of a reaction. While this is a very simple encapsulation of the complex non-equilibrium thermodynamics within a cell, it can have significant implications on constraining a network and reducing (or expanding) the number of possible steady state solutions. At the most elementary level, FBA can be applied to a single biosynthetic pathway [3]. While this might be interesting in some cases, the benefits of this approach are really appreciated when one makes the jump to the organelle-, cell-, and genome-scale models [4, 5].

The data (in-)completeness problem is likely to always be present at all spatial and temporal hierarchies in biology. High-throughput technologies have been progressing to help close the data incompleteness gap. Genomics was the first "omics" field in biology and has been followed by numerous other high-throughput measurement technologies, including proteomics, and metabolomics and a seemingly innumerable

array of other "omics" sub-fields. We focus the discussion on the technologically driven by high-throughput measurements.

Aside from technical challenges associated with the analysis of voluminous datasets, there is a more pressing challenge regarding the context and the manner in which the data are analyzed. There is a demonstrated need to move away from black-box modeling approaches to understand these data, towards mechanistic or partially mechanistic (gray box) models. The development of multiple "omics" data fields has further compounded this problem, further highlighting the need for analysis of large datasets that often include orthogonal types of data to be analyzed in a biologically relevant and biologically meaningful context.

Constraint-based modeling has provided one approach to organizing and analyzing this data from a biological viewpoint, while paying heed to physico-chemical constraints. The application of these methods to the deadly pathogen *Mycobacterium tuberculosis* [6] has resulted in advancements in the understanding of its metabolic capabilities and opened potential avenues for new or alternative treatments.

2 The Reconstruction Process

The quality and scope of metabolic network reconstructions have continued to evolve during the last 15 years, with current descriptions involving a detailed, iterative, quality-controlled process [7]. Progress has been made in the automation of reconstruction [8]. Nevertheless, a key hallmark of quality reconstructions however has been the need for manual curation, on some level. General steps employed in the process of network reconstruction are outlined in Fig. 1.1.

The starting point for genome-scale reconstructions is a sequenced and annotated genome. This serves as the scaffold onto which the biochemical transformations carried out by enzymes in the organism are mapped. Manual curation follows, which includes gathering evidence and critically evaluating the primary (and review) literature for information about the genes, proteins, and metabolites. There have been an increasing number of organism specific databases that have been developed during the past decade that were very helpful for fleshing out the general network architecture of *M. tuberculosis* [9–11].

Following manual curation, there is conversion of the set of biochemical reactions into a stoichiometric matrix. The stoichiometric matrix is unique compared to many other types of matrices in biological systems, as it has integer entries, thus there is no noise associated with the values [2, 12]. The conversion to a mathematical format also involves the application of (qualitative) thermodynamic constraints. Quantitative constraints have been explored [13, 14] and have been used for expanding the scope to dynamic models [15–17]. However, for the purpose of developing a basic model with which to carry out constraint-based modeling, only directionality needs to be specified (i.e., reversible or irreversible). The debugging and functionality testing stage is another step that is a critical step in the process, as it ensures network functionality. It is unfortunately also a time-consuming process.

For micro-organisms the primary functionality test involves biomass production. The biomass "reaction" is actually a pseudo-reaction that is included to account for the growth as well as non-growth associated demands [18]. Since this reaction requires the production and utilization of a large number of metabolites, the ability to produce biomass implicitly accounts for the biosynthesis of a large number of compounds. Construction of iNJ661 involved testing individual components (i.e., non-essential amino acids, mycolic acids, etc.) prior to testing the complete biomass function. Since the biomass function involves so many different compounds, in practice, many of the compounds (such as non-essential amino acids, vitamin products, etc.) are tested individually (see "Reaction and Pathway Function Testing" in Fig. 1.1). There are other considerations and issues to evaluate during the quality control process, such as revising constraints in order to eliminate "free energy" producing loops [19]. This issue and others are discussed in more detail by Thiele and Palsson in [7].

It should be noted that the iterative loop in Fig. 1.1 involves further manual curation and more detailed investigation into a particular functionality, in order to understand why the test failed. For iNJ661 this included multiple rounds of revisiting and reevaluating the primary literature as well as detailed evaluation of the relevant pathways. In some situations there may be no direct evidence to support incorporation of a particular reaction (such as a transport reaction), which is an intermediate in a pathway whose endpoint is known to occur in the organism. In order to produce a functional model, the transport reaction may need to be added. This is one example of why "confidence level" scores are an important quality control measure in network reconstructions, because they denote the type of evidence that was used to justify incorporation of the reaction. These can then be used to determine future experiments and to also re-evaluate model content when additional datasets are generated for the organism.

3 Network Characterization

There are a myriad of ways to test or assess functionality of a model [2, 20–22]. Once a functional model has been constructed, one of the first steps of analysis is to understand how a particular objective, such as growth, varies on varying substrate utilization. Phase-plane diagrams can address such questions by plotting two different network fluxes (uptake or exchange reactions) while optimizing for an objective. This can be used to assess the trade-off associated with the use of one substrate versus another. The fatty acid constitution of Mycobacteria and other acid fast organisms is complex and unique compared to most other prokaryotes. These fatty acids also constitute a significant portion of the biomass. Glycerol is a required substrate for this and the trade-off between glucose and glycerol is demonstrated in Fig. 1.2, while optimizing for biomass.

The ability to carry out genome-wide screening of gene essentiality in microbes, enables testing of in silico predictions using network models. Results can be categorized