University of Manchester

Doctoral Thesis

Metabolic modelling of *Streptomyces* species for secondary metabolites production

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Science and Engineering

2019

Adam Amara

School of Chemistry, Faculty of Science and Engineering

Blank page

List of contents

List of figures

List of supplementary figures

List of tables

List of supplementary tables

List of equations

List of electronic supplementary files

Electronic Supplementary 2.1

- Additional file 1: iAA1259 metabolic model in SBML format. (XML)
- Additional file 2: Excel file specifying metabolites, reactions, genes contained, and databases IDs present in the iAA1259 metabolic model. (XLS)
- Additional file 3: Biomass modifications and calculation of ATP consumption. (XLS)
- Additional file 4: Full detailed data used for the Fig. 4. (XLS)
- Additional file 5: High-resolution version of Figure 2.4. Validation by integrated transcriptomics and proteomics analysis. (PNG)

Electronic Supplementary 3.1

- Additional File 1: Modifications done on the iAA1259 model to build the iSLT1240 model, proteome comparisons with the low identity genes analysis, isoenzymes deleted, model modifications, and biomass modifications. (XLSX)
- Additional File 2: Genome scale metabolic network in SBML format. (XML)
- Additional File 3: Set of constraints for the biomass and the R2YE media.
- Additional File 4: This file contains the sorted and raw flux predictions. (XLSX)
- Additional File 5: Detailed processed data used to constrain the models under the conditions used for the proteomics experiments. (XLSX)

Electronic Supplementary 4.1

Reconstruction pipeline notebook: R and Python script with output for the reconstruction pipeline. (HTML)

Reconstruction pipeline notebook: R and Python script with output for the reconstruction pipeline. (R script)

Reconstruction pipeline notebook: R and Python script with output for the reconstruction pipeline. (PDF)

Electronic Supplementary 4.2

Interactive_plot_1: Interactive bubble plot with the number of reactions added versus the number of reactions deleted in the reconstructed metabolic models. (HTML)

Interactive_plot_2: Interactive bubble plot with the sum of reactions deleted and added versus ratio of reactions change to model size of the metabolic models. (HTML)

Interactive_plot_3: Interactive bubble plot for the ratio of blocked reactions versus the distance of the model to the reference strain *S. coelicolor*. (HTML)

Interactive_plot_4: Interactive bubble plot for the number of essential reactions versus the distance of the model to the reference strain *S. coelicolor*. (HTML)

Electronic Supplementary 4.3

This folder contains the 51 models reconstructed by the comparative metabolic model reconstruction pipeline in SBML format. (XML)

Electronic Supplementary 5.1

Notebook ACT 1: Notebook *R* script with plot outputs for the acquired actinorhodin data. (HTML)

Notebook ACT 2: Notebook *R* script with plot outputs for the chemostat actinorhodin published by Nieselt et al. (HTML)

Notebook RED 1: Notebook *R* script with plot outputs for the acquired undecylprodigisin data. (HTML)

Notebook RED 2: Notebook *R* script with plot outputs for the chemostat undecylprodigiosin published by Nieselt et al. (HTML)

Notebook Glucose: Notebook *R* script with plot outputs for the chemostat glucose published by Nieselt et al. (HTML)

Notebook Glutamate: Notebook *R* script with plot outputs for the chemostat glutamate published by Nieselt et al. (HTML)

Notebook Pi: Notebook *R* script with plot outputs for the chemostat phosphate published by Nieselt et al. (HTML)

Electronic Supplementary 6.1

M145_R-tool_output: The output data and figures from the flux R tool for *S. coelicolor* M145. (Folder)

M145-ACT_R-tool_output: The output data and figures from the flux R tool for *S. coelicolor* M145+ACT. (Folder)

M1146_R-tool_output: The output data and figures from the flux R tool for *S. coelicolor* M1146. (Folder)

M1146-ACT_R-tool_output: The output data and figures from the flux R tool for *S. coelicolor* M1146+ACT. (Folder)

Additional_File_1: Predicted metabolic fluxes in the four different strains and correlation of fluxes to actinorhodin production. (XLSX)

Additional_File_2: Trade-off curves between all metabolites secretions and ACT production or growth rate for all four strains. (DOCX)

Additional_File_3: Metabolic reactions highly correlated to actinorhodin production in *S. coelicolor* metabolic model. (XLSX)

Additional_File_4: High resolution of the metabolic map of the central metabolism of *S. coelicolor* displaying the reactions highly correlated with actinorhodin. (SVG)

List of abbreviations

6PGC: 6-Phospho-D-gluconate
6PGL: 6-phospho-D-glucono-1,5-lactone
ACC: acetyl-CoA carboxylase
ACP: Acyl-Carrier-Protein
ACT: Actinorhodin
ADP: Adenosine DiPhosphate
AMP: Adenosine MonoPhosphate
ATP: Adenosine TriPhosphate
BCFA: Branched-Chain Fatty Acid
BGC: Biosynthetic Gene Clusters
CDA: Calcium-Dependant Antibiotic
CITL: citrate lyase
CoA: Coenzyme A
COBRA: COnstraint-Based Reconstruction and Analysis
CPK: yellow Coelicolor Polyketide or Coelimycin P1
DHAP: Dihydroxyacetone Phosphate
DNA: Deoxyribonucleic acid
E4P: D-Erythrose 4-phosphate.
F6P: Fructose-6-phosphate
FAD: Flavin Adenine Dinucleotide
FBA: Flux Balance Analysis
FBP: fructose-1,6-biphosphate
FMN: Flavin MonoNucleotide
G6P: glucose-6phosphate
GA3P: glyceraldehyde-3-phosphate
GDP: Guanosine DiPhosphate
gDW: gram of Dry Weight
GSMM: Genome-scale metabolic model
MALS: malate synthase
MDH: Malate DeHydrogenase
ME1/2: malic enzymes
NADH: Nicotinamide Adenine DiNucleotide
NADPH: Nicotinamide Adenine DiNucleotide Phosphate
NRPS: Nonribosomal peptide
PDH: Pyruvate DeHydrogenase
PEP: Phosphoenolpyruvate
PEPC: Phosphoenolpyruvate carboxylase
PEPCK: Phosphoenolpyruvate carboxykinase
pFBA: parsimonious Flux Balance Analysis
PFK: phosphofructokinase
PFL: Pyruvate Formate Lyase
Pi: Phosphate inorganic
PKS: PolyKetide Synthase
PPP: Pentose Phosphate Pathway
PPP: Pentose-Phosphate Pathway
PRPP: phosphoribosyl pyrophosphate
PRPPS: PRPP synthase
R5P: Ribose 5-phosphate
RED: Undecylprodigiosin
RiPP: Ribosomally synthesized and post-translationally modified peptides
Ru5P: D-Ribulose 5-phosphate
S7P: sedoheptulose-7-phosphate
SCFA: Straight-Chain Fatty Acid
SCO: *Streptomyces coelicolor*
SLIV: *Streptomyces lividans*
TAG: TriAcylGlycerol
TALA: transaldolase.
TCA Cycle: TriCarboxylic Acid Cycle
UDP: Uridine DiPhosphate
WT: Wild-Type
Xu5P: D-Xylulose 5-phosphate

Abstract

Streptomyces species produce a vast diversity of secondary metabolites of clinical and biotechnological importance, including antibacterial that are increasingly sought after to tackle the spread of resistance. Recent developments in metabolic engineering, synthetic and systems biology have opened new opportunities to exploit *Streptomyces* secondary metabolism, but achieving industrial levels of production without time-consuming optimisation has remained challenging.

In this thesis, we present the reconstruction and analysis of constraint-based genome-scale metabolic models to study and engineer primary and secondary metabolism of *Streptomyces* species. The aim of this work is to better understand so that it can aid in the increase of antibiotic production in *Streptomyces* spp. This would ultimately help in the discovery and production of new antibiotics to face the rise of antimicrobial resistance.

This thesis starts with an introduction on *Streptomyces* (and other Actinobacteria) primary and secondary metabolism, and on synthetic and systems biology principles and methods relevant to *Streptomyces* strains engineering (*Chapter I*). The work presented here involved the update and validation of a high-quality genome-scale metabolic model of *Streptomyces coelicolor* to study its primary and secondary metabolism (*Chapter II*). A metabolic model of *Streptomyces lividans* was reconstructed and compared to *S. coelicolor* model to identify the metabolic differences potentially responsible for the differences in antibiotics production between and the two closely related strains (*Chapter III*). The comparative reconstruction and comparison method used for *S. lividans* was automated to compare about 50 different Actinobacteria strains and study the metabolic differences between these organisms (*Chapter IV*). We automated the essential process used to constrain metabolic exchanges in the metabolic model that allows condition-specific predictions, by building an *R* tool (*Chapter V*). The models and tools developed and validated in this thesis were used to help design better *S. coelicolor* antibiotic producing strains, by applying it to the production of actinorhodin and the heterologous production of the GE2270A antibiotic (*Chapter VI*). Finally, the thesis concludes with a discussion on the future applications of the models and tools developed here, as well as future developments and applications of metabolic modelling for synthetic and systems biology of antibiotic production in *Streptomyces* species.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property University IP Policy (see http://documents.manchester.ac.uk/display.aspx?DocID=24420, in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.

Acknowledgements

First and foremost, I would like to thank my two supervisors, Professor Eriko Takano and Professor Rainer Breitling, who offered me this exceptional opportunity to carry my PhD project. I would like to thank both for their support, their open-door, the many discussions, their encouragements, and for teaching me to become a scientist.

I would like to thank the School of Chemistry of The University of Manchester for providing me support and funding to carry my PhD project in this institution with an incredible heritage and at the forefront of science and technology.

During the last 4 years I met many people with whom I had a chance to carry my research and exchange, and I would like to thank all them. I would like to thank people from SYNBIOCHEM for helping me during my first years, such as Dr. Nicholas Rattray, Dr. Maria Vinaixa, and Dr. Pablo Carbonell. I would like to thank people from the Breitling group, Dr. Aliah Hawari, Dr. Areti Tsigkinopoulou; special thanks to Dr. Francesco Del Carratore for the many discussions and advices on programming. In the Takano group, I had the chance to work every day with amazing people; thank you to all the members I had the chance to meet and work with, such as Dr. Ashley Chessher, Dr. William Finnigan, Dr. Suhui Ye Huang, Dr. Meli Tsai, Dr. Kamila Schmidt, Dr. Jane Gilsenan, Dr. Harshwardhan Poddar, Dr. Oksana Bilyk, and Dr. Richard Balint. Special thanks go to Dr. Gajendar Komati-Reddy, Dr. Matthew Cummings, Katherine Baker, and Nico Prandi. Thank you to Dr. Marc Biarnes-Carrera for the conversations and the fun we had (and countless memes).

During my time in Manchester, I had the chance to meet so many amazing people, but I would like to give a special thanks to my best mate Florian.

J'aimerais bien évidemment remercier mes parents pour leur soutient durant ma thèse, mes études, et ma vie de manière général. J'aimerais les remercier pour ces précieux conseils qu'ils m'ont toujours prodigué, de l'éthique de travail qu'ils m'ont transmis, et de m'avoir rappelé que lorsque une porte se ferme une autre s'ouvre. J'espère vous rendre fière en faisant ce que j'aime. Merci papa! Merci maman!

Ik wil heel graag mijn grote liefde Myrthe bedanken! Bedankt voor je geduld en dat je in mijn leven bent. Ik hoop dat we samen geweldige avonturen mogen beleven in de toekomst.

Blank page

Chapter I

1. Introduction

A major challenge in engineering biological systems is to understand and control the information from genotypes to functional phenotypes of interest [1–3]. The capability to understand and modify biological systems has accelerated these last two decades, due to the development of modern high-throughput measurements technologies such as genome sequencing [1,4]. This pushed biology as a data-rich field, in a similar way, to the shift of astronomy toward a data-rich field 30 years ago with the development of technologies such as the Hubble space telescope [5]. This explosion of complex and large datasets has driven the need for analytical tools, and modelling tools to represent complex biological systems and help interpret the data. The objective of modifying biological systems to have desired functions (e.g., bioproduction of industrially relevant compounds) has required interdisciplinary approaches [6,7]. To achieve this objective, researchers and engineers use synthetic biology to engineer biological systems and systems biology to design organisms, then analyse and interpret the data from the engineered biological systems. These methods are now applied to discover and produce antibiotics and other secondary metabolites of clinical or industrial interest [8,9]. This thesis presents the application of constraint-based genome-scale metabolic modelling with the integration of omics data, and the creation of new metabolic models and tools to study and improve the production of antibiotics in *Streptomyces* species and other Actinobacteria.

1.1 Actinobacteria and *Streptomyces* species metabolism

The emergence of pathogen's resistant to last-resort antibiotics, and the rapidly decreasing rate of antibiotics discovery, it is essential to develop new methods to discover and produce novel antibiotics [10,11]. The most prolific producer of antibiotics are the soil bacteria of the genus *Streptomyces* from the Actinobacteria phylum, which are the source of around 70% of all antibiotics on the market [12]. These organisms are still a promising source of novel antibiotics [12,13], as well as a bioproduction hosts [14–16]. Actinobacteria are Gram positive bacteria with a high genomic GC content, the organisms from this phylum are able to survive and thrive in various environmental conditions such as in soil and aquatic environments (e.g., *Streptomyces*, *Salinispora*, and *Micromonospora* species), or in plant and animals (e.g., *Corynebacterium*, *Mycobacterium*, *Nocardia*, *Frankia* species) [17]. The adaptation to different ecological niches makes the Actinobacteria a very diverse phylum phylogenetically [18], genetically [19,20], and physiologically (e.g., metabolism, or morphology) [17,21,22]. These adaptations to various ecological niches lead to the selection of a diverse range of secondary metabolites to survive (e.g., chelators) and compete against various microorganisms (e.g., antibiotics) [17,23].

1.1.1 Primary and secondary metabolism

Because of the industrial and clinical importance of secondary metabolites produced by *Streptomyces* species and in Actinobacteria, their primary and secondary metabolism has been widely studied [24–27]. However, little is known about the full intricacies of metabolic regulation, response to environmental cues, secondary metabolites roles, or metabolic differences responsible for high or low production

levels [28–31]. Primary metabolism corresponds to the anabolic and catabolic metabolism used for biomass production; these are reactions to use energy and synthetize the biochemical building blocks such as lipids, nucleic acids, proteins, carbohydrates, and storage molecules (e.g., glycogen, or polyphosphate) [24,32]. While the secondary metabolism produces non-essential metabolites for the life of the organism [24,32]. However, these molecules can give a survival advantage; for example, by giving a competitive advantage in the interaction with other organisms (e.g antibiotics) or by protecting from environmental and internal stress (e.g. osmoprotectants, and pigments) [24,31]. The bioactivity of these molecules gives them a therapeutic interest; many pharmaceutical compounds are derived from secondary metabolites, of course antibiotics [33,34], but also anti-cancerous [35], antifungal [36], and other top-seller drugs such as lovastatin [37]. There are still many secondary metabolites, of which may have a potential to be a new antibiotics. Just in *Streptomyces* species, the conservative estimates for the number of natural products produced are of about 150,000, of which only 3% have been characterized [12]. Many secondary metabolites remain uncharacterized, due to multiple limitations such as compounds produced below the detection limit, or compound not produced in laboratory cultivation conditions or compound found in unculturable strains [38,39]. However, the recent developments of bioinformatics tools such as antiSMASH [9,40] have enabled the prediction of diverse secondary metabolites encoded in microorganisms genome sequences.

In general, the genes associated with the biosynthesis, transport, transcription, or resistance of secondary metabolites are found clustered in the genome (*Figure 1.1*); this corresponds to a biosynthetic gene cluster (BGC) [41,42]. The BGCs are tightly

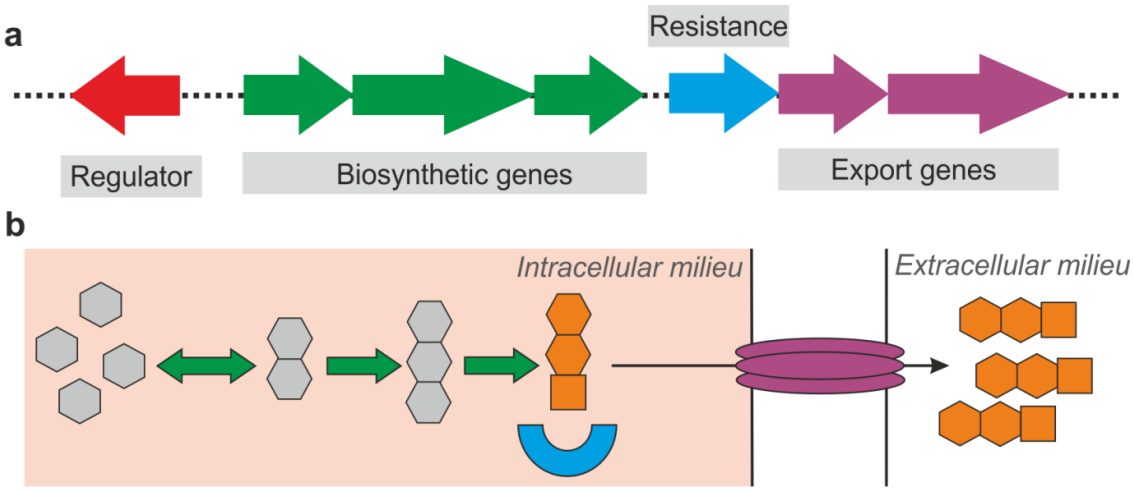


Figure 1.1: Schematic representation of a typical secondary metabolite biosynthetic gene cluster with the corresponding metabolic functions

a) A simplistic representation of a BGC producing a secondary metabolite. The regulator gene is in red, the biosynthetic genes are in green, the resistance gene is in blue, and the export genes are in purple.

b) A simplistic representation of the metabolic pathway and mechanisms encoded by the BGC. The biosynthetic genes (in green) in the metabolic pathway encode enzymes converting primary metabolites-hexagon- (e.g., malonyl-CoA) into intermediates, progressively catalysing reactions toward the final product (in orange). The final compound is toxic for the organism, so the BGC encodes resistant proteins (in blue) to protect the cell from the compound toxicity. In general, the compounds are exported by efflux pumps (in purple) to avoid toxicity or to ensure the metabolite function in the extracellular milieu (e.g., chelation of metals, or antimicrobial activity).

regulated, and the co-localisation of these genes makes regulation easier for the organism [24,41]. While the genes associated with primary metabolism are more frequently spread around the genome and duplicated [25,43]. Many primary metabolic reactions have multiple isoenzymes; these can be expressed differentially depending on life cycle or on environmental cues. For example, the depletion of phosphate triggers the entry into stationary growth phase for *Streptomyces coelicolor*, which is coordinated with a metabolic switch leading to multiple primary metabolic pathways expression to decrease [44–47]. During this metabolic switch multiple isoenzymes are differentially expressed, such as the genes associated to the NADH dehydrogenase, the *ndh* genes expression decreased while the *nuo* genes expression increased [47,48]. The

duplication of the primary metabolic genes makes the metabolism more robust and help the organism maintain essential metabolic function despite metabolic perturbations (e.g., nutrient limitation) [25].

The role and origins of secondary metabolisms are still unclear. But this metabolism certainly has a major biological importance, as strains dedicate a significant part of their genome to this metabolism. As it requires more energy to duplicate and maintain this DNA, this function would have been lost if it did not have a significant evolutionary advantage [24]. For example, the calcium-dependent antibiotic (CDA) BGC alone represents 1.1% of the genome of *Streptomyces coelicolor* [24]. There are many theories on the origin of secondary metabolisms; such as the final product selection theory, which suggests that secondary metabolites are selected for their biological function. This theory considers that selection is focused on the final metabolite because the BGCs are clustered, tightly regulated and coordinated, with a final compound well adapted to the organism [24,49]. However, how could large secondary metabolic pathways evolve step by step? The pathway intermediates rarely bring a biological advantage; furthermore, secondary metabolites are produced from primary metabolites that could be used for growth or survival. So another theory suggested is that the production of these compounds was due to overflow metabolism, with the excess of the uncontrolled primary metabolic intermediates leading to the production of secondary metabolites [24,26]. However, the secondary metabolites cannot be considered as just by-products of the overflow metabolism because their BGCs are tightly regulated and controlled. So a hypothesis brought these two theories together suggesting that when growth is limited by a nutrient (e.g., nitrogen or phosphate), the secondary metabolism was activated to maintain primary

metabolic pathways active until the limiting nutrient was available [24,26]. Then, the overflow metabolism coming from primary metabolism could be diverted toward secondary metabolic pathways. The final compounds ended up selected because of their evolutionary advantage [24]. These theories and their implications are well described in a review on primary and secondary metabolism of Streptomycetes by Hodgson [24].

The evolution of secondary metabolism helps to understand the direct link of these pathways to the primary metabolism, from the metabolic precursors to their regulations. Central metabolism has a critical role in secondary metabolism control and precursors production. For example, many secondary metabolites, such as polyketide synthases metabolites (see *Section 1.1.2*), use acetyl-CoA and malonyl-CoA as building blocks, so the pathways producing and consuming these metabolites have a major impact on secondary metabolism[24,50,51]. Glycolysis is important to generate these precursors while the fatty-acid biosynthesis is a major consumer of these precursors [52,53]. In parallel, the Embden-Meyerhof-Parnas section of glycolysis and oxidative phosphorylation is important to generate reductive and oxidative power needed for some enzymes catalysis in the secondary metabolic pathways [50,54]. The direct dependence of secondary metabolism on the primary metabolism makes primary metabolism an important target to study and engineer to understand and increase secondary metabolites production [50,51,54–57].

1.1.2 Different types of Actinobacteria secondary metabolites

Actinobacteria have diverse secondary metabolites, the main ones with major industrial or therapeutics interest are (I) the polyketides and lipids compounds [58], (II)

the terpenes [59], (III) the non-ribosomal peptides (and amino-acids compounds) [60], (IV) the ribosomally synthesized and post-translationally modified peptides (RiPP) [61]. Their biosynthesis relies on particular enzymes, respectively, (I) polyketides synthases (PKSs) [58], (II) terpene synthases (TS) [59], (III) non-ribosomal peptides synthases (NRPS) [60].

PKSs produce compounds that include many antibiotics compounds (e.g., tetracycline and erythromycin), immunosuppressants (e.g., rapamycin), or anticancer (e.g., doxorubicin) [62,63]. The PKSs use acyl-CoA building blocks to build complex molecules. PKSs are classified into three types, types I, II, and III. The type I are large multi-domain biosynthetic proteins organised in modules (*Figure 1.2a*), the type II are multi-enzyme complexes carrying a set of iteratively acting enzymes (*Figure 1.2b*), the type III PKSs are homo-dimeric enzymes that iteratively condense the building blocks to build the molecule (*Figure 1.2c*). All the PKSs use acyl-CoA precursors as building blocks (e.g., acetyl-CoA and malonyl-CoA), but the type I and II PKSs requires ACP to use the acyl-CoA precursors (*Figure 1.2a* and *b*) [58]. The type I PKSs are widely produced in Actinobacteria strains, such as the erythromycin compound (*Figure 1.2a*) first isolated in *Saccharopolyspora erythraea* [33], this compound is used to treat skin and respiratory tract infections. Type II PKSs are also frequently identified in *Streptomyces* and Actinobacteria, such as actinorhodin (*Figure 1.2b*) which is produced by *Streptomyces coelicolor* (and other species) [64,65]; this blue-pigmented compound has an antibiotic activity [66] but is mostly used to study antibiotics production as it is a coloured compound and it is easy to assay [50,67,68]. Type III PKSs are also present in many Actinobacteria, such as tetrahydroxynaphthalene (THN) and flavolin (*Figure 1.2c*) in *Streptomyces antibioticus* [69].

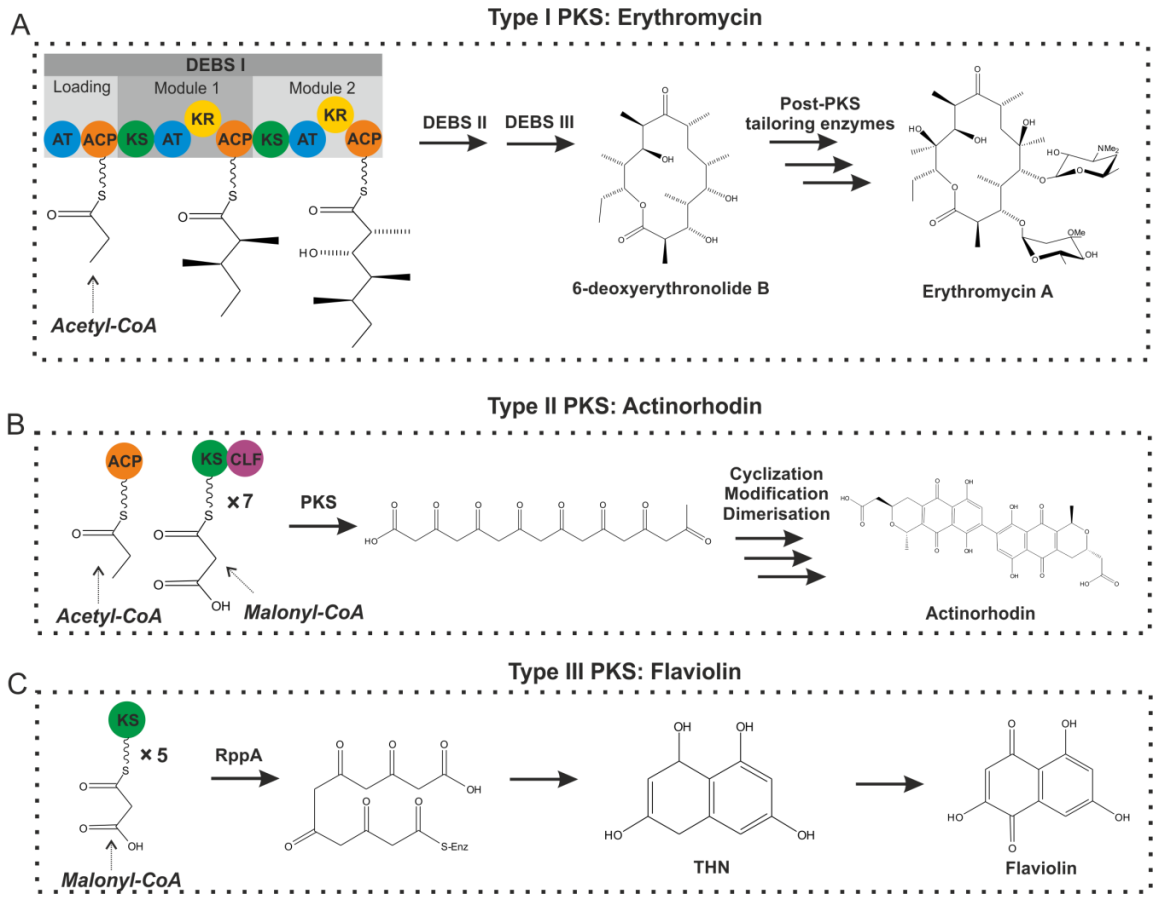


Figure 1.2: Representation of the three different types of PK biosynthesis pathways

A) Example of a type I PK biosynthetic pathway of erythromycin. The main building blocks of the type I PKS are acyl-CoA metabolites (here acetyl-CoA), this metabolite is loaded on an ACP and modified by the different enzymes in the modules of the PKS. This constitute the minimal PKS leading to 6-deoxyerythonolide B (DEBS), which can be modified by tailoring enzymes to lead to the final compound erythromycin A. Domains shown here are composed of acetyltransferases (AT), acyl carrier proteins (ACP), keto synthases (KS) and ketoreductases (KR). The example here is a modular type I PKS, but iterative versions exists using one module iteratively to produce the secondary metabolite, mostly found in fungi [37,58].

B) Example of a type II PK biosynthetic pathway of actinorhodin. The main building blocks of the type II PKS are also acyl-CoA metabolites (here acetyl-CoA and malonyl-CoA), these metabolites are loaded on the minimal PKS and elongated to form a long chain fatty acid which is then modified by other enzymes (e.g., cyclization, modification, dimerization) to lead to the final compound [58,65].

C) Example of a type III PK biosynthesis pathway of flaviolin. The main building blocks are also acyl-CoA metabolites (here malonyl-CoA), this process is ACP-independent and iteratively uses the enzyme to produce the precursor. The condensation of the malonyl-CoA is catalysed by RppA. The precursor is transformed in tetrahydroxynaphthalene (THN) by dehydration and decarboxylation. The THN is then converted to flaviolin by a monooxygenase (MomA) [58,69].

Terpenes compounds include many molecules used in aroma and flavours (e.g., limonene, and menthol) as well as food additives and complements (e.g., carotenoids, and retinol) [70]. Terpenes are widespread across the Actinobacteria, for example, in *S. coelicolor* alone, there are four terpenes BGCs, corresponding to isorenieratene, albaflavenone, geosmin, and hopene [40]. The terpenes biosynthesis is divided in two, the isoprene phase and the geranyl pyrophosphate phase and further [59]. In the isoprene phase, the precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are produced either from the mevalonate pathway (MEV) or from the non-mevalonate pathway using 2-C-methyl-D-erythritol 4-phosphate (MEP) (*Figure 1.3a*). Then, the IPP and DMAPP are converted into geranyl pyrophosphate (GPP) entering the second phase of terpenes biosynthesis, where GPP can be converted into a wide range of terpenes compounds, such as monoterpenes and monoterpenoids (e.g., menthol, and linalool), or sesquiterpenes (e.g., geosmin), and diterpene (e.g., retinol) [59].

NRPS compounds have diverse functions such as antibiotics (e.g., vancomycin, and calcium dependent antibiotics), to immunosuppressant (cyclosporine A), or biosurfactants (e.g., surfactin) [71]. NRPS compounds are frequently found in Actinobacteria, for example, vancomycin is a large-spectrum antibiotic produced by *Amycolatopsis orientalis* (also known as *Nocardia orientalis*) [72,73]. The NRPS biosynthetic machinery is similar to the type I PKSs, as it uses modular multi-domain enzymes (*Figure 1.3b*), each domain has a function to modify the precursors amino-acids to condense in the NRPS final compound. Many NRPS BGCs contain tailoring enzymes that allow modification of the compound, such as the addition of sugar groups in vancomycin (*Figure 1.3b*) [73].

The RiPP compounds are used for diverse applications, such as food additive and preservative (e.g., nicin), or antibiotics (e.g., thiostrepton, GE2270A). Many RiPP compounds are produced in Actinobacteria, such as GE2270A a thiopeptide produced in *Planobispora rosea* [74], with derivatives in clinical trials as antibiotics [75,76]. In opposition to NRPS peptides, RiPP peptides are produced by the ribosome using the transcript from the BGC. The ribosome produces a precursor peptide containing a leader peptide acting as a signal system and a core peptide that will be modified to become the final compound (*Figure 1.3c*). The leader peptide is then cleaved and the core peptide is further modified by different enzymes encoded in the BGC (e.g., cyclodehydration) which creates the final compound (e.g., GE2270A, or thiostrepton) [75].

Secondary metabolites biosynthesis consumes metabolic precursors from the primary metabolism. Ones of the most important precursors are the acyl-CoA metabolites, particularly acetyl-CoA and malonyl-CoA, that are used for PKS and terpenes (via MVA pathway) biosynthesis [58,59,77]. The amino-acids are also important precursors for the peptides secondary metabolites NRPS and RiPP [61,73]. Most of the secondary metabolites pathways have enzymes using cofactors such as NADPH, NADH, FMNH, FADH, ATP, or S-adenosyl methionine (SAM) [16,78]. This explains the importance of the primary metabolism regulation and metabolic precursor availability for secondary metabolites production [24].

1.1.3 *Streptomyces coelicolor:* the model strain of antibiotic production

Streptomyces coelicolor is a well-studied *Streptomyces* strain; it is the model organism for antibiotics production in the *Streptomyces* genus [79]. It was the first

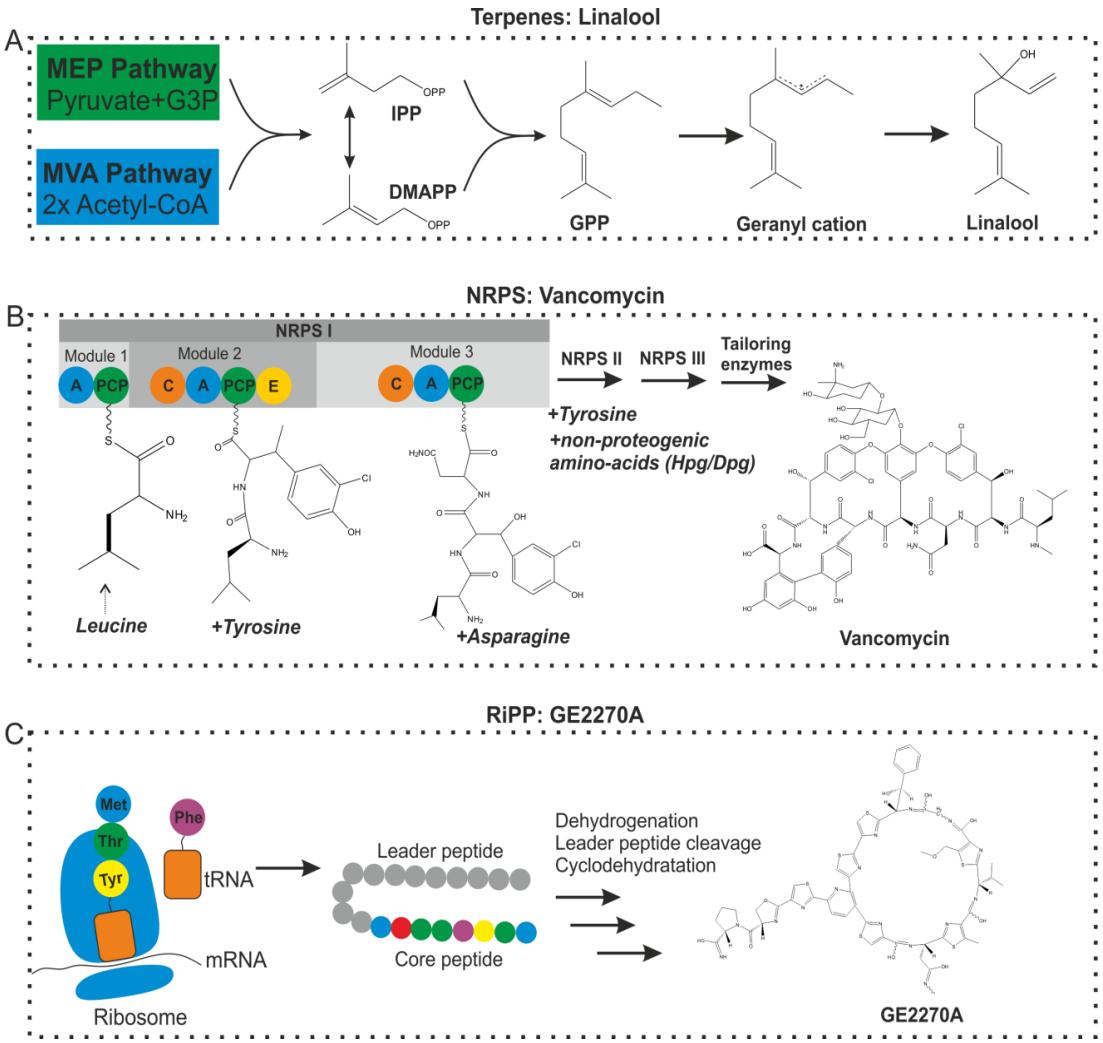


Figure 1.3: Schematic representation of terpenes, NRPS, and RiPP biosynthetic pathways

A) Example of a terpenes biosynthetic pathway for linalool [59]. The terpenes precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are produced from the mevalonate (MVA) pathway and/or from the non-mevalonate pathways (MEP). The precursors are transformed into the common intermediates geranyl pyrophosphate (GPP) subsequently transformed in a wide range of terpenoids.

B) Example of a non-ribosomal peptides synthase (NRPS) for vancomycin [73]. The NRPS is divided in multiple modules using different proteogenic amino-acids like leucine and tyrosine, or using non-proteogenic amino-acids like the 4-hydroxyphenylglycine (Hpg) and the 3,5-dihydroxyphenylglycine (DPG). The compound is then modified by tailoring enzymes. A = activation domain; C = condensation domain; E = epimerization ; PCP = peptide carrier protein domain.

C) Example of ribosomally synthesized and post-translationally modified peptides (RiPP) for GE2270A [75]. The RiPP precursor is produced by ribosomes using the BGC transcript. The precursor peptide is constituted by a leader peptide that will be cleaved and a core peptide that will be modified by diverse enzymes encoded in the BGC (e.g., dehydrogenation and cyclodehydratation).

Streptomyces with its genome fully sequenced [79]. More than 20 BGCs were identified in its genome sequence, including type I, II & III PKSs, terpenoids, NRPS, and RiPP. The most studied secondary metabolites are Actinorhodin (ACT) — a type II polyketide synthase (PKS) product (*Equation 1.1*), Coelimycin P1 or Coelicolor PolyKetide (CPK) — a type I PKS product (*Equation 1.2*), Calcium-Dependent Antibiotic (CDA) — a non-ribosomal peptide synthase (NRPS) product (*Equation 1.3*), and Undecylprodigiosin (RED) — a hybrid NRPS-PKS product (*Equation 1.4*). A major advantage of this organism is the production of multiple coloured antibiotics allowing fast visual assays of the antibiotic production by the organism [79]. ACT is a pH-dependent blue-pigmented antibiotic [64], CPK is a yellow-pigmented antibiotic [80], undecylprodigiosin is a pH-dependent tripyrrole red-pigmented antibiotic [81]; all three are quantifiable by spectrophotometry. All these compounds are produced from metabolic precursors coming from the central metabolites such as acetyl-CoA and malonyl-CoA, cofactors such as NADH, NADPH, $FADH_2$, or $FMNH_2$ (*Equation 1.1, Equation 1.2, Equation 1.3, Equation 1.4*), and amino-acids such as glutamate, or alanine (*Equation 1.3*).

The wide range of molecular tools and protocols [23,82] used and tested in the *S. coelicolor* strain, makes it an attractive organism for synthetic biology [15]. The *S. coelicolor* molecular toolbox is expanding, with the recent addition of an engineered CRISPR/Cas system for genome-editing in *Streptomyces* species [83]. Multiple

Actinorhodin (ACT)

$$16\ \text{acetyl-CoA} + 15\ \text{ATP} + 4\ \text{NAD}^+ + 12\ \text{NADPH} + 5.5\ \text{O}_2 + 4\ \text{FMNH}_2$$

$$\downarrow$$

$$\textit{Actinorhodin} + 15\ \text{ADP} + 15\ \text{Pi} + 4\ \text{NADH} + 12\ \text{NADP}^+ + 16\ \text{CoA-SH} + 4\ \text{FMN}^+ + 13\ \text{H}_2\text{O}$$

Equation 1.1: Stoichiometric equation of Actinorhodin

Coelimycin P1 (CPK)

6 malonyl-CoA + NADH + alanine + glutamate + $FADH_2$ + 2 O_2 + acetyl-cysteine

⬇

Coelimycin P1 + 6 CO_2 + 7 H_2O + pyruvate + α-ketoglutarate + 4 CoA-SH + SH-ACP + FAD^+

Equation1.2: Stoichiometric equation of Coelimycin P1

Calcium Dependent Antibiotic form 1b (CDA1b)

2 ribose-5-phosphtae + 3 erythrose-4-phosphate + 4 3-phosphoglycerate + 6 phosphoenolpyruvate + 4 acetyl-CoA+ 5 oxaloacetate + 2 α-ketoglutarate + 26 ATP + 5 NAD^+ + 20 NADPH + 14 NH_3 + 4 O_2 + tetrahydrofolate + $FADH_2$ + FMN^+ + pyridoxal-phosphate + HS-ACP

⬇

Calcium-Dependent Antibiotics 1b + 2 glyceraldehyde-3-phosphate + 2 pyruvate + 11 ADP + 15 AMP + 47 Pi + 3 PPi + 5 NADH + 20 $NADP^+$ + 5 CO_2 + methyl-tetrahydrofolate + 4 CoA-SH + succinate + FAD^+ + $FMNH_2$ + phenyl-methyl-pyrazolone + H_2O_2 + Acetyl-S-ACP + 27 H_2O

Equation 1.3: Stoichiometric equation of Calcium-Dependent Antibiotic form 1b

Undecylprodigiosin (RED)

2 3-phosphoglycerate + 8 acetyl-CoA + α-ketoglutarate+ 11 ATP + 2 NAD^+ + 15.5 NADPH + 3 NH_3 + tetrahydrofolate + 2 FAD^+ + 0.5 FMN^+ + S-adenosylmethionine

⬇

Undecylprodigiosin + 11 ADP + 13 Pi + 2 NADH + 15.5 $NADP^+$ + 2 CO_2 + methyl-tetrahydrofolate + 8 CoA-SH + 2 $FADH_2$ + 0.5 $FMNH_2$ + S-adenosylhomocysteine + 13 H_2O

Equation 1.4: Stoichiometric equation of Undecylprodigiosin

S. coelicolor mutant strains have been developed [84–86], such as the M1146 strain where the BGCs for the highly active secondary metabolites ACT, RED, CPK, and CDA were deleted from the chromosome [84]. This frees metabolic resources for heterologous pathways introduced in the strain [87]. The M1146 strain is widely used as a host strain for heterologous expression of BGCs from uncultivable or genetically intractable

Actinobacteria. The *S. coelicolor* strain is a useful chassis strain for screening and as a production host for secondary metabolic pathways [84,87]. There are other strains that are potential chassis strains, such as genome-minimized strains of *Streptomyces albus* [88] and *Streptomyces avermitilis* [14], or *Streptomyces clavuligerus* [89].

1.2 Introduction to Synthetic and Systems Biology

Synthetic biology definitions are very diverse and are mostly linked to the many goals of the discipline, from biomanufacturing of high-value chemicals, to the engineering of living therapeutics, or the creation of minimal cells and protocells [1,7,90]. Furthermore, the field draws methodologies from different disciplines, such as biochemistry, microbiology, chemistry, medicine, engineering, mathematics, computational biology, and systems biology [6,7,90]. However, despite the ambiguities around the boundaries of what is and what is not synthetic biology, many groups of scientists attempted to reach a general definition of the field. This includes the definition by the new and emerging science and technologies (NEST) group (Synthetic Biology: Applying Engineering to Biology: Report of a NEST High Level Expert Group) [91]:

> *'Synthetic biology is the engineering of biology: the synthesis of complex, biologically based (or inspired) systems, which display functions that do not exist in nature. This engineering perspective may be applied at all levels of the hierarchy of biological structures—from individual molecules to whole cells, tissues and organisms. In essence, synthetic biology will enable the design of 'biological systems' in a rational and systematic way'*

While synthetic biology enables the rational engineering of biological systems, it needs systems biology to understand existing and engineered systems from a

quantitative point of view [92–94]. Systems biology uses computational and mathematical methods to understand biological systems using experimental data (e.g., omics data) and biological knowledge (e.g., biochemistry, or molecular biology) [92,95,96]. So, combining the two disciplines creates benefits to both, as synthetic biology gains from the analysis and tools of systems biology to design and engineer organisms with the desired function (e.g., bioproduction of antibiotics) [92,97,98], while systems biology gains knowledge from synthetic biology experiments that modify biological organisms and reveal new functions, as well as creating new biological systems to study [98,99].

1.2.1 Synthetic biology engineering principles

The ambitious objective of synthetic biology to engineer biological systems requires to standardise and systematise the organism's modification in a similar manner as software or electrical engineering [7]. Synthetic biology uses the iterative design-build-test-learn (DBTL) cycle to engineer biological systems [100,101] (*Figure 1.4***)**. This cycle builds a framework to engineer biological systems in a more systematic and efficient way [102]. The DBTL cycle described here focuses on synthetic biology for secondary metabolites production.

In the design phase, to engineer a given function, such as bioproduction of an antibiotic, the synthetic biologist either selects a native producing strain or a chassis capable to produce the compound [103,104] (*Figure 1.4*). In the latter case, the metabolic pathway is either readapted from the native pathway by refactoring and redesigning the genetic construct for the pathway, or by designing and selecting new pathways and enzymes [105,106]. The metabolic network of the producing organism is also redesigned to increase fluxes toward compound production [107,108].

The organism designs are implemented in the build phase; multiple experimental methods are used, particularly molecular biology and biochemistry methods such as gene synthesis, pathway assembly, and gene editing (*Figure 1.4***)**. For example, multiple pathways can be built at the same time using automated combinatorial assembly [8] with different combinations of genetic components and the best constructs will be identified in the test phase. The designed BGC pathways are built by gene synthesis, pathway assembly, and expressed in the organism of choice [8]. The metabolic network is modified using diverse methods such as overexpression and gene deletion, applying gene editing methods such as CRISPR/Cas [109] and molecular biology to overexpress a target gene [110].

The test of the reengineered organisms is necessary to assess if the desired function is optimised or to identify potential issues and bottlenecks in the engineered organism (*Figure 1.4*). The organism phenotype characteristics can be assessed by measuring growth parameters, nutrients consumption, and compound export. Some very valuable information on the engineered organism state are measured by omics data (transcriptomics, proteomics, and metabolomics) [111]. Multiple screening methods are used to study if the function is optimised or if a new function is acquired, for example, new antibiotics produced are tested for their bioactivity, or a bioassay used to identify the most expressed pathway using a fluorescent reporter (e.g., GFP) embedded in the construct [8,112].

The test results are used to redesign the organism to further optimise it or determine if it reached its maximal potential (*Figure 1.4***)**. This is the learn phase, which uses *in-silico* methods to analyse and interpret the data from the test phase, such as statistical analysis of the data (e.g., growth curve, or pathway enzymes

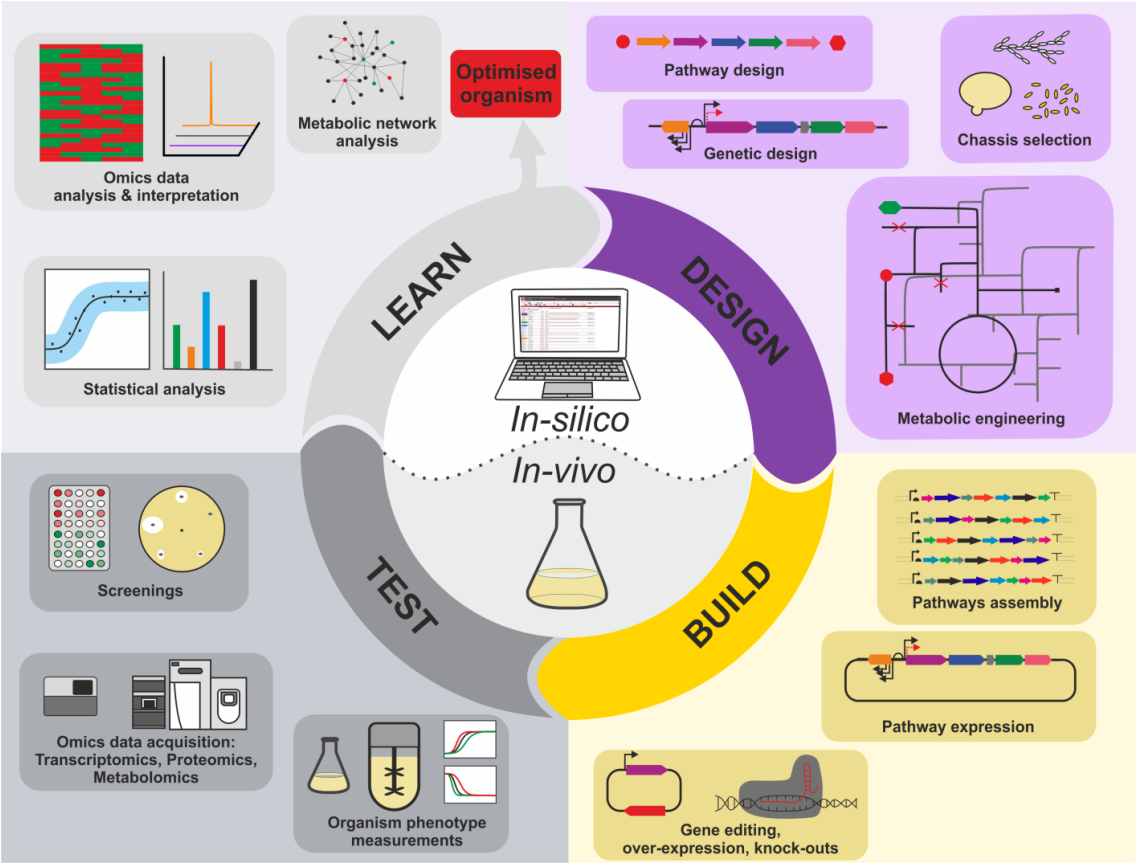


Figure 1.4: Design-Build-Test-Learn cycle in synthetic biology

This DBTL cycle focuses on the synthetic biology of secondary metabolites production. The Design and Learn phases are dominated by *in-silico* methods, while the Build and Test phase are dominated by *in-vivo* methods.

- Design: the design of engineered organism spawn from the selection of the chassis strain, to the metabolic pathway design and genetic construct design to the organism metabolic network engineering.
- Build: the build phase executes the design phase plans by assembling multiple pathways, expressing the genetic constructs, and engineering the strains using gene editing and molecular biology methods to over-express or knock-out target genes identified in the design phase.
- Test: here the organisms built in the previous phase are tested, such as their phenotype (e.g., growth, and nutrient exchanges), the omics data is acquired (transcriptome, proteome, metabolome), and the organism function is optimized (e.g., bioassays).
- Learn: the data from the test phase is analysed and interpreted, from statistical analysis to omics data analysis and integrative analysis in metabolic models. The data is also used to predict and design a new round of modifications on the organism or assess if the function is optimised.

The DBTL cycle is repeated until the function, here high production of a secondary metabolite, is optimised in the target organism [101,102].

expression) to understand the link between design and the resulting function measured in the test phase [101,113]. The analysis and interpretation of the omics data help to identify the impact of a design on the organism, and identify undesirable effects (e.g., by-product measured by metabolomics) [95,101]. The omics data is integrated into a genome-scale metabolic model to further interpret the data, and predict a new design strategy (e.g., deletion of a by-product producing pathway) [107,114].

These steps are repeated as many times as necessary until the organism function is optimised (e.g., high production of a secondary metabolite). The automation of this cycle is critical to accelerate the cycle and rapidly engineer biological systems [2,101]. However, the cycle can be often stalled by the pace of interpreting and integrating the test data into the design phase [1]. Hence, there is a need to develop new automated methods in the learn phase to accelerate the DBTL cycle [1,99,101].

1.2.2 Synthetic biology and metabolic engineering of secondary metabolism

The recent and rapid expansion of synthetic biology is facilitating a new route to natural product discovery by rational engineering of secondary metabolites BGCs [8,115], and to bioproduction optimization by rationally engineering production hosts [116,117]. Synthetic biology of secondary metabolism is bringing a more systematic approach to natural products discovery and production [8]. For example, by moving away from traditional methods such as random mutagenesis to activate cryptic BGCs toward systematic methods based on computational genome-mining and gene-editing to identify new BGCs and then delete the regulator with tools like CRISPR/Cas9 to awaken the production [109,118]. There are three main approaches in the synthetic

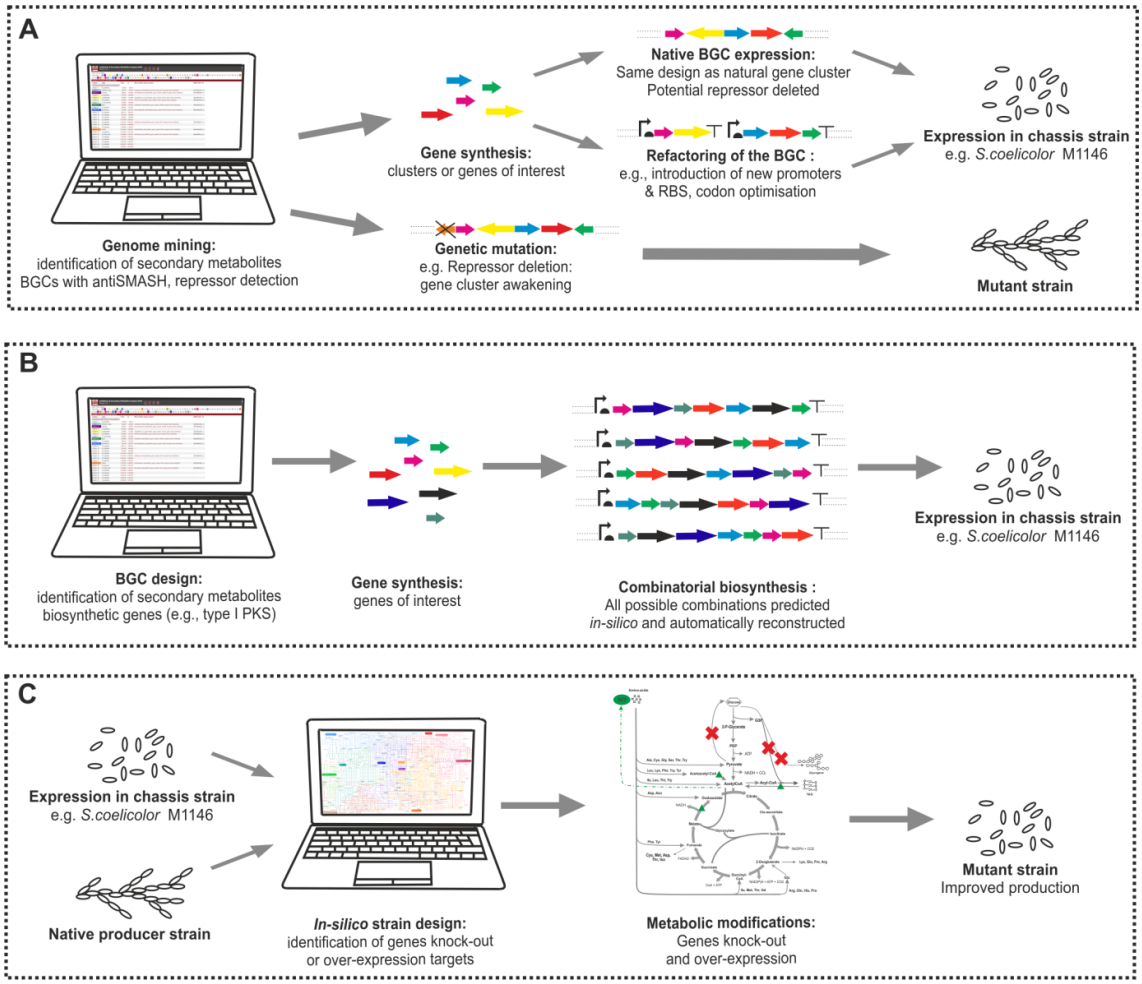


Figure 1.5: Different synthetic biology approaches to initiate and improve production of secondary metabolites

a) The discovery of new biosynthetic gene clusters secondary metabolites. After identification of new BGCs using *in-silico* methods (e.g., genome mining). If the native producer is cultivable and genetically tractable, the cluster expression can be activated by deleting any potential repressor. Otherwise, the gene sequence of the biosynthetic genes can be synthesised and expressed in a chassis strain. Either by expressing the native BGC with potential repressors deleted or by refactoring the whole BGC and putting it under control of an artificial inducing system. The product can be then identified and screened.

b) Combinatorial biosynthesis to discover new chemistries. The possible combinations of modular biosynthetic genes (e.g., type I PKS and NRPS) are predicted *in-silico* then the gene are synthesised and combined on a large scale in vectors. Then these are expressed in the chassis strain to be screened.

c) Improving production in chassis strains and native producers. The production in a chassis strain or a native strain is simulated and studied *in-silico*, where gene knock-outs and overexpression targets are predicted. The strain metabolic design is optimised and a set of key genes are knocked-out and/or overexpressed to increase production.

biology of secondary metabolism: (1) discovery of new BGCs secondary metabolites, (2) design of new secondary metabolites, and (3) production optimization (*Figure 1.5*) [8,119,120].

The discovery of new BGCs secondary metabolites (1) using synthetic biology is driven by a combination of computational methods (i.e*., in-silico* identification of BGCs), molecular biology tools (e.g., golden-gate assembly), and decreasing cost of gene sequencing and synthesis (*Figure 1.5a*) [8,117]. The increasing availability of genome sequences and meta-genomic data combined with the development of genome mining tools [121] enabled the discovery of thousands of unknown secondary metabolites BGCs [13]. A first method, already widely used in the natural product research consists of deleting the BGC regulator identified *in-silico* to trigger the secondary metabolite production in the native strain [122–124]. The secondary metabolite is identified and characterized using analytical methods such as liquid chromatography mass-spectrometry (LC-MS), where the producer and non-producer chromatogram and mass-spectrum are compared to identify the secondary metabolite produced. Then, if the compound has a different mass-spectrum and fragmentation patterns from existing secondary metabolites (dereplication), the compound can be purified and its structure defined by nuclear magnetic resonance (NMR) [122,125,126]. This method was successfully applied in the activation of CPK in *S. coelicolor* which was awakened by targeted inactivation of the pathway repressor (ScbR2) within the BGC, leading to the identification of CPK and an antimicrobial metabolite (abCpk) [127]. However, these methods to identify secondary metabolites is still low-throughput and are not adapted to allow the rapid characterization of the thousands of unknown BGCs [13]. Furthermore, this method only works if the metabolite is produced at levels high enough for

purification / identification and if the native strain is culturable and genetically-tractable [38,39]. However, the recent developments in synthetic biology and computational biology opened the opportunity to accelerate this process, by identifying the BGCs *in-silico* then re-design and refactor the BGCs for optimal expression in the host strains. The BGCs are refactored by introducing different promoters and ribosome binding sites (RBS), replacing regulatory systems with inducible systems, and by codon optimization [128,129]. The redesigned BGCs are expressed and produced in a chassis strain to allow characterization of the compound. Then, if the secondary metabolite is of interest (e.g., bioactivity against resistant microorganisms) the production can be improved by engineering a chassis strain adapted for high-titers production. This process of plug-and-play discovery and production of secondary metabolites is described in detail in a review by Medema, et al. [104].

The engineering of BGCs to create new chemistry (2) is another application of synthetic biology, where the modularity of some biosynthetic genes (e.g., type I PKS and NRPS) is exploited to create new chemical structures [104,115] (*Figure 1.5b*). Once the biosynthetic gene sequences from the BGCs are synthesized, high-throughput DNA assembly methods help to test multiple designs in parallel; the constructs are then expressed and screened in a chassis strain. This method can be used to exploit the modularity of type I PKS and NRPS for combinatorial biosynthesis [60,115,130]. The redesigned BGCs are then expressed in a chassis strain and screened to identify new molecules. The chassis strain has to be adapted to produce the type of compounds screened (e.g., type I PKS) to ensure its metabolism produces the right precursors (e.g., acyl-CoA) [104]. The screening uses analytical methods (e.g., LC-MS, and NMR) to identify

and characterize the novel chemical structures, and bioassays to identify the bioactivity of the compounds (e.g., antibacterial spectrum).

The optimization of secondary metabolites production (3) is necessary to reach the highest production titers possible for downstream tests of the metabolites (e.g., clinical tests of the compound) and to decrease the costs of bioproduction (*Figure 1.5c*). The production optimization of existing and engineered BGCs is done in the native strain [51] or in a heterologous host [87]. Multiple approaches are used to increase metabolic production, such as refactoring of the BGCs [8,128], metabolic engineering [131], and regulatory machinery refactoring [128]. The BGC refactoring is frequently necessary to ensure optimal expression in a heterologous host as described above (1). Metabolic engineering methods to improve production in native and heterologous hosts help to engineer the strain primary metabolism to generate higher metabolic fluxes toward the precursors of the secondary metabolic pathways [50,51,132]. As well as cutting metabolic fluxes toward unessential metabolisms wasting resources that could be directed toward production and growth [77,133,134]. For example, metabolic engineering in *Streptomyces argillaceus* enabled a 229% increase of production of mithramycin (a polyketide antitumor); by cutting fluxes going toward triacylglycerides and glycogen storage, while increasing fluxes toward the precursors glucose-1-phosphate and malonyl-CoA [77].

Heterologous production of a secondary metabolite requires a sufficient understanding and control over the host organism, to ensure straightforward expression and production of the secondary metabolite, hence the use of well-studied chassis strains such as *S. coelicolor* [84,87,103]. When optimising production in a native strain, sufficient understanding of the strain metabolism is required to rationally modify its metabolism and redirect fluxes toward secondary metabolites production. These

approaches are now accelerated by the development of genome-scale metabolic modelling methods that helps to identify targets to redirect metabolic fluxes toward secondary metabolites [56,135,136].

1.2.3 Metabolic modelling

The recently emerging high-throughput DNA sequencing technologies have resulted in an explosion in the number of genome sequences available for organisms across the entire domain of life [74]. In parallel, new computational methods were developed to rapidly reconstruct genome-scale metabolic models (GSMMs) from available genome sequences [137–140]. Therefore, this allowed the reconstruction of many GSMMs, which have been applied for strain engineering through synthetic biology [98,141,142]. Similarly, the rapid developments in high-throughput technologies to acquire 'omics data have extended the scope of GSMMs to integrate and interpret these data to better understand cellular systems[111,137,142,143]. The increasing availability of 'omics' datasets and high-quality metabolic models helps in the rapid testing and debugging of engineered biological systems [94,144,145]. Nevertheless, a major challenge nowadays is not to generate more data, but rather to interpret and analyse these data in the context of biological systems, in order to understand and predict major biological behaviours or phenotypes [1,146,147].

1.2.3.1 Reconstruction and analysis of genome-scale metabolic models

Biological organisms are complex systems formed of networks of biological parts (e.g., metabolites, proteins, or genes) with subsystems of different complexity levels interacting with each other [148]. For example, metabolites are consumed and produced by metabolic reactions forming metabolic pathways which are part of the

whole metabolism, this metabolic network interacts with a network of proteins and genes [137]. Systems biology is a holistic approach to study and understand natural and engineered biological systems, by modelling and analysing the networks of interacting biological parts [92,97,137,149]. During the last 15 to 20 years various modelling approaches to study metabolic networks have been developed in the field of systems biology [92,97,98]. A particularly interesting approach is constraint-based metabolic modelling [138], which applies constraints (e.g. thermodynamic or stoichiometric) on an entire metabolic network to limit the possible metabolic fluxes distribution and predict the optimal metabolic fluxes of an entire organism *in silico* [137]. This approach has been increasingly applied to understand and predict metabolism, particularly to engineer better secondary metabolites producers strain [56,114,135,150]. A major advantage of constraint-based metabolic modelling compared to other modelling methods is the possibility to model an organism's metabolism from its genome sequence without the need of precise knowledge of parameters such as enzymes kinetic or metabolite concentrations [140,151,152]. Therefore, this method helps to predict phenotype characteristics (e.g., ability to grow using certain nutrients, genes essential for growth, etc.) from an organism genotype [146]. As a large number of prokaryotic genomes are available, including of *Streptomyces* and Actinobacteria strains, these available genome sequences opens the possibility to reconstruct genome-scale metabolic networks for a large number of Actinobacteria strains to study their metabolism *in silico* [22,153].

In short, four main steps are involved in the process of metabolic network reconstruction [137]: an automated draft reconstruction from a genome-annotation, a curation of this draft reconstruction by using different data sources (literature,

textbooks, and specialized databases), the conversion of this curated model into a mathematical representation as a stoichiometric matrix to enable metabolic fluxes prediction by constraining the model; and, finally, the validation and improvement of the model using experimental data [137,139] (*Figure 1.6*). The detailed standard protocol to reconstruct a high-quality genome-scale metabolic model (GSMM) is described in a step-by-step protocol by Thiele and Palsson [139]. A genome sequence annotation or a similarity-based annotation is used to build a draft reconstruction, with an initial set of biochemical reactions corresponding to putative metabolic genes annotated, the models built automatically will contain misannotations and gaps. Which necessitates applying a semi-manual curation to fill gaps, the gaps are filled with known reactions semi-manually (using databases) and other gaps are automatically filled with appropriate algorithms [154]. Then, the model is manually curated, by adding and modifying metabolic pathways and genes associated by using the literature. Once the obvious misannotations are corrected and the gaps are filled, the next step is to translate the model into a mathematical model, in a format compatible with available software [155–157]. The model goes through initial testing steps to ensure the quality of the model, such as mass and charge balance of the network [158]. The model is further validated to verify the ability to predict physiological functions (e.g. biomass or ATP production) depending on the environmental constraints (e.g. media and nutrients uptakes). Finally, to improve the model, different omics datasets are used, such as fluxomics, metabolomics, proteomics or transcriptomics, these data are in general used in an iterative way to identify incorrect predictions by the model and correct it [94] (*Figure 1.6*).

1.2.3.2 Constraint-based metabolic modelling

Once the GSMM is ready for analysis, the metabolic reactions are represented in a mathematical format as a stoichiometric matrix (*S*) of size $m \times n$, with *m* (number of rows) representing the metabolites, *n* (number of columns) representing the different reactions in the model, and the entries of the matrix represents the stoichiometry of the metabolites involved in the metabolic reactions [137,139] . When a substrate is consumed by a reaction, it has a negative coefficient, while the products of a reaction have a positive coefficient. Stoichiometric matrices (*S*) for GSMMs are sparse matrices; most of the entries are zero because the vast majority of the metabolites (with a few exceptions, such as cofactors) are involved in a limited number of reactions (*Figure 1.7*). *v* represents the vector containing all the fluxes going through the network of biochemical reactions [137]. At steady state, the flux through each reaction is given by $S \times v = 0$, which corresponds to a system of linear equations. As the number of metabolites is larger than the number of reactions ($m > n$), there is more than only one solution to these linear equations. Although, by constraining the system and optimizing the solution for a particular objective function (e.g., growth), so the algorithm can find an optimal point in the space of possible solutions when solving the linear equations [137] (*Figure 1.7*). In general, the objective function for a microbial model is growth maximisation, going from the hypothesis that the main aim of a bacterium is to proliferate; other objective functions can be maximization of ATP production, minimization of nutrient consumption, or multiple sets of objectives such as optimizing biomass and antibiotic production at the same time. The main constraints applied in the model are mass balance and reactions upper and lower fluxes constraints. The mass balance ensures that the numbers and the mass of the

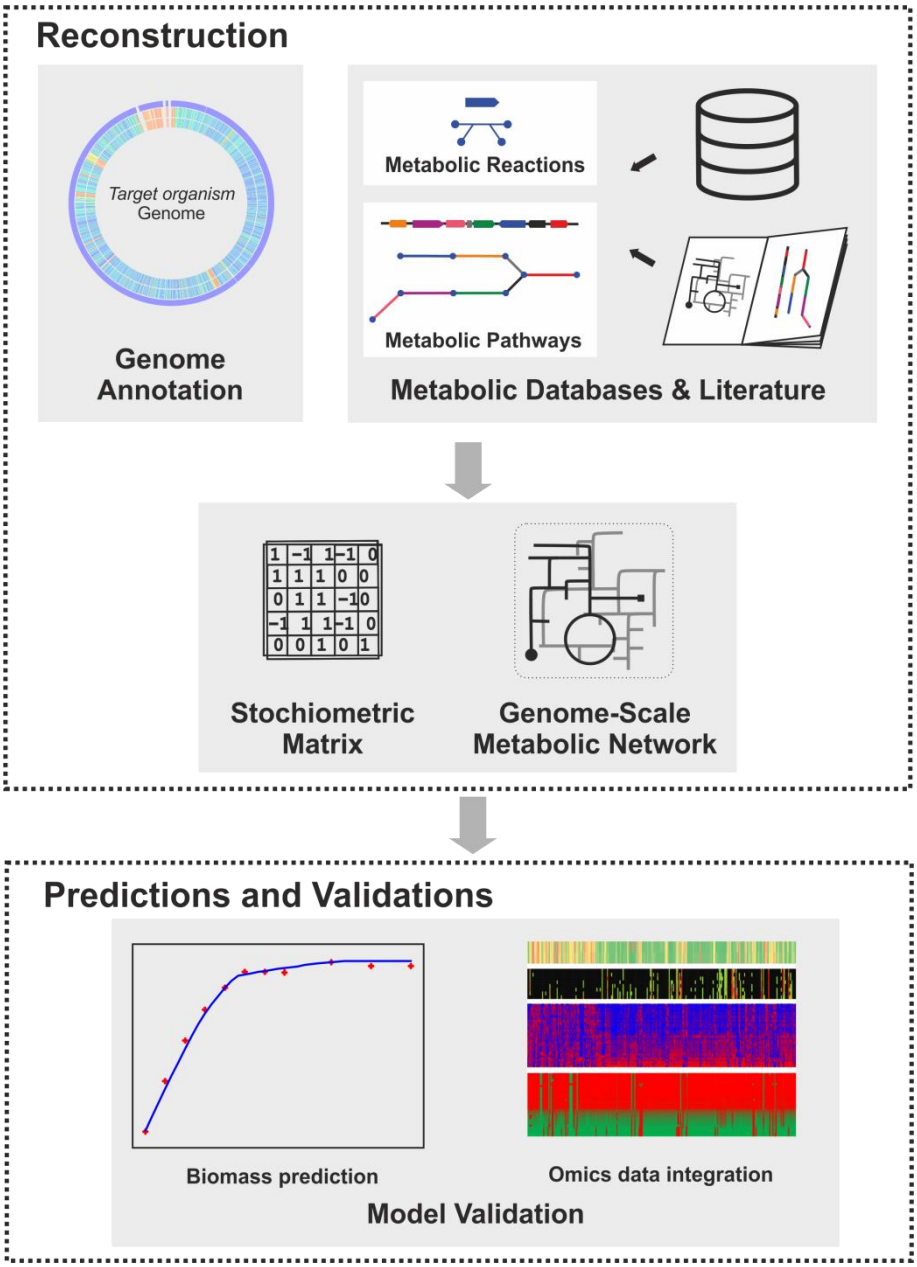


Figure 1.6: Summary of a genome-scale metabolic model reconstruction

The information extracted from genome annotation of the target organism, from specialised metabolic databases (e.g., BiGG), and from literature are collated together to identify the metabolic reactions and metabolic pathways to build the genome-scale metabolic model. The genome-scale metabolic network is transformed in a stoichiometric matrix [137,139].

The metabolic model objective function (e.g., biomass) is optimised under a given set of constraints representing the environmental conditions, the prediction are compared to experimental data to validate the model. Omics data integration is used to validate and improve the model [137,139].

atoms of the metabolites produced correspond to the numbers and mass of the atoms of the metabolites consumed; this respects the basic principle of mass conservation. The flux bounds contain thermodynamic constraints such as reversibility or irreversibility of a reaction taken into account in the model, avoiding biologically impossible fluxes in the model. Flux bounds are also used to fix minimum or maximum values of fluxes through a reaction, depending on experimental or environmental conditions (e.g. medium or flux measurements) [137,139].

Flux balance analysis (FBA) is the main constraint-based metabolic modelling method used [139,152]. FBA is a very useful approach to study large and complex metabolic networks, as it does not require a massive computational power even for very large networks, and this is particularly useful for systems that lacks information such as enzyme kinetic parameters. FBA assumes that metabolites production and consumption are balanced; this is the steady-state assumption [152]. This assumption is derived from the fact that metabolism reaches a stable state faster than other cellular processes (e.g., gene expression), so at a given time point metabolism should be at a quasi-steady state. FBA is performed on a constrained stoichiometric matrix S and maximises or minimises an objective function. In general, the objective function chosen for a model is the biomass production, as this is the function representing the major cellular processes of the cell (e.g., DNA/RNA synthesis, protein synthesis, or membrane synthesis) [139,159,160]. The biomass reaction represents the cell biomass as defined ratios of macromolecules produced from metabolites [140]. The substrate-to-biomass conversion is described quantitatively in the model using mmol for the metabolites consumed and gram of dry weight (gDW) for the cells' biomass. The stoichiometry of the biomass produced in the reaction have a molecular weight of

1 g.mmol^{-1} to compare with the specific growth rates [140]. As an example, the different metabolites consumed and produced to generate 1 gram of dry weight of *Streptomyces coelicolor* in the *i*AA1259 metabolic model are shown in the *Table 1.1*.

Reactants metabolites for the biomass reaction (to generate 1 gram of Dry Weight)							
Fatty acids composition	**mmol/g DCW**	***Triacylglycerol***	**mmol/g DCW**	**Vitamines & cofactors**	**mmol/g DCW**	**Amino-acids**	**mmol/g DCW**
Phospholipid		Triacylglycerol - Myristiate(n14:0)	0.000364	Thiamine diphosphate	0.00481	Alanine	0.60978
Cardiolipin - Myristiate(n14:0)	0.000098	Triacylglycerol - Pentadecanoate(n15:0)	0.000227	NAD	0.003066	Arginine	0.10195
Cardiolipin - Pentadecanoate(n15:0)	0.000061	Triacylglycerol - Palmitate(n16:0)	0.00416	NADH	0.003062	Asparagine	0.18343
Cardiolipin - Palmitate(n16:0)	0.001116	Triacylglycerol - Palmitoleate(n16:1)	0.001068	NADP	0.002744	Aspartate	0.18599
Cardiolipin - Palmitoleate(n16:1)	0.000287	Triacylglycerol - Anteisopentadecanoate(ai15:0)	0.003933	NADPH	0.00274	Cysteine	0.08069
Cardiolipin - Anteisopentadecanoate(ai15:0)	0.001055	Triacylglycerol - Anteisoheptadecanoate(ai17:0)	0.003614	Heme A	0.002387	Glutamine	0.1418
Cardiolipin - Anteisoheptadecanoate(ai17:0)	0.00097	Triacylglycerol - Isomyristate(i14:0)	0.000796	Heme O	0.002427	Glutamate	0.25085
Cardiolipin - Isomyristate(i14:0)	0.000213	Triacylglycerol - Isopentadecanoate(i15:0)	0.002273	Pyridoxal 5'-phosphate	0.008287	Glycine	0.55317
Cardiolipin - Isopentadecanoate(i15:0)	0.00061	Triacylglycerol - Isopalmitate(i16:0)	0.004614	Riboflavin	0.005397	Histidine	0.07991
Cardiolipin - Isopalmitate(i16:0)	0.001238	Triacylglycerol - Isoheptadecanoate(i17:0)	0.001682	Protoheme	0.003306	Isoleucine	0.24831
Cardiolipin - Isoheptadecanoate(i17:0)	0.000451			Siroheme	0.002236	Leucine	0.13981
Phosphatidic acid - Myristiate(n14:0)	0.000068			Menaquinol 9	0.0118	Lysine	0.0708
Phosphatidic acid - Pentadecanoate(n15:0)	0.000042	**Polyamines**		Mycothiol	0.004185	Methionine	0.12059
Phosphatidic acid - Palmitate(n16:0)	0.000778	Spermidine	0.013699	5,6,7,8-Tetrahydrofolate	0.004581	Phenylalanine	0.13521
Phosphatidic acid - Palmitoleate(n16:1)	0.0002	Spermine	0.009842	FAD	0.002592	Proline	0.37332
Phosphatidic acid - Anteisopentadecanoate(ai15:0)	0.000735	Putrescine	0.022527	Inorganic triphosphate	0.000141	Serine	0.21339
Phosphatidic acid - Anteisoheptadecanoate(ai17:0)	0.000676	Cadaverine	0.019494	Lipoate	0.010787	Threonine	0.26691
Phosphatidic acid - Isomyristate(i14:0)	0.000149			Molybdopterin guanine dinucleotide	0.00235	Tryptophan	0.03275
Phosphatidic acid - Isopentadecanoate(i15:0)	0.000425			Undecaprenyl diphosphate	0.002198	Tyrosine	0.06539
Phosphatidic acid - Isopalmitate(i16:0)	0.000863	**Inorganic ions**		5-Methyltetrahydrofolate	0.004431	Valine	0.22152
Phosphatidic acid - Isoheptadecanoate(i17:0)	0.000314	Calcium(II)	0.004952	Adenosylcobalamin	0.001286		
Phosphatidylethanolamine - Myristiate(n14:0)	0.000349	Chloride	0.004952	S-Adenosyl-L-methionine	0.005085		
Phosphatidylethanolamine - Pentadecanoate(n15:0)	0.000218	Cobalt(II)	0.000024	Bis-molybdopterin GDP	0.001282	**Cell wall composition**	
Phosphatidylethanolamine - Palmitate(n16:0)	0.003997	Copper(II)	0.000674	Biotin	0.008349	***Peptidoglycan***	
Phosphatidylethanolamine - Palmitoleate(n16:1)	0.001026	Iron(II)	0.006388	Coenzyme A	0.00266	Murein tetrapeptide	0.02984
Phosphatidylethanolamine - Anteisopentadecanoate(ai15:0)	0.003778	Iron(III)	0.007428	2-demethylmenaquinol 9	0.0118	Murein tripeptide	0.030136
Phosphatidylethanolamine - Anteisoheptadecanoate(ai17:0)	0.003472	Potassium	0.18569	5,10-Methylenetetrahydrofolate	0.00446		
Phosphatidylethanolamine - Isomyristate(i14:0)	0.000764	Magnesium(II)	0.008253	10-Formyltetrahydrofolate	0.004309	***Carbohydrate***	
Phosphatidylethanolamine - Isopentadecanoate(i15:0)	0.002184	Magnesium	0.000658	[2Fe-2S] iron-sulfur cluster	0.011698	UDP-N-acetyl-D-glucosamine	0.083227
Phosphatidylethanolamine - Isopalmitate(i16:0)	0.004433	Molybdate	0.000007	[4Fe-4S] iron-sulfur cluster	0.005849	UDPgalactose	0.166453
Phosphatidylethanolamine - Isoheptadecanoate(i17:0)	0.001616	Ammonium	0.012379				
Phosphatidylinositol - Myristiate(n14:0)	0.000027	Nickel(II)	0.000307			***Teichoic acid***	
Phosphatidylinositol - Pentadecanoate(n15:0)	0.000017	Sulfate	0.004126	**Glycogen and trehalose**		glycerol teichoic acid (n=25)	0.003536
Phosphatidylinositol - Palmitate(n16:0)	0.00031	Zinc(II)	0.000324	Trehalose	0.087643	glycerol teichoic acid (n=25)	0.003536
Phosphatidylinositol - Palmitoleate(n16:1)	0.00008			Glycogen	0.154187	glycerol teichoic acid (n=25)	0.003536
Phosphatidylinositol - Anteisopentadecanoate(ai15:0)	0.000293						
Phosphatidylinositol - Anteisoheptadecanoate(ai17:0)	0.000269	**Other metabolites**		**DNA composition**		**RNA composition**	
Phosphatidylinositol - Isomyristate(i14:0)	0.000059	H+	76.585	dATP	0.016297	ATP	0.11045
Phosphatidylinositol - Isopentadecanoate(i15:0)	0.000169	H2O	7.162	dCTP	0.042326	CTP	0.13226
Phosphatidylinositol - Isopalmitate(i16:0)	0.000344	ATP	75.7	dTTP	0.016297	GTP	0.17554
Phosphatidylinositol - Isoheptadecanoate(i17:0)	0.000125			dGTP	0.042326	UTP	0.10035

Product metabolites for the biomass reaction (to generate 1 gram of Dry Weight)		
Metabolic products	**mmol/g DCW**	
ADP	75.79	1 gram of dry weight of Biomass
UDP	0.24968	
Pi	75.79	
PPi	0.635847	

Table 1.1: Metabolic reactants and products needed to produce 1 gram of dry weight of *Streptomyces coelicolor* in the *i*AA1259 metabolic model

Also, one of the assumptions is that microorganisms naturally try to grow as much as possible, particularly microorganisms to outcompete other organisms [160,161]. Thus, growth is frequently the objective function with the most accurate metabolic predictions overall, particularly under normal laboratory conditions, where a single strain is grown in flask or a bioreactor, where the organism will grow until it consumes all the nutrients [94,160]. However, it is useful to also consider other possible objective functions, such as overproduction of a compound (e.g., antibiotics) which can create a new metabolic burden and compete with the biomass function [160,162] Under different environmental pressure organisms can evolve to adapt to slow growth rather than

maximised growth, for example, pathogenic microorganisms can benefit from slow growth to escape the immune system and propagate without being detected [163]. So, the maximisation of an objective function, such as biomass, is not always adapted to all organisms. To enable analysis of this type of biological systems, a method was derived from FBA to minimise the sum of total fluxes at fixed values of a target flux (e.g., biomass) replacing the objective function of FBA [164]. The objective function is a linear combination of flux vectors. The objective is optimised by FBA using linear programming to solve the equation $S*v=0$, with a set of upper and lower bounds on the vector v [152]. The FBA output is a flux distribution (v) needed to maximise or minimise the objective function. Other methods were derived from FBA, such as parsimonious flux balance analysis (pFBA) that minimises the sum of total fluxes and maximises the objective. For pFBA, the assumption is that under maximal growth the individuals selected in a population are the ones capable to rapidly and efficiently grow on a substrate using a minimum amount of enzymes [165]. FBA methods enable to do a lot of *in-silico* experiments quickly and test many different conditions or hypotheses (e.g., medium composition or gene essentiality) [108,152]. Hence, genome-scale metabolic models and FBA are extensively used to predict phenotypes and metabolism to design organisms for metabolic engineering and synthetic biology, particularly for the production of secondary metabolites [48,56,131,150,166]. For example, the production of rapamycin was increased by more than 140% in *Streptomyces hygroscopicus* by predicting influential gene knockouts using FBA [57].

Obviously, the constraint-based metabolic modelling approach presents some limitations: metabolite concentrations cannot be predicted, it is only suitable for fluxes predicted at steady state (but the dynamic can be simulated by time points

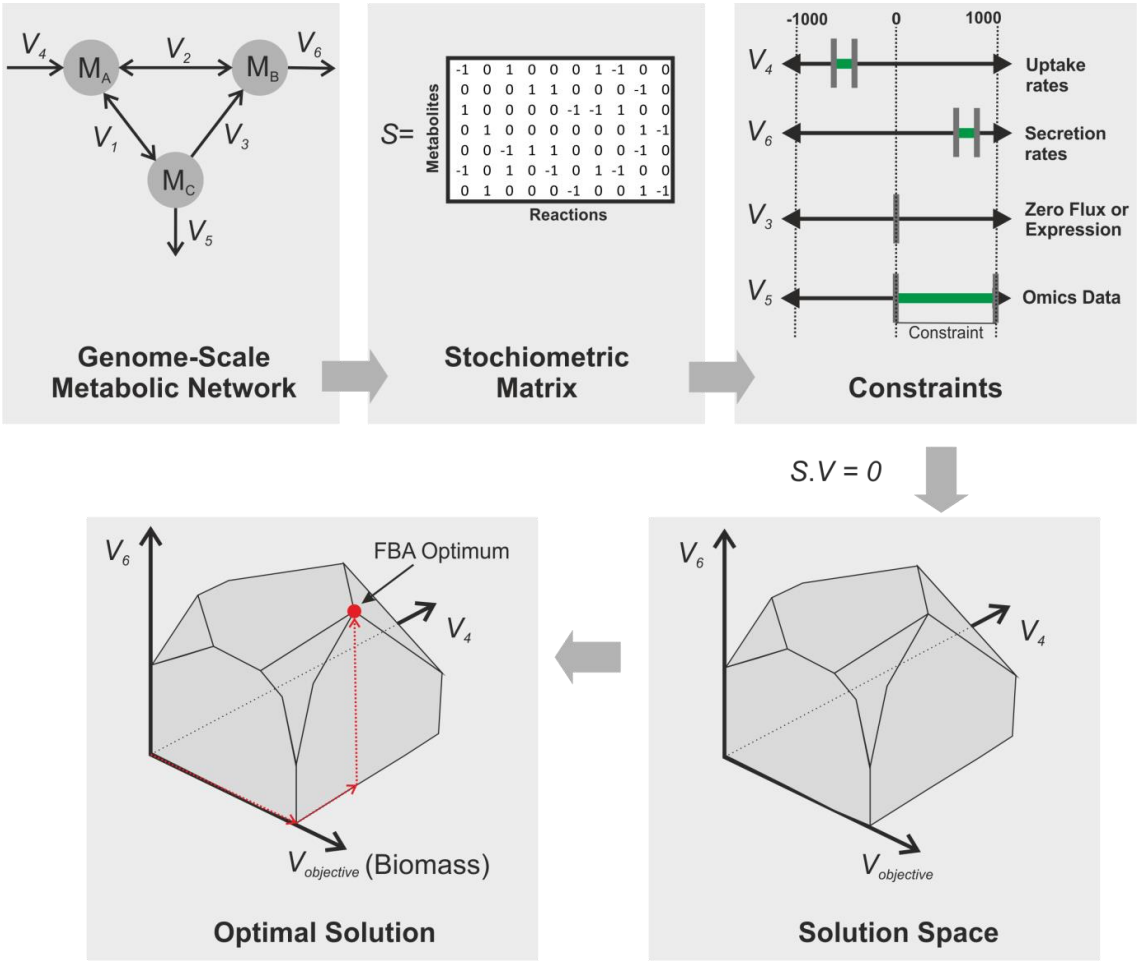


Figure 1.7: Representation of constraint-based metabolic modelling

The genome-scale metabolic network contains metabolites (*M*) as nodes and reactions represented by a flux vector (*V*). This network mathematical representation is a stoichiometric matrix (S) with metabolites and reactions. The flux vectors direction is a thermodynamic constraint on the reaction direction. The matrix rows represent the metabolites consumed (-1), produced (1), and untouched (0), and the columns represent the reactions. The fluxes vectors can be constrained, such as the uptake and secretion rates measured. The fluxes are set to zero if the flux measured (from fluxomics or metabolomics data) are measured as null or the genes or protein associated with the reaction are measured as null (from transcriptomics or proteomics data). The omics data can be used to set loose constraints such as direction of a reaction, for example, a reaction can be switched on to force production of a metabolite if it was measured in the metabolomics data. The constraints from the vectors create a constrained solution space, where the multiplication of S and V is null. The optimal solution to the objective (e.g., biomass) in this space is identified by flux balance analysis (FBA) [137,139,152].

experiments), and it does not account for regulatory effects in its standard implementation. However, during the last years efforts have been made to develop models that integrate prediction of regulatory effect and kinetic information enabling more quantitative predictions [151].

1.2.4 Omics data and metabolic modelling

The availability of omics data for many organisms under different conditions has increased with the rapid development of analytical methods such as LC-MS for metabolomics and proteomics data and whole transcriptome shotgun sequencing for transcriptomics data [95,143]. Omics data gives a system-level snapshot of an organism's cellular processes under a given condition. The application of genome-scale metabolic models for integrative analysis of omics data helps to have a snapshot of the organism metabolic state [140,167]. The different types of omics data bring a different layer of information on an organism metabolic state and are integrated differently in genome-scale metabolic models (*Figure 1.8*). The genomics data is used to build the genome-scale metabolic network by identifying the metabolic genes encoded in the genome [168]. The transcriptomics data can be used to build a condition-specific metabolic model by constraining the reactions associated with unexpressed genes as null fluxes and ensure the reactions with genes expressed can carry fluxes [94,114]. This is used to constrain and improve metabolic pathways fluxes predictions even under complex genetic regulation [94]. The proteomics data can be used in a similar way to transcriptomics by constraining null fluxes for the reactions with no protein detectable and ensure the reactions can carry fluxes if associated to proteins detected in significant levels [146]. Also proteomics is used to identify metabolic enzymes post-

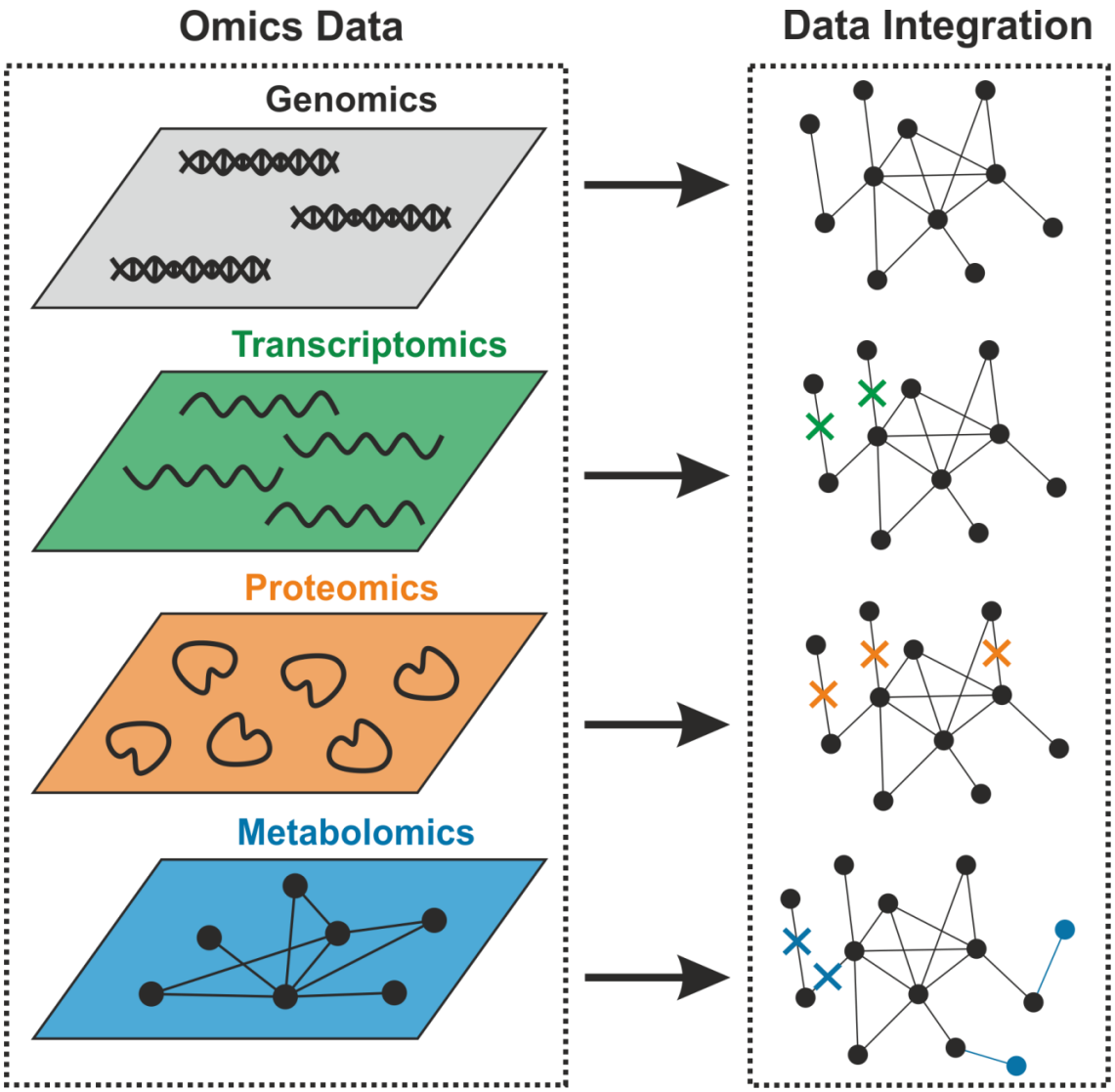


Figure 1.8: Representation of multi-omics data integration in the genome-scale metabolic models

The genomics data is used to generate the metabolic network by identifying the encoded metabolic genes during the metabolic model reconstruction.

The transcriptomics data can be used to identify the expression level of genes associated to reactions, if the genes are not expressed the reactions fluxes are constrained to 0.

The proteomics data can be used to identify if the enzymes associated to a reaction are translated. The reactions with no associated enzyme identified in the proteomics data have their fluxes constrained to 0.

The metabolomics data can be used to identify if a reaction or pathway is active as these produce some specific metabolites. The presence of new metabolites can be used to identify metabolic reactions absent from the model.

translationally regulated; for example, a metabolic gene identified as expressed may not be translated and the reaction associated can be constrained with a null flux boundary [146]. Quantitative proteomics data is also used to identify if the abundance of some enzymes may correlate with the fluxes predictions and identify inaccurate predictions such as overestimation of fluxes [146]. Finally, the omics method directly representing the metabolic state is metabolomics and other derivative methods such as fluxomics. Metabolomics can be used to identify if a pathway is active in an organism by measuring the pathway intermediates, if the pathway is inactive it can be constrained in the model [169]. Untargeted metabolomics can help to identify metabolic pathways that are not included in the metabolic model. Fluxomics directly measures fluxes in an organism by using isotope labelled-metabolites consumed by the organism (e.g., C-labelled glucose, or labelled phosphate) to track the level of metabolites consumed and produced, then calculating the metabolic fluxes, that are integrated as constraints in the metabolic model [135,170]. The integration of C-labelled metabolomics data in a genome-scale metabolic model was applied to study the carbon-flux distribution in *S. coelicolor* strains with (M145) and without (M1146) the main antibiotic gene clusters [171]. This helped to identify key metabolic differences between the two strains, such as a higher glycolytic metabolism in the antibiotic producer, and a higher pentose-phosphate metabolism in the non-producer. Quantitative metabolomics data is used to measure extracellular metabolite levels, the metabolite exchange rates are then calculated and constrained in the model and predict the intracellular metabolic fluxes [94,172–174]. This is particularly interesting for Actinobacteria metabolic modelling that excretes diverse secondary metabolites and waste products, for example, excreted antibiotics can be used to constrain the corresponding

metabolic pathway giving better insight on the production of these compounds. The integration of multi-omics data in genome-scale metabolic models is increasingly attractive as it gives a more holistic picture of the biological system and different omics data can highlight different mechanisms [111].

1.3 Aims of the thesis

With the rise of antimicrobial resistance and the decreasing number of antibiotics discovered [10], it is crucial to better understand and improve production of antibiotics to facilitate production of new compounds during the discovery process [16]. At the same time, increasing the yield of antibiotics makes production more cost-effective [11,16]. In this thesis, I have created constrained-based genome-scale metabolic models of several *Streptomyces* species and Actinobacteria (which are the main source of commercial antibiotics [12,27]), to better understand and improve secondary metabolite production in these organisms. This work was finally used to identify potential metabolic engineering targets to design better *Streptomyces* antibiotics producing strains.

The *Chapter II* of this thesis presents the reconstruction of the genome-scale metabolic model for the model of *Streptomyces* antibiotic producing strain, *S. coelicolor* A3(2). The metabolic model *i*AA1259 [48] is reconstructed on previous versions of *S. coelicolor* GSMMs [44,175,176]. The metabolic model is validated and improved using experimental chemostat and multi-omics data. The metabolic model is benchmarked against the previously published models and shows better predictions of metabolic fluxes. Furthermore, the *i*AA1259 was also improved by introducing multiple databases in the model to facilitate automated integration and analysis of omics data.

This chapter was published in "Development and validation of an updated computational model of *Streptomyces coelicolor* primary and secondary metabolism" in *BMC genomics* in 2018. In *Chapter III*, comparative metabolic modelling was applied to explore and understand, from a systems point of view, the metabolic differences between the two phylogenetically and genetically very close strains *S. coelicolor* and *S. lividans*. For this purpose, a GSMM for *S. lividans* TK24 was generated from the reconstruction of *S. coelicolor* A3(2) (from *Chapter II*) [48]. The metabolic model was used to identify metabolic functions lost and gained by *S. lividans*. The *S. lividans* metabolic model was validated with published experimental chemostat data. The metabolic models of the two strains and their antibiotic production profile was studied in a complex media (R2YE) [177], using published data. A comparative proteomics dataset [178] of the two strains was integrated into the two models to identify metabolic differences, again using published data. The impact of antibiotics production was predicted in the two strains. Finally, from collating the metabolic modelling results, and integrative analysis of published omics and experimental data, potential targets to increase antibiotics production (gene knockouts, gene overexpression, or heterologous enzyme expression) were identified.

Chapter IV covers the comparative reconstruction and analysis of 50 biotechnologically relevant Actinobacteria strains. For this purpose, I built a comparative metabolic model reconstruction pipeline to reconstruct metabolic models of Actinobacteria based on the high-quality *S. coelicolor* metabolic model previously reconstructed and validated in *Chapter II*. This pipeline is entirely based on open-source software parts and scripts in Python and *R*, making it easy and fast to use or modify. Metabolic models were reconstructed and compared for multiple

Actinobacteria strains, including *S. lividans* (from *Chapter III*), to help identify major metabolic differences between these strains. The metabolic distance between the strains was analysed, and the quality of the reconstructed models was verified. The core and accessory metabolism of the Actinobacteria were determined and studied, as well as their active core metabolism. Finally, the metabolic differences between the strains were further investigated by predicting and comparing their growth ability in different media conditions, *in-silico*.

In *Chapter V*, I developed an automated *R* tool to integrate exometabolome fluxes in constraint-based metabolic models. The tool was tested with experimental growth and antibiotics data collected for *S. coelicolor*, then the tool was further tested with a published chemostat dataset [47]. The tools output of this chemostat dataset was used to constrain the metabolic model of *S. coelicolor* (*Chapter II*). Finally, the tool was used to apply an ensemble modelling approach by sampling exchange fluxes within the confidence interval estimated, and these exchange fluxes were constrained in the metabolic model. This enabled to predict an ensemble of hundreds of plausible metabolic states, which were compared and validated with published experimental growth data [47].

Chapter VI uses the metabolic model of *S. coelicolor* (*Chapter II*) and metabolomics data acquired by collaborators to design a better production host. The first part focuses on the production of actinorhodin in *S. coelicolor* under four different conditions, which was studied in the WT strain, the WT with ACT cosmid, M1146, and M1146 with ACT cosmid. Exometabolomic data were automatically integrated using the flux tool (*Chapter V*) in the *S. coelicolor* model (*Chapter I*). The metabolic models constrained under the four different conditions were validated with experimental

data. The metabolic impact of producing actinorhodin was compared in the four models. Potential targets to increase antibiotic production were then identified. The second part focuses on the heterologous production of a RiPP antibiotic (GE2270A) in the *S. coelicolor* M1146 chassis strain for the TOPCAPI project. The production of GE2270A was predicted and studied *in silico* in *S. coelicolor*. The metabolic model was used to identify single and double knockouts that were non-essential to either production or growth under a complex industrial media. Then, a set of potential genes knockouts was identified to force the production of GE2270A in *S. coelicolor* by coupling the production of GE2270A to growth through the production of $FMNH_2$ cofactor.

Finally, *Chapter VII* presents the concluding remarks of this thesis, as well as the future perspectives of genome-scale metabolic modelling to study and design Actinobacteria secondary metabolites producers.

1.4 References

(1) Cameron, D. E.; Bashor, C. J.; Collins, J. J. A Brief History of Synthetic Biology. *Nat. Rev. Microbiol.* **2014**, *12*, 381–390.

(2) Khalil, A. S.; Collins, J. J. Synthetic Biology: Applications Come of Age. *Nat. Rev. Genet.* **2010**, *11* (5), 367–379.

(3) Heinemann, M.; Panke, S. Synthetic Biology--Putting Engineering into Biology. *Bioinformatics* **2006**, *22* (22), 2790–2799.

(4) Mardis, E. R. A Decade's Perspective on DNA Sequencing Technology. *Nature* **2011**, *470* (7333), 198–203.

(5) Bell, G.; Hey, T.; Szalay, A. Beyond the Data Deluge. *Science (80-.).* **2009**, *323* (5919), 1297 LP – 1298.

(6) Krohs, U.; Bedau, M. A. Interdisciplinary Interconnections in Synthetic Biology. *Biol. Theory* **2013**, *8* (4), 313–317.

(7) Andrianantoandro, E.; Basu, S.; Karig, D. K.; Weiss, R. Synthetic Biology: New Engineering Rules for an Emerging Discipline. *Mol. Syst. Biol.* **2006**, *2* (1).

(8) Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. Synthetic Biology to Access and Expand Nature's Chemical Diversity. *Nat. Rev. Microbiol.* **2016**, *14* (3), 135–149.

(9) Medema, M. H.; Blin, K.; Cimermancic, P.; Jager, V. de; Zakrzewski, P.; Fischbach,

M. A.; Weber, T.; Takano, E.; Breitling, R. AntiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. **2011**.

(10) Organization, W. H. *Antimicrobial Resistance: Global Report on Surveillance*; World Health Organization, 2014.

(11) Nathan, C.; Goldberg, F. M. The Profit Problem in Antibiotic R& D. *Nat. Rev. Drug Discov.* **2005**, *4* (11), 887–891.

(12) Watve, M. G.; Tickoo, R.; Jog, M. M.; Bhole, B. D. How Many Antibiotics Are Produced by the Genus *Streptomyces*? *Arch. Microbiol.* **2001**, *176* (5), 386–390.

(13) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Brown, L. C. W.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **2014**, *158* (2), 412–421.

(14) Komatsu, M.; Komatsu, K.; Koiwai, H.; Yamada, Y.; Kozone, I.; Izumikawa, M.; Hashimoto, J.; Takagi, M.; Omura, S.; Shin-ya, K.; et al. Engineered *Streptomyces Avermitilis* Host for Heterologous Expression of Biosynthetic Gene Cluster for Secondary Metabolites. *ACS Synth. Biol.* **2013**, *2* (7), 384–396.

(15) Medema, M. H.; Breitling, R.; Takano, E. Synthetic Biology in *Streptomyces* Bacteria. *Methods Enzym.* **2011**, *497*, 485–502.

(16) Rokem, J. S.; Lantz, A. E.; Nielsen, J. Systems Biology of Antibiotic Production by Microorganisms. *Nat. Prod. Rep.* **2007**, *24* (6), 1262–1287.

(17) Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; van Wezel, G. P. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **2016**, *80* (1), 1 LP – 43.

(18) Verma, M.; Lal, D.; Kaur, J.; Saxena, A.; Kaur, J.; Anand, S.; Lal, R. Phylogenetic Analyses of Phylum Actinobacteria Based on Whole Genome Sequences. *Res. Microbiol.* **2013**, *164* (7), 718–728.

(19) Doroghazi, J. R.; Metcalf, W. W. Comparative Genomics of Actinomycetes with a Focus on Natural Product Biosynthetic Genes. *BMC Genomics* **2013**, *14* (1), 611.

(20) Kim, J.-N.; Kim, Y.; Jeong, Y.; Roe, J.-H.; Kim, B.-G.; Cho, B.-K. Comparative Genomics Reveals the Core and Accessory Genomes of *Streptomyces* Species. *J. Microbiol. Biotechnol.* **2015**, *25* (10), 1599–1605.

(21) Genilloud, O. Physiology of Actinobacteria. In *Biology and Biotechnology of Actinobacteria*; Springer International Publishing: Cham, 2017; pp 151–180.

(22) Alam, M. T.; Medema, M. H.; Takano, E.; Breitling, R. Comparative Genome-scale Metabolic Modeling of *Actinomycetes*: The Topology of Essential Core Metabolism. *FEBS Lett.* **2011**, *585* (14), 2389–2394.

(23) Dyson, P. J. *Streptomyces : Molecular Biology and Biotechnology*; Caister Academic Press, 2011.

(24) Hodgson, D. A. Primary Metabolism and Its Control in Streptomycetes: A Most Unusual Group of Bacteria. *Adv. Microb. Physiol.* **2000**, *42*, 47–238.

(25) Schniete, J. K.; Cruz-Morales, P.; Selem-Mojica, N.; Fernández-Martínez, L. T.; Hunter, I. S.; Barona-Gómez, F.; Hoskisson, P. A. Expanding Primary Metabolism Helps Generate the Metabolic Robustness To Facilitate Antibiotic Biosynthesis in *Streptomyces*; *MBio* **2018**, *9* (1), e02283-17.

(26) Bu'Lock, J. D. Intermediary Metabolism and Antibiotic Synthesis. *Adv. Appl. Microbiol.* **1961**, *3*, 293–342.

(27) Hopwood, D. A. *Streptomyces* in Nature and Medicine : The Antibiotic Makers; Oxford University Press, 2007.
(28) Zhu, H.; Sandiford, S. K.; van Wezel, G. P. Triggers and Cues That Activate Antibiotic Production by *Actinomycetes*. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 371–386.
(29) Craney, A.; Ahmed, S.; Nodwell, J. Towards a New Science of Secondary Metabolism. *J. Antibiot. (Tokyo).* **2013**, *66* (7), 387–400.
(30) van Wezel, G. P.; McDowall, K. J. The Regulation of the Secondary Metabolism of *Streptomyces*: New Links and Experimental Advances. *Nat. Prod. Rep.* **2011**, *28* (7), 1311.
(31) Bibb, M. J. Regulation of Secondary Metabolism in Streptomycetes. *Curr. Opin. Microbiol.* **2005**, *8* (2), 208–215.
(32) White, D.; Drummond, J.; Fuqua, C. *The Physiology and Biochemistry of Prokaryotes*; 2007.
(33) Oliynyk, M.; Samborskyy, M.; Lester, J. B.; Mironenko, T.; Scott, N.; Dickens, S.; Haydock, S. F.; Leadlay, P. F. Complete Genome Sequence of the Erythromycin-Producing Bacterium *Saccharopolyspora Erythraea* NRRL23338. *Nat. Biotechnol.* **2007**, *25* (4), 447–453.
(34) SZABÓ, G.; VÁLYI-NAGY, T.; BARABÁS, G.; BÁSSLER, G. Comaprison of Strains of *Streptomyces Griseus* Which Produce Streptomycin and Those Which Do Not. *Nature* **1960**, *188* (4748), 428–428.
(35) Niewerth, D.; Jansen, G.; Riethoff, L. F. V.; van Meerloo, J.; Kale, A. J.; Moore, B. S.; Assaraf, Y. G.; Anderl, J. L.; Zweegman, S.; Kaspers, G. J. L.; et al. Antileukemic Activity and Mechanism of Drug Resistance to the Marine *Salinispora Tropica* Proteasome Inhibitor Salinosporamide A (Marizomib). *Mol. Pharmacol.* **2014**, *86* (1), 12–19.
(36) Manivasagan, P.; Venkatesan, J.; Sivakumar, K.; Kim, S.-K. Pharmaceutically Active Secondary Metabolites of Marine Actinobacteria. *Microbiol. Res.* **2014**, *169* (4), 262–278.
(37) Vaishnav, P.; Demain, A. L. Unexpected Applications of Secondary Metabolites. *Biotechnol. Adv.* **2011**, *29* (2), 223–229.
(38) Zhang, M. M.; Wang, Y.; Ang, E. L.; Zhao, H. Engineering Microbial Hosts for Production of Bacterial Natural Products. *Nat. Prod. Rep.* **2016**, *33* (8), 963–987.
(39) Zarins-Tutt, J. S.; Barberi, T. T.; Gao, H.; Mearns-Spragg, A.; Zhang, L.; Newman, D. J.; Goss, R. J. M. Prospecting for New Bacterial Metabolites: A Glossary of Approaches for Inducing, Activating and Upregulating the Biosynthesis of Bacterial Cryptic or Silent Natural Products. *Nat. Prod. Rep.* **2016**, *33* (1), 54–72.
(40) Blin, K.; Medema, M. H.; Kazempour, D.; Fischbach, M. A.; Breitling, R.; Takano, E.; Weber, T. AntiSMASH 2.0—a Versatile Platform for Genome Mining of Secondary Metabolite Producers. *Nucleic Acids Res.* **2013**, gkt449.
(41) Fischbach, M. A.; Walsh, C. T.; Clardy, J. The Evolution of Gene Collectives: How Natural Selection Drives Chemical Innovation. *Proc. Natl. Acad. Sci.* **2008**, *105* (12), 4601–4608.
(42) Medema, M. H.; Cimermancic, P.; Sali, A.; Takano, E.; Fischbach, M. A. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. *PLoS Comput. Biol.* **2014**, *10* (12), e1004016.

(43) Cavalier-Smith, T. Origins of Secondary Metabolism. *Ciba Found. Symp.* **1992**, *171*, 64–80; discussion 80-7.
(44) Alam, M. T.; Merlo, M. E.; Hodgson, D. A.; Wellington, E. M. H.; Takano, E.; Breitling, R. Metabolic Modeling and Analysis of the Metabolic Switch in *Streptomyces Coelicolor*. *BMC Genomics* **2010**, *11* (1), 1.
(45) Wentzel, A.; Sletta, H.; Ellingsen, T. E.; Bruheim, P. Intracellular Metabolite Pool Changes in Response to Nutrient Depletion Induced Metabolic Switching in *Streptomyces Coelicolor*. *Metabolites* **2012**, *2* (1), 178–194.
(46) Thomas, L.; Hodgson, D. A.; Wentzel, A.; Nieselt, K.; Ellingsen, T. E.; Moore, J.; Morrissey, E. R.; Legaie, R.; Wohlleben, W.; Rodriguez-Garcia, A.; et al. Metabolic Switches and Adaptations Deduced from the Proteomes of *Streptomyces Coelicolor* Wild Type and PhoP Mutant Grown in Batch Culture. *Mol. Cell. Proteomics* **2012**, *11* (2), M111.013797-M111.013797.
(47) Nieselt, K.; Battke, F.; Herbig, A.; Bruheim, P.; Wentzel, A.; Jakobsen, Ø. M.; Sletta, H.; Alam, M. T.; Merlo, M. E.; Moore, J. The Dynamic Architecture of the Metabolic Switch in *Streptomyces Coelicolor*. *BMC Genomics* **2010**, *11* (1), 10.
(48) Amara, A.; Takano, E.; Breitling, R. Development and Validation of an Updated Computational Model of *Streptomyces Coelicolor* Primary and Secondary Metabolism. *BMC Genomics* **2018**, *19* (1), 519.
(49) Stone, M. J.; Williams, D. H. On the Evolution of Functional Secondary Metabolites (Natural Products). *Mol. Microbiol.* **1992**, *6* (1), 29–34.
(50) Butler, M. J.; Bruheim, P.; Jovetic, S.; Marinelli, F.; Postma, P. W.; Bibb, M. J. Engineering of Primary Carbon Metabolism for Improved Antibiotic Production in *Streptomyces Lividans*. *Appl. Environ. Microbiol.* **2002**, *68* (10), 4731 LP – 4739.
(51) Ryu, Y.-G.; Butler, M. J.; Chater, K. F.; Lee, K. J. Engineering of Primary Carbohydrate Metabolism for Increased Production of Actinorhodin in *Streptomyces Coelicolor*. *Appl. Environ. Microbiol.* **2006**, *72* (11), 7132–7139.
(52) Smirnova, N.; Reynolds, K. A. Engineered Fatty Acid Biosynthesis in Streptomyces by Altered Catalytic Function of Beta-Ketoacyl-Acyl Carrier Protein Synthase III. *J. Bacteriol.* **2001**, *183* (7), 2335–2342.
(53) Rodriguez, E.; Navone, L.; Casati, P.; Gramajo, H. Impact of Malic Enzymes on Antibiotic and Triacylglycerol Production in *Streptomyces Coelicolor*. *Appl. Environ. Microbiol.* **2012**, *78* (13), 4571 LP – 4579.
(54) Borodina, I.; Siebring, J.; Zhang, J.; Smith, C. P.; van Keulen, G.; Dijkhuizen, L.; Nielsen, J. Antibiotic Overproduction in *Streptomyces Coelicolor* A3 (2) Mediated by Phosphofructokinase Deletion. *J. Biol. Chem.* **2008**, *283* (37), 25186–25199.
(55) Wang, J.; Wang, C.; Song, K.; Wen, J. Metabolic Network Model Guided Engineering Ethylmalonyl-CoA Pathway to Improve Ascomycin Production in *Streptomyces Hygroscopicus* Var. Ascomyceticus. *Microb. Cell Fact.* **2017**, *16* (1), 169.
(56) Huang, D.; Li, S.; Xia, M.; Wen, J.; Jia, X. Genome-Scale Metabolic Network Guided Engineering of *Streptomyces Tsukubaensis* for FK506 Production Improvement. *Microb. Cell Fact.* **2013**, *12* (1), 52.
(57) Dang, L.; Liu, J.; Wang, C.; Liu, H.; Wen, J. Enhancement of Rapamycin Production by Metabolic Engineering in *Streptomyces Hygroscopicus* Based on Genome-Scale Metabolic Model. *J. Ind. Microbiol. Biotechnol.* **2017**, *44* (2), 259–

270.
(58) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chemie Int. Ed.* **2009**, *48* (26), 4688–4716.
(59) Oldfield, E.; Dr., P. D.; Lin, F.-Y.; Dr. Terpene Biosynthesis: Modularity Rules. *Angew. Chem. Int. Ed. Engl.* **2012**, *51* (5), 1124.
(60) Nguyen, K. T.; Ritz, D.; Gu, J.-Q.; Alexander, D.; Chu, M.; Miao, V.; Brian, P.; Baltz, R. H. Combinatorial Biosynthesis of Novel Antibiotics Related to Daptomycin. *Proc. Natl. Acad. Sci.* **2006**, *103* (46), 17462 LP – 17467.
(61) Hudson, G. A.; Mitchell, D. A. RiPP Antibiotics: Biosynthesis and Engineering Potential. *Curr. Opin. Microbiol.* **2018**, *45*, 61–69.
(62) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chemie Int. Ed.* **2009**, *48* (26), 4688–4716.
(63) Staunton, J.; Weissman, K. J. Polyketide Biosynthesis: A Millennium Review. *Nat. Prod. Rep.* **2001**, *18* (4), 380–416.
(64) Bystrykh, L. V; Fernández-Moreno, M. A.; Herrema, J. K.; Malpartida, F.; Hopwood, D. A.; Dijkhuizen, L. Production of Actinorhodin-Related" Blue Pigments" by *Streptomyces Coelicolor* A3 (2). *J. Bacteriol.* **1996**, *178* (8), 2238–2244.
(65) Okamoto, S.; Taguchi, T.; Ochi, K.; Ichinose, K. Biosynthesis of Actinorhodin and Related Antibiotics: Discovery of Alternative Routes for Quinone Formation Encoded in the Act Gene Cluster. *Chem. Biol.* **2009**, *16* (2), 226–236.
(66) Mak, S.; Nodwell, J. R. Actinorhodin Is a Redox-Active Antibiotic with a Complex Mode of Action against Gram-Positive Cells. *Mol. Microbiol.* **2017**, *106* (4), 597–613.
(67) Doull, J. L.; Vining, L. C. Nutritional Control of Actinorhodin Production by Streptomyces Coelicolor A3(2): Suppressive Effects of Nitrogen and Phosphate. *Appl Microbiol Biotechnol* **1990**, *32* (4), 449–454.
(68) Melzoch, K.; De Mattos, M. J. T.; Neijssel, O. M. Production of Actinorhodin by *Streptomyces Coelicolor* A3(2) Grown in Chemostat Culture. *Biotechnol. Bioeng.* **1997**, *54* (6), 577–582.
(69) Funa, N.; Funabashi, M.; Yoshimura, E.; Horinouchi, S. A Novel Quinone-Forming Monooxygenase Family Involved in Modification of Aromatic Polyketides. *J. Biol. Chem.* **2005**, *280* (15), 14514–14523.
(70) Silvestre, A. J. D. Terpenes: Major Sources, Properties and Applications. *Monomers, Polym. Compos. from Renew. Resour.* **2008**, 17–38.
(71) Martínez-Núñez, M. A.; López, V. E. L. y. Nonribosomal Peptides Synthetases and Their Applications in Industry. *Sustain. Chem. Process.* **2016**, *4* (1), 13.
(72) Jeong, H.; Sim, Y. M.; Kim, H. J.; Lee, D.-W.; Lim, S.-K.; Lee, S. J. Genome Sequence of the Vancomycin-Producing *Amycolatopsis Orientalis* Subsp. Orientalis Strain KCTC 9412T. *Genome Announc.* **2013**, *1* (3).
(73) Hubbard, B. K.; Walsh, C. T. Vancomycin Assembly: Nature's Way. *Angew. Chemie Int. Ed.* **2003**, *42* (7), 730–765.
(74) Tocchetti, A.; Bordoni, R.; Gallo, G.; Petiti, L.; Corti, G.; Alt, S.; Cruz, J. C. S.; Salzano, A. M.; Scaloni, A.; Puglia, A. M.; et al. A Genomic, Transcriptomic and Proteomic Look at the GE2270 Producer *Planobispora Rosea*, an Uncommon Actinomycete. *PLoS One* **2015**, *10* (7), e0133705.
(75) Fabbretti, A.; He, C.-G.; Gaspari, E.; Maffioli, S.; Brandi, L.; Spurio, R.; Sosio, M.;

Jabes, D.; Donadio, S. A Derivative of the Thiopeptide GE2270A Highly Selective against *Propionibacterium Acnes*. *Antimicrob. Agents Chemother.* **2015**, *59* (8), 4560–4568.

(76) LaMarche, M. J.; Leeds, J. A.; Amaral, A.; Brewer, J. T.; Bushell, S. M.; Deng, G.; Dewhurst, J. M.; Ding, J.; Dzink-Fox, J.; Gamber, G.; et al. Discovery of LFF571: An Investigational Agent for *Clostridium Difficile* Infection. *J. Med. Chem.* **2012**, *55* (5), 2376–2387.

(77) Zabala, D.; Braña, A. F.; Flórez, A. B.; Salas, J. A.; Méndez, C. Engineering Precursor Metabolite Pools for Increasing Production of Antitumor Mithramycins in *Streptomyces Argillaceus*. *Metab. Eng.* **2013**, *20*, 187–197.

(78) Gunnarsson, N.; Eliasson, A.; Nielsen, J. Control of Fluxes towards Antibiotics and the Role of Primary Metabolism in Production of Antibiotics. In *Molecular Biotechnolgy of Fungal beta-Lactam Antibiotics and Related Peptide Synthetases*; Springer, 2004; pp 137–178.

(79) Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; et al. Complete Genome Sequence of the Model Actinomycete *Streptomyces Coelicolor* A3(2). *Nature* **2002**, *417* (6885), 141–147.

(80) Pawlik, K.; Kotowska, M.; Chater, K. F.; Kuczek, K.; Takano, E. A Cryptic Type I Polyketide Synthase (Cpk) Gene Cluster in *Streptomyces Coelicolor* A3(2). *Arch. Microbiol.* **2007**, *187* (2), 87–99.

(81) Stankovic, N.; Senerovic, L.; Ilic-Tomic, T.; Vasiljevic, B.; Nikodinovic-Runic, J. Properties and Applications of Undecylprodigiosin and Other Bacterial Prodigiosins. *Appl. Microbiol. Biotechnol.* **2014**, *98* (9), 3841–3858.

(82) Kieser, T. *Practical Streptomyces Genetics*; John Innes Foundation, 2000.

(83) Cobb, R. E.; Wang, Y.; Zhao, H. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **2014**, *4* (6), 723–728.

(84) Gomez-Escribano, J. P.; Bibb, M. J. Engineering *Streptomyces Coelicolor* for Heterologous Expression of Secondary Metabolite Gene Clusters. *Microb. Biotechnol.* **2011**, *4* (2), 207–215.

(85) Shima, J.; Hesketh, A.; Okamoto, S.; Kawamoto, S.; Ochi, K. Induction of Actinorhodin Production by RpsL (Encoding Ribosomal Protein S12) Mutations That Confer Streptomycin Resistance in *Streptomyces Lividans* and *Streptomyces Coelicolor* A3(2). *J. Bacteriol.* **1996**, *178* (24), 7276–7284.

(86) OKAMOTO-HOSOYA, Y.; SATO, T.-A.; OCHI, K. Resistance to Paromomycin Is Conferred by RpsL Mutations, Accompanied by an Enhanced Antibiotic Production in *Streptomyces Coelicolor* A3(2). *J. Antibiot. (Tokyo).* **2000**, *53* (12), 1424–1427.

(87) Gomez-Escribano, J. P.; Bibb, M. J. Heterologous Expression of Natural Product Biosynthetic Gene Clusters in *Streptomyces Coelicolor*: From Genome Mining to Manipulation of Biosynthetic Pathways. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 425–431.

(88) Zaburannyi, N.; Rabyk, M.; Ostash, B.; Fedorenko, V.; Luzhetskyy, A. Insights into Naturally Minimised *Streptomyces Albus* J1074 Genome. *BMC Genomics* **2014**, *15* (1), 97.

(89) Medema, M. H.; Alam, M. T.; Heijne, W. H. M.; van den Berg, M. A.; Müller, U.;

Trefzer, A.; Bovenberg, R. A. L.; Breitling, R.; Takano, E. Genome-Wide Gene Expression Changes in an Industrial Clavulanic Acid Overproduction Strain of *Streptomyces Clavuligerus*. *Microb. Biotechnol.* **2011**, *4* (2), 300–305.

(90) Arkin, A. Setting the Standard in Synthetic Biology. *Nat. Biotechnol.* **2008**, *26* (7), 771–773.

(91) European Commission. Directorate General for Research. *Synthetic Biology Applying Engineering to Biology.*; Publications Office, 2005.

(92) Breitling, R. What Is Systems Biology? *Front. Physiol.* **2010**, *1*.

(93) Heinemann, M.; Sauer, U. Systems Biology of Microbial Metabolism. *Curr. Opin. Microbiol.* **2010**, *13* (3), 337–343.

(94) Schellenberger, J.; Que, R.; Fleming, R. M. T.; Thiele, I.; Orth, J. D.; Feist, A. M.; Zielinski, D. C.; Bordbar, A.; Lewis, N. E.; Rahmanian, S.; et al. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6* (9), 1290–1307.

(95) Gomez-Cabrero, D.; Abugessaisa, I.; Maier, D.; Teschendorff, A.; Merkenschlager, M.; Gisel, A.; Ballestar, E.; Bongcam-Rudloff, E.; Conesa, A.; Tegnér, J. Data Integration in the Era of Omics: Current and Future Challenges. *BMC Syst. Biol.* **2014**, *8* (Suppl 2), I1.

(96) Kohl, M. Standards, Databases, and Modeling Tools in Systems Biology; Humana Press, 2011; pp 413–427.

(97) Singh, V.; Dhar, P. K. *Systems and Synthetic Biology*; Springer, 2015.

(98) Barrett, C. L.; Kim, T. Y.; Kim, H. U.; Palsson, B. Ø.; Lee, S. Y. Systems Biology as a Foundation for Genome-Scale Synthetic Biology. *Curr. Opin. Biotechnol.* **2006**, *17* (5), 488–492.

(99) Purnick, P. E. M.; Weiss, R. The Second Wave of Synthetic Biology: From Modules to Systems. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (6), 410–422.

(100) Keasling, J. D. Synthetic Biology and the Development of Tools for Metabolic Engineering. *Metab. Eng.* **2012**, *14* (3), 189–195.

(101) Carbonell, P.; Jervis, A. J.; Robinson, C. J.; Yan, C.; Dunstan, M.; Swainston, N.; Vinaixa, M.; Hollywood, K. A.; Currin, A.; Rattray, N. J. W.; et al. An Automated Design-Build-Test-Learn Pipeline for Enhanced Microbial Production of Fine Chemicals. *Commun. Biol.* **2018**, *1* (1), 66.

(102) Opgenorth, P.; Costello, Z.; Okada, T.; Goyal, G.; Chen, Y.; Gin, J.; Benites, V.; de Raad, M.; Northen, T. R.; Deng, K.; et al. Lessons from Two Design–Build–Test–Learn Cycles of Dodecanol Production in *Escherichia Coli* Aided by Machine Learning. *ACS Synth. Biol.* **2019**, *8* (6), 1337–1351.

(103) Chi, H.; Wang, X.; Shao, Y.; Qin, Y.; Deng, Z.; Wang, L.; Chen, S. Engineering and Modification of Microbial Chassis for Systems and Synthetic Biology. *Synth. Syst. Biotechnol.* **2019**, *4* (1), 25–33.

(104) Medema, M. H.; Breitling, R.; Bovenberg, R.; Takano, E. Exploiting Plug-and-Play Synthetic Biology for Drug Discovery and Production in Microorganisms. *Nat. Rev. Microbiol.* **2011**, *9* (2), 131–137.

(105) Carbonell, P.; Parutto, P.; Baudier, C.; Junot, C.; Faulon, J.-L. Retropath: Automated Pipeline for Embedded Metabolic Circuits. *ACS Synth. Biol.* **2014**, *3* (8), 565–577.

(106) Carbonell, P.; Wong, J.; Swainston, N.; Takano, E.; Turner, N. J.; Scrutton, N. S.; Kell, D. B.; Breitling, R.; Faulon, J.-L. Selenzyme: Enzyme Selection Tool for

Pathway Design. *Bioinformatics* **2018**, *34* (12), 2153–2154.
(107) Fong, S. S. Computational Approaches to Metabolic Engineering Utilizing Systems Biology and Synthetic Biology. *Comput. Struct. Biotechnol. J.* **2014**, *11* (18), 28–34.
(108) Rocha, I.; Maia, P.; Evangelista, P.; Vilaça, P.; Soares, S.; Pinto, J. P.; Nielsen, J.; Patil, K. R.; Ferreira, E. C.; Rocha, M. OptFlux: An Open-Source Software Platform for in Silico Metabolic Engineering. *BMC Syst. Biol.* **2010**, *4* (1), 45.
(109) Tao, W.; Yang, A.; Deng, Z.; Sun, Y. CRISPR/Cas9-Based Editing of Streptomyces for Discovery, Characterization and Production of Natural Products. *Front. Microbiol.* **2018**, *9*, 1660.
(110) Lu, W.; Ye, L.; Lv, X.; Xie, W.; Gu, J.; Chen, Z.; Zhu, Y.; Li, A.; Yu, H. Identification and Elimination of Metabolic Bottlenecks in the Quinone Modification Pathway for Enhanced Coenzyme Q10 Production in *Rhodobacter Sphaeroides*. *Metab. Eng.* **2015**, *29*, 208–216.
(111) Fondi, M.; Liò, P. Multi -Omics and Metabolic Modelling Pipelines: Challenges and Tools for Systems Microbiology. *Microbiol. Res.* **2015**, *171*, 52–64.
(112) Phelan, R. M.; Sachs, D.; Petkiewicz, S. J.; Barajas, J. F.; Blake-Hedges, J. M.; Thompson, M. G.; Reider Apel, A.; Rasor, B. J.; Katz, L.; Keasling, J. D. Development of Next Generation Synthetic Biology Tools for Use in *Streptomyces Venezuelae*. *ACS Synth. Biol.* **2017**, *6* (1), 159–166.
(113) Camacho, D. M.; Collins, K. M.; Powers, R. K.; Costello, J. C.; Collins, J. J. Next-Generation Machine Learning for Biological Networks. *Cell* **2018**, *173* (7), 1581–1592.
(114) Kim, M.; Yi, J. S.; Lakshmanan, M.; Lee, D.-Y.; Kim, B.-G. Transcriptomics-Based Strain Optimization Tool for Designing Secondary Metabolite Overproducing Strains of *Streptomyces Coelicolor*. *Biotechnol. Bioeng.* **2016**, *113* (3), 651–660.
(115) Cummings, M.; Breitling, R.; Takano, E. Steps towards the Synthetic Biology of Polyketide Biosynthesis. *FEMS Microbiol. Lett.* **2014**, *351* (2), 116–125.
(116) Aslan, S.; Noor, E.; Bar-Even, A. Holistic Bioengineering: Rewiring Central Metabolism for Enhanced Bioproduction. *Biochem. J.* **2017**, *474* (23), 3935 LP – 3950.
(117) Carbonell, P.; Currin, A.; Jervis, A. J.; Rattray, N. J. W.; Swainston, N.; Yan, C.; Takano, E.; Breitling, R. Bioinformatics for the Synthetic Biology of Natural Products: Integrating across the Design–Build–Test Cycle. *Nat. Prod. Rep.* **2016**, *33* (8), 925–932.
(118) Tong, Y.; Weber, T.; Lee, S. Y. CRISPR/Cas-Based Genome Engineering in Natural Product Discovery. *Nat. Prod. Rep.* **2019**.
(119) Kim, E.; Moore, B. S.; Yoon, Y. J. Reinvigorating Natural Product Combinatorial Biosynthesis with Synthetic Biology. *Nat. Chem. Biol.* **2015**, *11* (9), 649–659.
(120) Breitling, R.; Takano, E. Synthetic Biology of Natural Products. *Cold Spring Harb. Perspect. Biol.* **2016**, *8* (10), 357–369.
(121) Blin, K.; Medema, M. H.; Kazempour, D.; Fischbach, M. A.; Breitling, R.; Takano, E.; Weber, T.; Geer, L.; Geer, R.; Gonzales, N.; et al. AntiSMASH 2.0--a Versatile Platform for Genome Mining of Secondary Metabolite Producers. *Nucleic Acids Res.* **2013**, *41* (W1), W204–W212.
(122) Rutledge, P. J.; Challis, G. L. Discovery of Microbial Natural Products by Activation of Silent Biosynthetic Gene Clusters. *Nat. Rev. Microbiol.* **2015**, *13* (8),

509–523.
(123) Oakley, C. E.; Ahuja, M.; Sun, W.-W.; Entwistle, R.; Akashi, T.; Yaegashi, J.; Guo, C.-J.; Cerqueira, G. C.; Russo Wortman, J.; Wang, C. C. C.; et al. Discovery of McrA, a Master Regulator of Aspergillus Secondary Metabolism. *Mol. Microbiol.* **2017**, *103* (2), 347–365.
(124) Reen, F.; Romano, S.; Dobson, A.; O'Gara, F.; Reen, F. J.; Romano, S.; Dobson, A. D. W.; O'Gara, F. The Sound of Silence: Activating Silent Biosynthetic Gene Clusters in Marine Microorganisms. *Mar. Drugs* **2015**, *13* (8), 4754–4783.
(125) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A. Molecular Networking as a Dereplication Strategy. *J. Nat. Prod.* **2013**, *76* (9), 1686–1699.
(126) Bouslimani, A.; Sanchez, L. M.; Garg, N.; Dorrestein, P. C. Mass Spectrometry of Natural Products: Current, Emerging and Future Technologies. *Nat. Prod. Rep.* **2014**, *31* (6), 718–729.
(127) Gottelt, M.; Kol, S.; Gomez-Escribano, J. P.; Bibb, M.; Takano, E. Deletion of a Regulatory Gene within the Cpk Gene Cluster Reveals Novel Antibacterial Activity in *Streptomyces Coelicolor* A3 (2). *Microbiology* **2010**, *156* (8), 2343–2353.
(128) Shao, Z.; Rao, G.; Li, C.; Abil, Z.; Luo, Y.; Zhao, H. Refactoring the Silent Spectinabilin Gene Cluster Using a Plug-and-Play Scaffold. *ACS Synth. Biol.* **2013**, *2* (11), 662–669.
(129) Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. Direct Cloning and Refactoring of a Silent Lipopeptide Biosynthetic Gene Cluster Yields the Antibiotic Taromycin A. *Proc. Natl. Acad. Sci.* **2014**, *111* (5), 1957–1962.
(130) Menzella, H. G.; Reid, R.; Carney, J. R.; Chandran, S. S.; Reisinger, S. J.; Patel, K. G.; Hopwood, D. A.; Santi, D. V. Combinatorial Polyketide Biosynthesis by de Novo Design and Rearrangement of Modular Polyketide Synthase Genes. *Nat. Biotechnol.* **2005**, *23* (9), 1171–1176.
(131) Huang, D.; Wen, J.; Wang, G.; Yu, G.; Jia, X.; Chen, Y. In Silico Aided Metabolic Engineering of *Streptomyces Roseosporus* for Daptomycin Yield Improvement. *Appl. Microbiol. Biotechnol.* **2012**, *94* (3), 637–649.
(132) Wohlleben, W.; Mast, Y.; Muth, G.; Röttgen, M.; Stegmann, E.; Weber, T. Synthetic Biology of Secondary Metabolite Biosynthesis in Actinomycetes: Engineering Precursor Supply as a Way to Optimize Antibiotic Production. *FEBS Lett.* **2012**, *586* (15), 2171–2176.
(133) Thykaer, J.; Nielsen, J.; Wohlleben, W.; Weber, T.; Gutknecht, M.; Lantz, A. E.; Stegmann, E. Increased Glycopeptide Production after Overexpression of Shikimate Pathway Genes Being Part of the Balhimycin Biosynthetic Gene Cluster. *Metab. Eng.* **2010**, *12* (5), 455–461.
(134) Liao, G.; Wang, L.; Liu, Q.; Guan, F.; Huang, Y.; Hu, C. Manipulation of Kynurenine Pathway for Enhanced Daptomycin Production in *Streptomyces Roseosporus*. *Biotechnol. Prog.* **2013**, *29* (4), 847–852.
(135) Breitling, R.; Achcar, F.; Takano, E. Modeling Challenges in the Synthetic Biology of Secondary Metabolism. *ACS Synth. Biol.* **2013**, *2* (7), 373–378.
(136) Simeonidis, E.; Price, N. D. Genome-Scale Modeling for Metabolic Engineering. *J. Ind. Microbiol. Biotechnol.* **2015**, *42* (3), 327–338.

(137) Palsson, B. *Systems Biology : Constraint-Based Reconstruction and Analysis*.
(138) Varma, A.; Palsson, B. Ø. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology* **1994**, *12* (10), 994–998.
(139) Thiele, I.; Palsson, B. Ø. A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction. *Nat. Protoc.* **2010**, *5* (1), 93–121.
(140) Feist, A. M.; Palsson, B. Ø. The Growing Scope of Applications of Genome-Scale Metabolic Reconstructions Using Escherichia Coli. *Nat. Biotechnol.* **2008**, *26* (6), 659–667.
(141) Esvelt, K. M.; Wang, H. H. Genome-scale Engineering for Systems and Synthetic Biology. *Mol. Syst. Biol.* **2013**, *9* (1), 641.
(142) O'Brien, E. J. J.; Monk, J. M. M.; Palsson, B. O. Ø. Using Genome-Scale Models to Predict Biological Capabilities. *Cell* **2015**, *161* (5), 971–987.
(143) Joyce, A. R.; Palsson, B. Ø. The Model Organism as a System: Integrating "omics" Data Sets. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (3), 198–210.
(144) Kim, W. J.; Kim, H. U.; Lee, S. Y. Current State and Applications of Microbial Genome-Scale Metabolic Models. *Curr. Opin. Syst. Biol.* **2017**, *2*, 10–18.
(145) Kim, B.; Kim, W. J.; Kim, D. I.; Lee, S. Y. Applications of Genome-Scale Metabolic Network Model in Metabolic Engineering. *J. Ind. Microbiol. Biotechnol.* **2015**, *42* (3), 339–348.
(146) Sánchez, B. J.; Zhang, C.; Nilsson, A.; Lahtvee, P.-J.; Kerkhoven, E. J.; Nielsen, J. Improving the Phenotype Predictions of a Yeast Genome-Scale Metabolic Model by Incorporating Enzymatic Constraints. *Mol. Syst. Biol.* **2017**, *13* (8), 935.
(147) King, Z. A.; Lloyd, C. J.; Feist, A. M.; Palsson, B. Ø. Next-Generation Genome-Scale Models for Metabolic Engineering. *Curr. Opin. Biotechnol.* **2015**, *35*, 23–29.
(148) Mesarovic Mihajlo, D.; Sreenath, S. N.; Keene, J. D. Search for Organising Principles: Understanding in Systems Biology. *Syst. Biol. (Stevenage).* **2004**, *1* (1), 19–27.
(149) Gunawardena, J. Biological Systems Theory. *Science (80-.).* **2010**, *328* (5978), 581 LP – 582.
(150) Licona-Cassani, C.; Marcellin, E.; Quek, L.-E.; Jacob, S.; Nielsen, L. K. Reconstruction of the *Saccharopolyspora Erythraea* Genome-Scale Model and Its Use for Enhancing Erythromycin Production. *Antonie Van Leeuwenhoek* **2012**, *102* (3), 493–502.
(151) Khodayari, A.; Maranas, C. D. A Genome-Scale *Escherichia Coli* Kinetic Metabolic Model k-Ecoli457 Satisfying Flux Data for Multiple Mutant Strains. *Nat. Commun.* **2016**, *7*, 13806.
(152) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat. Biotechnol.* **2010**, *28* (3), 245–248.
(153) Mohite, O. S.; Weber, T.; Kim, H. U.; Lee, S. Y. Genome-Scale Metabolic Reconstruction of Actinomycetes for Antibiotics Production. *Biotechnol. J.* **2019**, *14* (1), 1800377.
(154) Thiele, I.; Vlassis, N.; Fleming, R. M. T. FastGapFill: Efficient Gap Filling in Metabolic Networks. *Bioinformatics* **2014**, *30* (17), 2529–2531.
(155) Ebrahim, A.; Lerman, J. A.; Palsson, B. Ø.; Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7* (1), 74.
(156) Aurich, M. K.; Fleming, R. M. T.; Thiele, I. MetaboTools: A Comprehensive Toolbox for Analysis of Genome-Scale Metabolic Models. *Front. Physiol.* **2016**, *7*,

327.
(157) Gelius-Dietrich, G.; Desouki, A.; Fritzemeier, C.; Lercher, M. J. Sybil – Efficient Constraint-Based Modelling in R. *BMC Syst. Biol.* **2013**, *7* (1), 125.
(158) Chan, S. H. J.; Cai, J.; Wang, L.; Simons-Senftle, M. N.; Maranas, C. D. Standardizing Biomass Reactions and Ensuring Complete Mass Balance in Genome-Scale Metabolic Models. *Bioinformatics* **2017**, *33* (22), 3603–3609.
(159) Becker, S. A.; Feist, A. M.; Mo, M. L.; Hannum, G.; Palsson, B. Ø.; Herrgard, M. J. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox. *Nat. Protoc.* **2007**, *2* (3), 727–738.
(160) Feist, A. M.; Palsson, B. O. The Biomass Objective Function. *Curr. Opin. Microbiol.* **2010**, *13* (3), 344–349.
(161) Altafini, C.; Facchetti, G. Metabolic Adaptation Processes That Converge to Optimal Biomass Flux Distributions. *PLoS Comput. Biol.* **2015**, *11* (9), e1004434.
(162) Ow, D. S.-W.; Lee, D.-Y.; Yap, M. G.-S.; Oh, S. K.-W. Identification of Cellular Objective for Elucidating the Physiological State of Plasmid-Bearing *Escherichia Coli* Using Genome-Scale *in Silico* Analysis. *Biotechnol. Prog.* **2009**, *25* (1), 61–67.
(163) Garzoni, C.; Kelley, W. L. Return of the Trojan Horse: Intracellular Phenotype Switching and Immune Evasion by *Staphylococcus Aureus*. *EMBO Mol. Med.* **2011**, *3* (3), 115–117.
(164) Holzhütter, H.-G. The Principle of Flux Minimization and Its Application to Estimate Stationary Fluxes in Metabolic Networks. *Eur. J. Biochem.* **2004**, *271* (14), 2905–2922.
(165) Lewis, N. E.; Hixson, K. K.; Conrad, T. M.; Lerman, J. A.; Charusanti, P.; Polpitiya, A. D.; Adkins, J. N.; Schramm, G.; Purvine, S. O.; Lopez-Ferrer, D.; et al. Omic Data from Evolved *E. Coli* Are Consistent with Computed Optimal Growth from Genome-Scale Models. *Mol. Syst. Biol.* **2010**, *6*, 390.
(166) D'Huys, P.-J.; Lule, I.; Vercammen, D.; Anné, J.; Van Impe, J. F.; Bernaerts, K. Genome-Scale Metabolic Flux Analysis of *Streptomyces Lividans* Growing on a Complex Medium. *J. Biotechnol.* **2012**, *161* (1), 1–13.
(167) Dromms, R.; Styczynski, M. Systematic Applications of Metabolomics in Metabolic Engineering. *Metabolites* **2012**, *2* (4), 1090–1122.
(168) Lewis, N. E.; Nagarajan, H.; Palsson, B. O. Constraining the Metabolic Genotype–Phenotype Relationship Using a Phylogeny of *in Silico* Methods. *Nat. Rev. Microbiol.* **2012**, *10* (4), 291–305.
(169) Kell, D. B. Systems Biology, Metabolic Modelling and Metabolomics in Drug Discovery and Development. *Drug Discov. Today* **2006**, *11* (23), 1085–1092.
(170) Chokkathukalam, A.; Jankevics, A.; Creek, D. J.; Achcar, F.; Barrett, M. P.; Breitling, R. MzMatch–ISO: An R Tool for the Annotation and Relative Quantification of Isotope-Labelled Mass Spectrometry Data. *Bioinformatics* **2013**, *29* (2), 281–283.
(171) Coze, F.; Gilard, F.; Tcherkez, G.; Virolle, M.-J.; Guyonvarch, A. Carbon-Flux Distribution within *Streptomyces Coelicolor* Metabolism: A Comparison between the Actinorhodin-Producing Strain M145 and Its Non-Producing Derivative M1146. *PLoS One* **2013**, *8* (12), e84151.
(172) Cakir, T.; Efe, C.; Dikicioglu, D.; Hortacsu, A.; Kirdar, B.; Oliver, S. G. Flux Balance Analysis of a Genome-Scale Yeast Model Constrained by Exometabolomic Data Allows Metabolic System Identification of Genetically Different Strains.

Biotechnol Prog **2007**, *23* (2), 320–326.
(173) Mo, M. L.; Palsson, B. Ø.; Herrgard, M. J. Connecting Extracellular Metabolomic Measurements to Intracellular Flux States in Yeast. *BMC Syst Biol* **2009**, *3*, 37.
(174) Aurich, M. K.; Paglia, G.; Rolfsson, Ó.; Hrafnsdóttir, S.; Magnúsdóttir, M.; Stefaniak, M. M.; Palsson, B. Ø.; Fleming, R. M. T.; Thiele, I. Prediction of Intracellular Metabolic States from Extracellular Metabolomic Data. *Metabolomics* **2015**, *11* (3), 603–619.
(175) Borodina, I.; Krabben, P.; Nielsen, J. Genome-Scale Analysis of *Streptomyces Coelicolor* A3(2) Metabolism. *Genome Res.* **2005**, *15* (6), 820–829.
(176) Kim, M.; Yi, J. S.; Kim, J.; Kim, J. N.; Kim, M. W.; Kim, B. G. Reconstruction of a High-Quality Metabolic Model Enables the Identification of Gene Overexpression Targets for Enhanced Antibiotic Production in *Streptomyces Coelicolor* A3(2). *Biotechnol. J.* **2014**, *9* (9), 1185–1194.
(177) Esnault, C.; Dulermo, T.; Smirnov, A.; Askora, A.; David, M.; Deniset-Besseau, A.; Holland, I.-B.; Virolle, M.-J. Strong Antibiotic Production Is Correlated with Highly Active Oxidative Metabolism in *Streptomyces Coelicolor* M145. *Sci. Rep.* **2017**, *7* (1), 200.
(178) Millan-Oropeza, A.; Henry, C.; Blein-Nicolas, M.; Aubert-Frambourg, A.; Moussa, F.; Bleton, J.; Virolle, M.-J. Quantitative Proteomics Analysis Confirmed Oxidative Metabolism Predominates in *Streptomyces Coelicolor* versus Glycolytic Metabolism in *Streptomyces Lividans*. *J. Proteome Res.* **2017**, *16* (7), 2597–2613.

Chapter II

2. Development and validation of an updated computational model of *Streptomyces coelicolor* primary and secondary metabolism

Adam Amara, Eriko Takano, and Rainer Breitling*

Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, Faculty of Science and Engineering, University of Manchester, 131Princess Street, Manchester, M1 7DN, United Kingdom.

***Corresponding author**: Rainer Breitling, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom, Tel.: +44161 306 5117, E-mail: rainer.breitling@manchester.ac.uk

2.1 Preface:

"Development and validation of an updated computational model of *Streptomyces coelicolor* primary and secondary metabolism" was published in *BMC Genomics* in 2018. The published manuscript has been formatted in a style consistent with the thesis format, instead of *BMC Genomics* format.

Published manuscript is available here: https://doi.org/10.1186/s12864-018-4905-5

2.2 Abstract

Streptomyces species produce a vast diversity of secondary metabolites of clinical and biotechnological importance, in particular antibiotics. Recent developments in metabolic engineering, synthetic and systems biology have opened new opportunities to exploit *Streptomyces* secondary metabolism, but achieving industry-level production without time-consuming optimization has remained challenging. Genome-scale metabolic modelling has been shown to be a powerful tool to guide metabolic

engineering strategies for accelerated strain optimization, and several generations of models of *Streptomyces* metabolism have been developed for this purpose.

Here, we present the most recent update of a genome-scale stoichiometric constraint-based model of the metabolism of *Streptomyces coelicolor*, the major model organism for the production of antibiotics in the genus. We show that the updated model enables better metabolic flux and biomass predictions and facilitates the integrative analysis of multi-omics data such as transcriptomics, proteomics and metabolomics.

The updated model presented here provides an enhanced basis for the next generation of metabolic engineering attempts in *Streptomyces*.

2.3 Introduction

Streptomyces species are usually soil-dwelling bacteria, which have adapted to their competitive ecological niches by developing a notably diverse secondary metabolism (e.g., antimicrobials). Currently, more than two thirds of the antibiotics used have been derived from natural products discovered in *Streptomyces* and related species [1]; however, the antibiotic discovery pipeline is drying up, while the antimicrobial resistance threat is growing. *Streptomyces coelicolor* A3(2) is a well-studied model organism for the production of antibiotics in this genus. The genome of this soil-dwelling bacterium encodes more than twenty secondary metabolite biosynthetic gene clusters (BGCs) [2], and the species is known to produce multiple antibiotics such as Actinorhodin (Act), Undecylprodigiosin (Red), Calcium-Dependant Antibiotic (CDA) and a yellow Coelicolor Polyketide, Coelimycin P1 (CPK) [3]. Recent developments in metabolic engineering, synthetic and systems biology have opened new opportunities

to exploit *Streptomyces*' secondary metabolism diversity to discover novel antibiotics and natural product-derived drugs [4,5]. However, expensive and time-consuming strain optimization is usually required to achieve industrially competitive production levels. A major issue faced in strain design is the ability to integrate test data (e.g. metabolomics) to improve the design [6], and many of the issues encountered are related to metabolic optimization, such as metabolic bottlenecks to increase production [7], heterologous biosynthetic pathway precursors production [8], or accurate predictions for metabolic engineering [9].

Genome-scale metabolic models (GSMM) have been shown to be a powerful tool to guide metabolic engineering strategies for accelerated strain optimization [10–12], and several generations of models of *Streptomyces* metabolism have been developed for this purpose [13–17]. The use of constraint-based modelling, in particular with flux balance analysis (FBA), enables the reconstruction and analysis of large metabolic networks from the genome sequence as well as predictions of growth associated phenotypes (metabolic fluxes, growth rates, metabolic gene essentiality) [18]. Informative models for this purpose can be constructed even when enzyme kinetic data or metabolite concentrations are unknown in the target organism, making this approach particularly attractive for less well-studied organisms like *Streptomyces* strains. In 2005, the first generation GSMM of *S. coelicolor*, *i*IB711, was published [19], which was used to identify metabolic gene knock-outs to drive the enhanced production of antibiotics in the strain [20]. In 2010, an updated model, *i*MA789, was published [21], which introduced more detailed antibiotics metabolic pathways and was used to interpret time-course gene expression data, which was then used to improve the model and update the genome annotation of the organism in the area of

secondary metabolism. The most recent model update, *i*MK1208, was published by Kim *et al.* (2014); this model significantly expanded the number of reactions and genes, as well as updating the biomass reaction. This model was then used in a transcriptomics-based optimization for actinorhodin overproduction in *S. coelicolor* [15].

Furthermore, several genome-scale metabolic models for other biotechnologically relevant *Streptomyces* strains have been reconstructed since the first *S. coelicolor* model, *i*IB711. A model of the *Streptomyces tenebrarius* metabolic network, which was derived from the *i*IB711 model of Borodina, Krabben & Nielsen has been used to identify targets to optimize production of tobramycin [22]. A model of *Saccharopolyspora erythraea* has been reconstructed based on the *i*MA789 model of Alam *et al.* to improve the production of erythromycin [23]. One of the most recent model reconstructions derived from Kim *et al.*'s *S. coelicolor i*MK1208 model was used for model-guided engineering of ethylmalonyl-CoA pathways in *Streptomyces hygroscopicus* to increase production of ascomycin [14]. A large collection of minimally curated metabolic models of different *Streptomyces* strains and other actinomycetes was used to evaluate potential host strains for overproducing different chemical classes of secondary metabolites using comparative multi-objective modelling [24].

Based on recent advances in our understanding of *Streptomyces* metabolism and technical progress in the concepts of computational model building, we constructed and validated an updated GSMM of *S. coelicolor, i*AA1259, to provide a more precise metabolic flux and biomass predictions and to facilitate the integration of metabolomics, proteomics, and transcriptomics information with the model predictions.

2.4 Methods & Materials

2.4.1 Metabolic model reconstruction

The model reconstruction was initiated by updating the *i*MK1208 *S. coelicolor* model. The standard protocol for reconstruction of high-quality constraint-based GSMMs was followed when adding new genes, reactions, and metabolites [25].

In summary, the initial stoichiometric matrix was generated by comparing and using the iMK1208 model [26], KEGG [27], ScoCyc [28], and two automated reconstructions using RAST annotations and SEED reconstructions [29,30]. The resulting matrix was manually curated for specific pathways (e.g., secondary metabolites biosynthesis, oxidative phosphorylation), to add or correct missing reactions, metabolites, genes associated, or reversibility constraints; this was supported by extensive literature survey to identify new knowledge or gaps in the previous model. Comparative analysis of transcriptomics data with *i*MK1208 helped to identify gene mis-annotations to be corrected [21,31]. The biomass reaction was updated, as multiple reactions impacting biomass have been added (e.g., demethylmenaquinone, cytochrome oxidases or NADH dehydrogenase reactions). This was followed by a recalculation of the ATP fluxes for growth-associated and non-growth-associated maintenance using chemostat data [32], following the Varma & Palsson protocol [33]; the resulting values were very similar to those used in *i*MK1208 (with a GAM of 75.7 ATP in *i*MK1208 versus 75.79 ATP in *i*AA1259, and an NGAM of 2.65 in *i*MK1208 versus 2.64 in *i*AA1259). The detailed modifications on the biomass are available in the *Electronic Supplementary 2.1 - Additional file 3*.

Finally, multiple database identifiers were added either by automatic matching or by manual curation when necessary. The metabolites were annotated with multiple database identifiers; BiGG and KEGG identifications were already present in *i*MK1208, and other databases relevant to metabolomics data analysis were added: ChEBI, HMDB, CAS, IUPAC, ChemSpider, Metlin and PubChem identifications, wherever available. Chemical structure-related annotations (SMILES or InChi) were also introduced for all metabolites. Furthermore, when available, all reactions were annotated with EC code and CAS registry number, in addition to the BiGG annotation used in *i*MK1208. Additional gene annotations have been included to facilitate transcriptomics data integration with identifiers for Gene Ontology (GO), RefSeq, EMBL-ENA and Ensembl. For integrated proteomics analysis, annotations have been expanded to include identifiers for UniProt, Pfam and Panther, as well as data on protein length, mass, and amino acid sequence to support the direct mapping of mass-spectrometry-based proteomics data in the future. The expansion of these annotations also aims at helping fast reconstruction of metabolic models for other *Streptomyces* strains using comparative reconstruction and modelling methods. The final model has been named *i*AA1259 and is compliant with current metabolic model standards [34,35]. The final model is available in SBML format and Excel format in *Electronic Supplementary 2.1 - Additional files 1 and 2*.

2.4.2 Constraint-based modelling

The model was analysed by using Flux Balance Analysis (FBA) and parsimonious FBA (pFBA) to predict optimal *in silico* growth and metabolic flux using the COBRA toolbox in Matlab and Python [36,37] and further evaluated using OptFlux and Sybil [38,39]. To apply

condition-specific constraints corresponding to the media composition, the uptake fluxes for exometabolites not available in the medium were set to zero, while all metabolic by-products were always allowed to leave the metabolic system. The measured nutrient uptake rates from the fermenter datasets are used to define constraints of the nutrient uptake for the model. The objective function maximized in the modelling was the growth rate (steady-state flux towards biomass).

Despite the fact that FBA is not a dynamic modelling approach (its basic assumption being a steady-state flux distribution), using dynamic constraints on CO_2, O_2, glucose, phosphate, and glutamate uptake based on fermenter time-course data [31] enabled simulation of the growth and metabolic dynamics across time. In order to simulate the production of the main antibiotics, the biomass composition was varied dynamically depending on the observed concentration of γ-Act and Red secondary metabolites in the cultures [21].

2.4.3 Transcriptomics and proteomics data analysis

Multiple omics data types have been used to validate the model; the proteomics data [40] have been acquired from the same time-series experiment samples as the flux constraints data and the transcriptomics data [31]. The transcriptomics and proteomics data were matched to corresponding metabolic genes associated with reactions by matching the StrepDB gene annotations. The matching procedure was similar to the one used for iMA789 [21]. Gene expression levels and predicted fluxes were compared using Pearson and Spearman correlations. The data used are available in the *Supplementary 2.1 - Additional file 4*.

2.5 Results & Discussion

2.5.1 Genome-scale reconstruction and characteristics updated

The construction of the updated GSMM of *S. coelicolor* A3(2), iAA1259, was based on all three previously published iterative reconstructions of *S. coelicolor* metabolic models [19,21,26], by updating and adding data in the model based on new genetic (e.g., gene–protein–reaction relationships) or biochemical knowledge. A summary of the main updates and new features added is available in the *Supplementary Tables 1, 2, 3* and *4*.

Multiple pathways were added or updated. 1) Polysaccharide degradation pathways (e.g., for xylan, cellulose) were introduced to enable simulated growth in complex media containing these carbon sources. 2) The biosynthetic pathway for the secondary metabolite CPK [41–43] was added to the model. This cryptic BGC is awakened under phosphate-limited condition, in nitrogen and carbon rich media [41,44], such as in the minimal media used for systems biology studies of *S. coelicolor* [45]. 3) The biosynthetic pathways for the signalling molecules gamma-butyrolactones (SCB1, 2 and 3) were added [46,47]; secondary metabolite production in *Streptomyces* (e.g., CPK) can be activated through these small diffusible molecules, and they are an interesting target for synthetic biology engineering [48,49]. 4) The futalosine pathway, an alternative menaquinone biosynthesis pathway, which was highlighted as incomplete in the previous model [26], has now been updated following recently published studies [50,51]. 5) The oxidative phosphorylation associated reactions have been manually curated. 6) Following the above modifications, the biomass reaction has also been updated to reflect more detailed knowledge on biomass composition such as the presence of 2-

demethylmenaquinol in *S. coelicolor* (MetaCyc) [28], and organic polyphosphate storage [52], as well as an update in the stoichiometry of menaquinol based on *Mycobacterium tuberculosis* data [53] (see details in *Electronic Supplementary 2.1 - Additional file 3*).

In order to facilitate metabolomics data analysis, all metabolites in the model have now been annotated with standard identifiers for a variety of relevant databases (PubChem and ChEBI) [54,55], and chemical and structural information about each metabolite has been added (InChi and SMILES strings) to ensure unambiguous metabolite identification [56,57]. The model capacity to facilitate metabolomics data analysis has been tested by mapping metabolites annotated with mzMatch [58] from an untargeted metabolomics dataset of *S. coelicolor* [59]; the metabolites were mapped automatically onto the *i*AA1259 metabolic network (see details in *Supplementary Figure 2.1*). In addition, to facilitate transcriptomics data analysis and comparative modelling, gene annotation has been expanded to include identifiers for multiple standard databases (Gene Ontology, Ensembl, and RefSeq) [60–62]. Finally, to integrate proteomics data analysis, standard database identifiers (UniProt, Pfam, and Panther) [63–65] and key reference data, such as protein sequence, length, and mass, have been added. The final model, *i*AA1259, is fully compliant with the current standards for high-quality GSMMs [25,34,35], *i*AA1259 is available as a SBML file in *Electronic Supplementary 2.1 Additional File 1* and as an excel file in *Additional File 2*.

2.5.2 Validations of the metabolic model predictions

As the first step in model validation, chemostat data collected by Melzoch *et al. for S. coelicolor* in a glucose-limited minimal defined media [32] were used to compare biomass predictions by the four generations of model: *i*IB711, *i*MA789, iMK1208, and

*i*AA1259. Specific growth rates for each model were predicted *in silico*, using the known glucose and O_2 uptake rates as constraints on the model, along with the production rates of CO_2 and γ-actinorhodin, the extracellular lactone form of actinorhodin [66]. Biomass production was maximized to estimate the optimal predicted growth rate. Then, the growth rate predicted *in silico* was compared to the dilution rate that corresponds to the observed growth rate at steady state (*Figure 2.1*). Since the first published model *i*IB711, there have been some significant improvements in biomass predictions; *i*AA1259 shows a slight improvement in predictions compared to the previous model update, *i*MK1208 (8.2% average error for *i*MK1208 predictions versus 7.0% with *i*AA1259). This first validation confirms that the predictive performances of the updated model *i*AA1259 are at least as good as the previous models generations. However, the next validation step requires more complex and quantitative datasets. The data used as constraints and the predicted growth rates data for the different models are available in *Supplementary Table 7.*

A more substantial improvement in prediction quality is observed when comparing the dynamic growth predictions of the metabolic models *i*AA1259, *i*MK1208, and *i*MA789, to published experimental growth data (*Figure 2.2*) [31]. The dynamic growth was predicted by applying dynamic constraints from fermenter data (see *Methods* for details). The comparison of the predicted and experimental dynamic cell growth shows a significant improvement in quantitative and qualitative biomass prediction using the updated model *i*AA1259 (moving from an average absolute error of 37.6% with *i*MK1208 predictions to 5.3% with *i*AA1259; *Figure 2.2*, and *Supplementary Figure 2.2*). This improvement in biomass predictions is most likely due to the update of the biomass reaction and in the oxidative phosphorylation-related

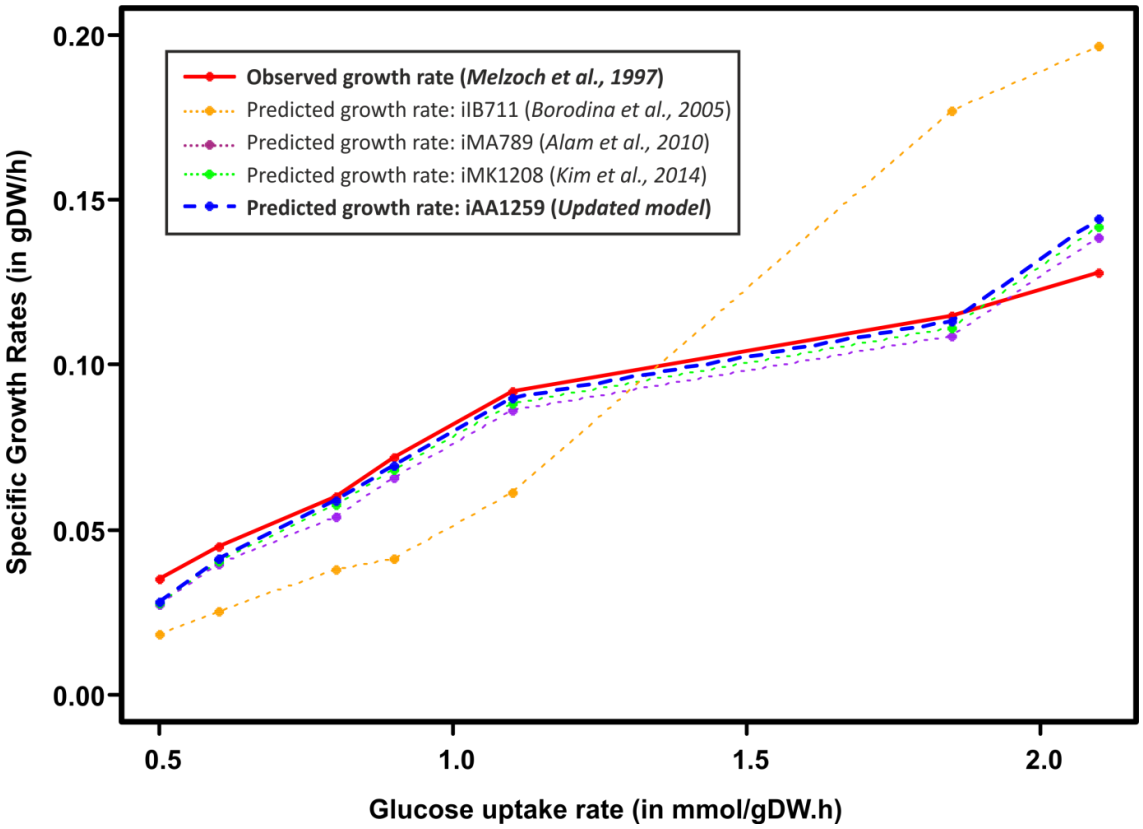


Figure 2.1: Initial model growth predictions validation.

Comparison of the specific growth rate predicted *in-silico* with different models to the measured growth rates in chemostat data [32] with a glucose-limited minimal defined media. The published data on the rate of glucose uptake, oxygen consumption, CO_2 production and γ-actinorhodin production for seven different conditions were used as metabolic constraints in the different models. Growth prediction by iAA1259 shows a slight improvement compared to its immediate predecessor, iMK1208 [26].

reactions updates (i.e., cytochrome oxidases and/or menaquinone pathway), as these are the main adjustments affecting biomass-related reactions directly.

The next validation step involved individual metabolic flux predictions across the model. For this purpose, the models were constrained with time series fermenter data for glucose and O_2 uptake rates and the production rates of γ-actinorhodin and CO_2 for 32 time points from Nieselt *et al.* using the method introduced by Alam *et al.* [21]. The metabolic flux predictions were compared to time series of gene-and protein expression reported by Lahtvee and colleagues (2017), as proxies for the relative metabolic flux across the time course [67]. For the majority of genes, the gene expression

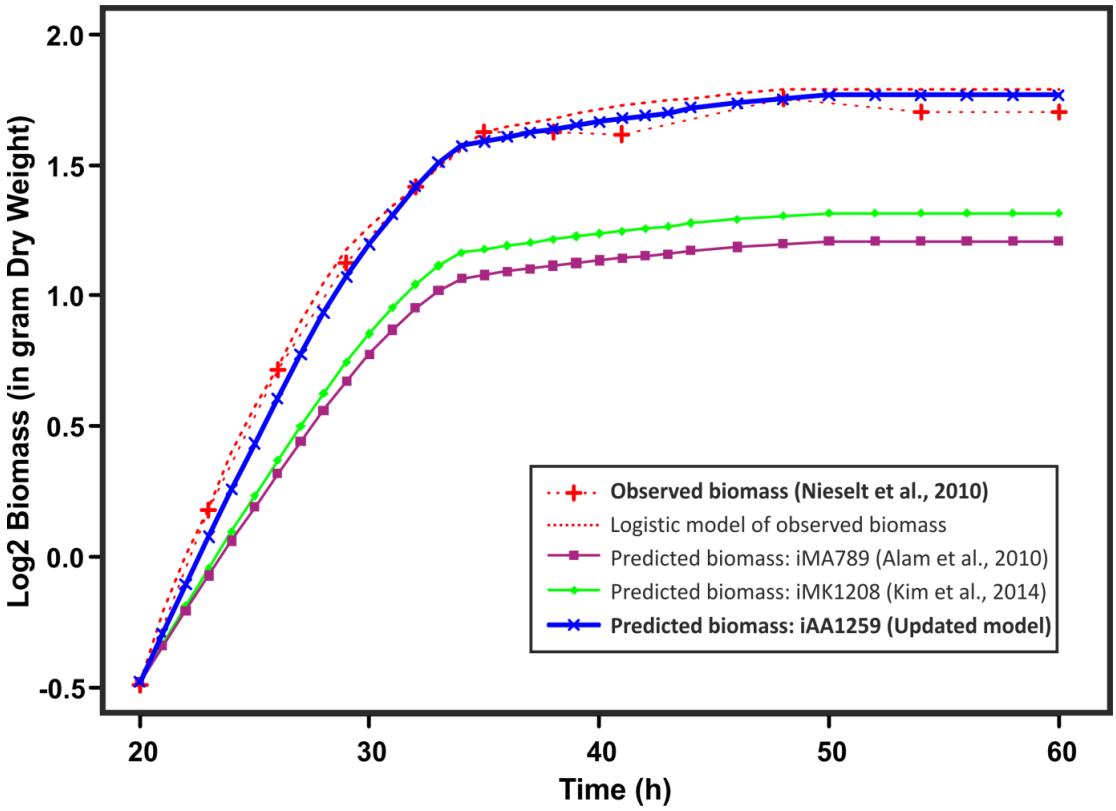


Figure 2.2: Comparison of dynamic cell growth predictions.
The quantitative *in silico* growth predictions are compared to measured biomass and predictions with previously published models across time points. The models were constrained using phosphate, glucose, and glutamate consumption, as well as production of the antibiotics actinorhodin and undecylprodigiosin measured in a fermenter experiment [31]. The updated model's predictions are closer to experimental observations than those of previous models, showing a significant improvement in growth prediction with the *i*AA1259 model.

changes over time are strongly correlated to the predicted metabolic fluxes through the associated reactions (*Figure 2.3*), and the correlation is substantially improved in the updated model presented here (median Spearman correlation coefficient 0.56, compared to 0.18 in the most recent predecessor, *i*MK1208). When focusing only on the correlation for genes that change at least 25% in expression across the time course (*Figure 2.3d*), i.e. those genes that should show correlation, the quality of the correlation is even more pronounced (the Pearson correlation coefficient increases by 39% from 0.56 to 0.78), and it becomes clear that only a very small number of genes

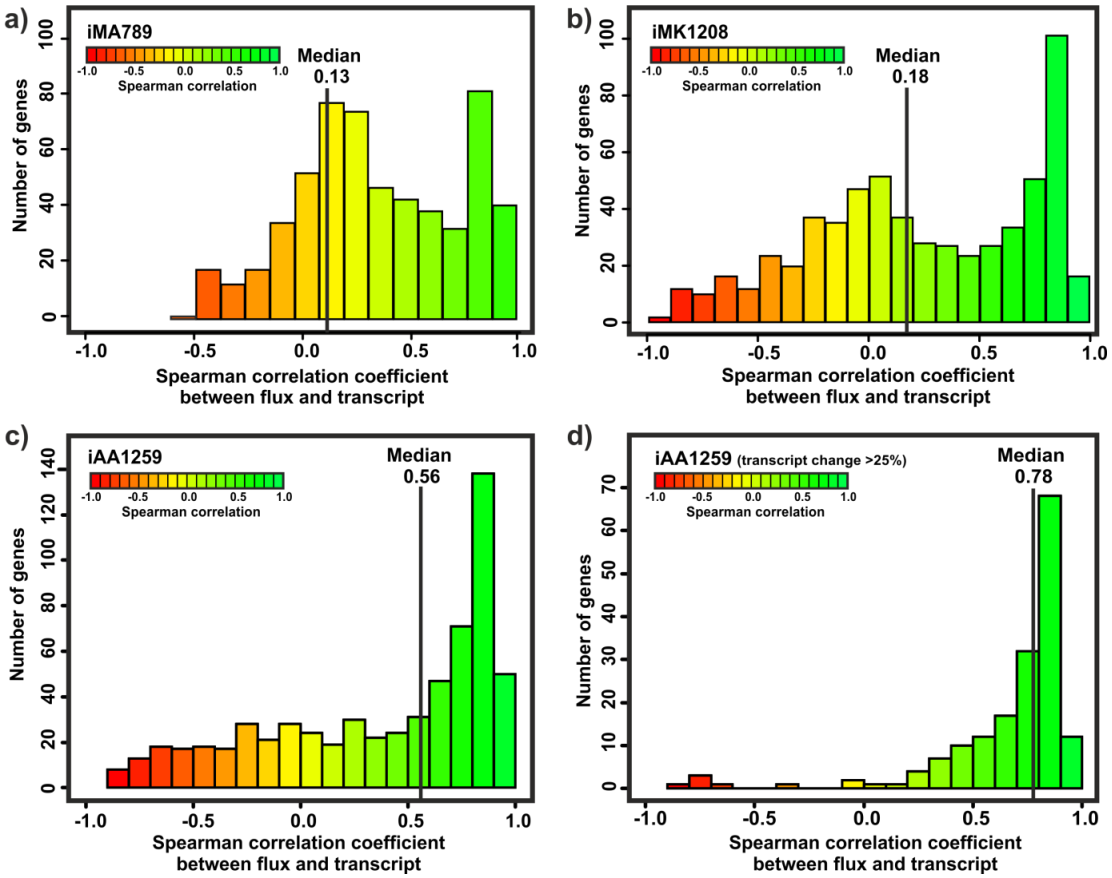


Figure 2.3: Correlation analysis between gene expression and predicted fluxes for the different models.

The histograms show correlations between gene expression and flux predicted for the metabolic genes present in the different published GSMMs of *S. coelicolor*. This approach has been used first by Alam et al., 2010, for the model iMA789.

a) Histogram of correlations for the model iMA789 by Alam et al. [21].

b) Histogram of correlations for the model iMK1208 by Kim et al. [26].

c) Histogram of correlations for the new model, iAA1259. The histogram shows a strong correlation between gene expression and predicted fluxes for metabolic genes present in the model iAA1259. Overall correlation is substantially higher than for the previous models, with a median Spearman correlation of 0.56 compared to 0.13 for iMA789 and 0.18 for iMK1208.

d) Histogram of correlations for the model iAA1259, but only taking into account genes with expression variation of more than 25% between the minimal and maximal transcript level.

show anti-correlated behaviour, i.e. a strong disagreement between gene expression and predicted fluxes. A similar trend is observed when applied to the fluxes predicted with *i*MA789 and *i*MK1208 models, both showing an increase of overall Pearson

correlation from 0.13 to 0.38, and from 0.18 to 0.56, respectively (*Supplementary Figure 2.1*). The trend of a progressive increase in predictive power is still observed from *i*MA789 to *i*MK1208 (47% increase in correlation), and from *i*MK1208 to *i*AA1259 (42% increase in correlation).

In the updated model, there are two major genetic features showing anti-correlation (*Figure 2.4, row A*), the genes associated to Calcium Dependant Antibiotics (CDA) biosynthesis, and the *nuo* operon genes associated to an NADH dehydrogenase (complex I). The CDA genes (*Figure 2.4, row E*) are anti-correlated because their gene expression unexpectedly increases during the transition phase (*Figure 2.4, row C*), whereas the model does not produce CDA. The metabolite production was not switched on in the model, as there is no calcium in the media conditions used by Nieselt *et al.*, 2010. It has been shown previously that CDA could not be detected at significant levels if there was no calcium in the media [68]. Furthermore, production of the associated proteins is not confirmed by the proteomics data [40] (*Figure 2.4, row B*). Thus, in this case, the model prediction (no flux increase during the transition phase, *Figure 2.4 row D*) appears to be correct, and gene expression in this exceptional case might not be correlating with metabolic flux. The Figure 4 is available in high-definition in the *Additional File 5* as an *Electronic Supplementary 2.1*.

Regarding the second major anti-correlation, which is seen for the genes encoding the fourteen subunits of NADH dehydrogenase I (NDH-I), the nuo complex, the disagreement between gene expression and predicted flux is due to a regulatory phenomenon, which in general is difficult to capture in a constraint-based model: two isoenzyme complexes are present in *S. coelicolor*, the relative expression of which is controlled by a regulatory loop dependent on NADH/NAD^+ ratio [69]. Of these, the *nuo*

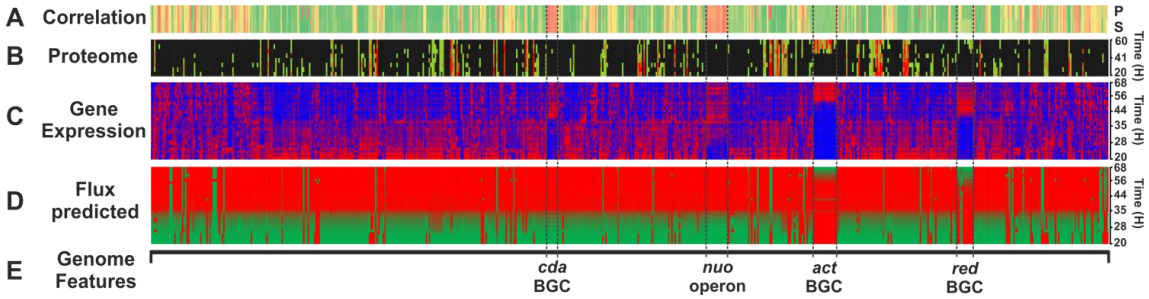


Figure 2.4: Validation by integrated transcriptomics and proteomics analysis.

Gene expression and proteomics data were mapped to metabolic enzyme-coding genes and the associated metabolic fluxes predicted over time. Overall, the predicted flux trends are strongly correlated (green colour in the top bar) to the observed gene expression trend across the metabolic switch event (around between 35 and 36 hours). Two highly anti-correlated gene clusters are highlighted (red colour in the top bar).

Correlation: Pearson (P) and Spearman (S) correlation coefficient between the experimental gene expression level and the predicted fluxes through the corresponding reaction for each individual metabolic gene (green: good correlation; yellow: no correlation; red: anti-correlation.

Proteome: Protein abundance observed in experimental time course data: red: high: green: low abundance, black: missing data (only a small subset of enzymes was quantified). Proteomics data from Thomas, et al. [40].

Gene expression: Gene expression levels observed in the same experimental time course (red: high, blue: low expression). A much larger number of time course were studied than in the proteomics analysis. Gene expression data from Nieselt et al. [31].

Predicted flux: Flux predicted during a simulated time course (green: high; red: low predicted flux).

Genome features: Selected genomic regions discussed in the text are annotated. The data is ordered based on the position of analysed genes in the reference genome (from left to right, from 161,237 bp to 8,468,158 bp). Genome sequence from Bentley et al. [2].

genes are preferentially expressed during fumarate respiration (stationary phase [70]). So, while fluxes through the reaction catalysed by the NADH-dehydrogenase are reduced after the transition phase, *nuo* gene expression increases and results in an anti-correlation of *nuo* gene expression with the flux prediction (which does not distinguish between the isoenzyme complexes). The second NADH dehydrogenase, NDH-II, encoded by three copies of the *ndh* gene, is preferentially expressed during exponential phase and switched off after the transition phase; hence, the *ndh* genes

show high correlation with the predicted flux. While regulatory phenomena like this are not considered in this type of model, the misprediction highlights an interesting phenomenon for future study, i.e. the impact of the relative role of the two sources of reducing co-factors on secondary metabolism in *S. coelicolor*.

2.6 Conclusions

Here, we have presented an updated computational model of *S. coelicolor* primary and secondary metabolism, *i*AA1259; this model shows improved predictive abilities compared to previous model generations for metabolic changes at different scales, from overall biomass dynamics to fluxes through individual reactions.

Another important improvement is that the model has been also updated to enable integrative multi-omics data analysis, to be used for designing and debugging of engineered *Streptomyces* strains using a synthetic biology approach [6], and is now fully compliant with current modelling standards [34,35].

The model presented here will be a good basis for the next round of computer-aided design of metabolically enhanced *Streptomyces* strains. The principled construction of the model using standard identifiers will facilitate the transfer of information to related strains beyond *S. coelicolor* (e.g., recently emerging popular biotechnological hosts, such as *Streptomyces albus* and *Streptomyces venezuelae* [71,72]). It will also serve as a solid starting point for the next generation of updated metabolic models, which will address the challenge of including kinetic and regulatory constraints, in a similar way as the recently published genome-scale metabolic models for the well-studied microorganisms *Escherichia coli* [73] and *Saccharomyces cerevisiae* [74].

2.7 References

(1) Hopwood, D. A. *Streptomyces* in Nature and Medicine : The Antibiotic Makers; Oxford University Press, 2007.

(2) Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; et al. Complete Genome Sequence of the Model Actinomycete *Streptomyces Coelicolor* A3(2). *Nature* **2002**, *417* (6885), 141–147.

(3) van Keulen, G.; Dyson, P. J. Production of Specialized Metabolites by Streptomyces Coelicolor A3 (2). *Adv Appl Microbiol* **2014**, *89*, 217–266.

(4) Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. Synthetic Biology to Access and Expand Nature's Chemical Diversity. *Nat. Rev. Microbiol.* **2016**, *14* (3), 135–149.

(5) Medema, M. H.; Breitling, R.; Bovenberg, R.; Takano, E. Exploiting Plug-and-Play Synthetic Biology for Drug Discovery and Production in Microorganisms. *Nat. Rev. Microbiol.* **2011**, *9* (2), 131–137.

(6) Carbonell, P.; Currin, A.; Jervis, A. J.; Rattray, N. J. W.; Swainston, N.; Yan, C.; Takano, E.; Breitling, R. Bioinformatics for the Synthetic Biology of Natural Products: Integrating across the Design–Build–Test Cycle. *Nat. Prod. Rep.* **2016**, *33* (8), 925–932.

(7) Lu, W.; Ye, L.; Lv, X.; Xie, W.; Gu, J.; Chen, Z.; Zhu, Y.; Li, A.; Yu, H. Identification and Elimination of Metabolic Bottlenecks in the Quinone Modification Pathway for Enhanced Coenzyme Q10 Production in *Rhodobacter Sphaeroides*. *Metab. Eng.* **2015**, *29*, 208–216.

(8) Gomez-Escribano, J. P.; Bibb, M. J. Heterologous Expression of Natural Product Biosynthetic Gene Clusters in *Streptomyces Coelicolor*: From Genome Mining to Manipulation of Biosynthetic Pathways. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 425–431.

(9) Fong, S. S. Computational Approaches to Metabolic Engineering Utilizing Systems Biology and Synthetic Biology. *Comput. Struct. Biotechnol. J.* **2014**, *11* (18), 28–34.

(10) Simeonidis, E.; Price, N. D. Genome-Scale Modeling for Metabolic Engineering. *J. Ind. Microbiol. Biotechnol.* **2015**, *42* (3), 327–338.

(11) Kim, B.; Kim, W. J.; Kim, D. I.; Lee, S. Y. Applications of Genome-Scale Metabolic Network Model in Metabolic Engineering. *J. Ind. Microbiol. Biotechnol.* **2015**, *42* (3), 339–348.

(12) Xu, C.; Liu, L.; Zhang, Z.; Jin, D.; Qiu, J.; Chen, M. Genome-Scale Metabolic Model in Guiding Metabolic Engineering of Microbial Improvement. *Appl. Microbiol. Biotechnol.* **2013**, *97* (2), 519–539.

(13) Huang, D.; Li, S.; Xia, M.; Wen, J.; Jia, X. Genome-Scale Metabolic Network Guided Engineering of *Streptomyces Tsukubaensis* for FK506 Production Improvement. *Microb. Cell Fact.* **2013**, *12* (1), 52.

(14) Wang, J.; Wang, C.; Song, K.; Wen, J. Metabolic Network Model Guided Engineering Ethylmalonyl-CoA Pathway to Improve Ascomycin Production in *Streptomyces Hygroscopicus* Var. Ascomyceticus. *Microb. Cell Fact.* **2017**, *16* (1), 169.

(15) Kim, M.; Yi, J. S.; Lakshmanan, M.; Lee, D.-Y.; Kim, B.-G. Transcriptomics-Based Strain Optimization Tool for Designing Secondary Metabolite Overproducing

Strains of *Streptomyces Coelicolor*. *Biotechnol. Bioeng.* **2016**, *113* (3), 651–660.
(16) Toro, L.; Pinilla, L.; Avignone‐rossa, C.; Ríos‐estepa, R.; Avignone-Rossa, C.; Ríos-Estepa, R. An Enhanced Genome-Scale Metabolic Reconstruction of *Streptomyces Clavuligerus* Identifies Novel Strain Improvement Strategies. *Bioprocess Biosyst. Eng.* No. 5.
(17) Fondi, M.; Pinatel, E.; Talà, A.; Damiano, F.; Consolandi, C.; Mattorre, B.; Fico, D.; Testini, M.; De Benedetto, G. E.; Siculella, L.; et al. Time-Resolved Transcriptomics and Constraint-Based Modeling Identify System-Level Metabolic Features and Overexpression Targets to Increase Spiramycin Production in *Streptomyces Ambofaciens*. *Front. Microbiol.* **2017**, *8*, 835.
(18) Schellenberger, J.; Que, R.; Fleming, R. M. T.; Thiele, I.; Orth, J. D.; Feist, A. M.; Zielinski, D. C.; Bordbar, A.; Lewis, N. E.; Rahmanian, S.; et al. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6* (9), 1290–1307.
(19) Borodina, I.; Krabben, P.; Nielsen, J. Genome-Scale Analysis of *Streptomyces Coelicolor* A3(2) Metabolism. *Genome Res.* **2005**, *15* (6), 820–829.
(20) Borodina, I.; Siebring, J.; Zhang, J.; Smith, C. P.; van Keulen, G.; Dijkhuizen, L.; Nielsen, J. Antibiotic Overproduction in *Streptomyces Coelicolor* A3 (2) Mediated by Phosphofructokinase Deletion. *J. Biol. Chem.* **2008**, *283* (37), 25186–25199.
(21) Alam, M. T.; Merlo, M. E.; Hodgson, D. A.; Wellington, E. M. H.; Takano, E.; Breitling, R. Metabolic Modeling and Analysis of the Metabolic Switch in *Streptomyces Coelicolor*. *BMC Genomics* **2010**, *11* (1), 1.
(22) Borodina, I.; Schöller, C.; Eliasson, A.; Nielsen, J. Metabolic Network Analysis of *Streptomyces* Tenebrarius, a *Streptomyces* Species with an Active Entner-Doudoroff Pathway. *Appl. Environ. Microbiol.* **2005**, *71* (5), 2294–2302.
(23) Licona-Cassani, C.; Marcellin, E.; Quek, L.-E.; Jacob, S.; Nielsen, L. K. Reconstruction of the *Saccharopolyspora Erythraea* Genome-Scale Model and Its Use for Enhancing Erythromycin Production. *Antonie Van Leeuwenhoek* **2012**, *102* (3), 493–502.
(24) Zakrzewski, P.; Medema, M. H.; Gevorgyan, A.; Kierzek, A. M.; Breitling, R.; Takano, E. MultiMetEval: Comparative and Multi-Objective Analysis of Genome-Scale Metabolic Models. *PLoS One* **2012**, *7* (12), e51511.
(25) Thiele, I.; Palsson, B. Ø. A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction. *Nat. Protoc.* **2010**, *5* (1), 93–121.
(26) Kim, M.; Yi, J. S.; Kim, J.; Kim, J. N.; Kim, M. W.; Kim, B. G. Reconstruction of a High-Quality Metabolic Model Enables the Identification of Gene Overexpression Targets for Enhanced Antibiotic Production in *Streptomyces Coelicolor* A3(2). *Biotechnol. J.* **2014**, *9* (9), 1185–1194.
(27) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27–30.
(28) Caspi, R.; Billington, R.; Ferrer, L.; Foerster, H.; Fulcher, C. A.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L. A.; et al. The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2016**, *44* (D1), D471–D480.
(29) Henry, C. S.; DeJongh, M.; Best, A. A.; Frybarger, P. M.; Linsay, B.; Stevens, R. L. High-Throughput Generation, Optimization and Analysis of Genome-Scale

Metabolic Models. *Nat. Biotechnol.* **2010**, *28* (9), 977–982.
(30) Overbeek, R.; Olson, R.; Pusch, G. D.; Olsen, G. J.; Davis, J. J.; Disz, T.; Edwards, R. A.; Gerdes, S.; Parrello, B.; Shukla, M.; et al. The SEED and the Rapid Annotation of Microbial Genomes Using Subsystems Technology (RAST). *Nucleic Acids Res.* **2014**, *42* (D1), D206–D214.
(31) Nieselt, K.; Battke, F.; Herbig, A.; Bruheim, P.; Wentzel, A.; Jakobsen, Ø. M.; Sletta, H.; Alam, M. T.; Merlo, M. E.; Moore, J. The Dynamic Architecture of the Metabolic Switch in *Streptomyces Coelicolor*. *BMC Genomics* **2010**, *11* (1), 10.
(32) Melzoch, K.; De Mattos, M. J. T.; Neijssel, O. M. Production of Actinorhodin by *Streptomyces Coelicolor* A3(2) Grown in Chemostat Culture. *Biotechnol. Bioeng.* **1997**, *54* (6), 577–582.
(33) Varma, A.; Palsson, B. Ø. Parametric Sensitivity of Stoichiometric Flux Balance Models Applied to Wild-Type *Escherichia Coli* Metabolism. *Biotechnol. Bioeng.* **1995**, *45* (1), 69–79.
(34) Le Novere, N.; Finney, A.; Hucka, M.; Bhalla, U. S.; Campagne, F.; Collado-Vides, J.; Crampin, E. J.; Halstead, M.; Klipp, E.; Mendes, P.; et al. Minimum Information Requested in the Annotation of Biochemical Models (MIRIAM). *Nat. Biotechnol.* **2005**, *23* (12), 1509–1515.
(35) King, Z. A.; Lu, J.; Drager, A.; Miller, P.; Federowicz, S.; Lerman, J. A.; Ebrahim, A.; Palsson, B. Ø.; Lewis, N. E.; J., H. BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models. *Nucleic Acids Res.* **2016**, *44* (D1), D515–D522.
(36) Ebrahim, A.; Lerman, J. A.; Palsson, B. Ø.; Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7* (1), 74.
(37) Becker, S. A.; Feist, A. M.; Mo, M. L.; Hannum, G.; Palsson, B. Ø.; Herrgard, M. J. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox. *Nat. Protoc.* **2007**, *2* (3), 727–738.
(38) Rocha, I.; Maia, P.; Evangelista, P.; Vilaça, P.; Soares, S.; Pinto, J. P.; Nielsen, J.; Patil, K. R.; Ferreira, E. C.; Rocha, M. OptFlux: An Open-Source Software Platform for *in Silico* Metabolic Engineering. *BMC Syst. Biol.* **2010**, *4* (1), 45.
(39) Gelius-Dietrich, G.; Desouki, A.; Fritzemeier, C.; Lercher, M. J. Sybil – Efficient Constraint-Based Modelling in R. *BMC Syst. Biol.* **2013**, *7* (1), 125.
(40) Thomas, L.; Hodgson, D. A.; Wentzel, A.; Nieselt, K.; Ellingsen, T. E.; Moore, J.; Morrissey, E. R.; Legaie, R.; Wohlleben, W.; Rodriguez-Garcia, A.; et al. Metabolic Switches and Adaptations Deduced from the Proteomes of *Streptomyces Coelicolor* Wild Type and PhoP Mutant Grown in Batch Culture. *Mol. Cell. Proteomics* **2012**, *11* (2), M111.013797-M111.013797.
(41) Pawlik, K.; Kotowska, M.; Chater, K. F.; Kuczek, K.; Takano, E. A Cryptic Type I Polyketide Synthase (Cpk) Gene Cluster in *Streptomyces Coelicolor* A3(2). *Arch. Microbiol.* **2007**, *187* (2), 87–99.
(42) Kotowska, M.; Ciekot, J.; Pawlik, K. Type II Thioesterase ScoT Is Required for Coelimycin Production by the Modular Polyketide Synthase Cpk of *Streptomyces Coelicolor* A3(2). *Acta Biochim. Pol.* **2014**, *61* (1), 141–147.
(43) Awodi, U. R.; Ronan, J. L.; Masschelein, J.; de los Santos, E. L. C.; Challis, G. L. Thioester Reduction and Aldehyde Transamination Are Universal Steps in Actinobacterial Polyketide Alkaloid Biosynthesis. *Chem. Sci.* **2017**, *8* (1), 411–415.

(44) Gottelt, M.; Kol, S.; Gomez-Escribano, J. P.; Bibb, M.; Takano, E. Deletion of a Regulatory Gene within the Cpk Gene Cluster Reveals Novel Antibacterial Activity in *Streptomyces Coelicolor* A3 (2). *Microbiology* **2010**, *156* (8), 2343–2353.

(45) Wentzel, A.; Bruheim, P.; Øverby, A.; Jakobsen, Ø. M.; Sletta, H.; Omara, W. A. M.; Hodgson, D. A.; Ellingsen, T. E. Optimized Submerged Batch Fermentation Strategy for Systems Scale Studies of Metabolic Switching in *Streptomyces Coelicolor* A3(2). *BMC Syst. Biol.* **2012**, *6* (1), 59.

(46) Kato, J.; Funa, N.; Watanabe, H.; Ohnishi, Y.; Horinouchi, S. Biosynthesis of Gamma-Butyrolactone Autoregulators That Switch on Secondary Metabolism and Morphological Development in *Streptomyces*. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (7), 2378–2383.

(47) Takano, E. γ-Butyrolactones: Streptomyces Signalling Molecules Regulating Antibiotic Production and Differentiation. *Curr. Opin. Microbiol.* **2006**, *9* (3), 287–294.

(48) Biarnes-Carrera, M.; Breitling, R.; Takano, E. Butyrolactone Signalling Circuits for Synthetic Biology. *Curr. Opin. Chem. Biol.* **2015**, *28*, 91–98.

(49) Biarnes-Carrera, M.; Lee, C.-K.; Nihira, T.; Breitling, R.; Takano, E. Orthogonal Regulatory Circuits for *Escherichia Coli* Based on the γ-Butyrolactone System of *Streptomyces Coelicolor*. *ACS Synth. Biol.* **2018**, acssynbio.7b00425.

(50) Ogasawara, Y.; Kondo, K.; Ikeda, A.; Harada, R.; Dairi, T. Identification of Tirandamycins as Specific Inhibitors of the Futalosine Pathway. *J. Antibiot. (Tokyo).* **2017**, *70* (6), 798–800.

(51) Zhi, X.-Y.; Yao, J.-C.; Tang, S.-K.; Huang, Y.; Li, H.-W.; Li, W.-J. The Futalosine Pathway Played an Important Role in Menaquinone Biosynthesis during Early Prokaryote Evolution. *Genome Biol. Evol.* **2014**, *6* (1), 149–160.

(52) Esnault, C.; Dulermo, T.; Smirnov, A.; Askora, A.; David, M.; Deniset-Besseau, A.; Holland, I.-B.; Virolle, M.-J. Strong Antibiotic Production Is Correlated with Highly Active Oxidative Metabolism in *Streptomyces Coelicolor* M145. *Sci. Rep.* **2017**, *7* (1), 200.

(53) Kavvas, E. S.; Seif, Y.; Yurkovich, J. T.; Norsigian, C.; Poudel, S.; Greenwald, W. W.; Ghatak, S.; Palsson, B. O.; Monk, J. M. Updated and Standardized Genome-Scale Reconstruction of Mycobacterium Tuberculosis H37Rv, IEK1011, Simulates Flux States Indicative of Physiological Conditions. *BMC Syst. Biol.* **2018**, *12* (1), 25.

(54) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; et al. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44* (D1), D1202–D1213.

(55) Degtyarenko, K.; de Matos, P.; Ennis, M.; Hastings, J.; Zbinden, M.; McNaught, A.; Alcantara, R.; Darsow, M.; Guedj, M.; Ashburner, M. ChEBI: A Database and Ontology for Chemical Entities of Biological Interest. *Nucleic Acids Res.* **2007**, *36* (Database), D344–D350.

(56) Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I. InChI - the Worldwide Chemical Structure Identifier Standard. *J. Cheminform.* **2013**, *5* (1), 7.

(57) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *281413* (2715), 31–36.

(58) Scheltema, R. A.; Jankevics, A.; Jansen, R. C.; Swertz, M. A.; Breitling, R. PeakML/MzMatch: A File Format, Java Library, R Library, and Tool-Chain for Mass Spectrometry Data Analysis. *Anal. Chem.* **2011**, *83* (7), 2786–2793.

(59) Jankevics, A.; Merlo, M. E.; de Vries, M.; Vonk, R. J.; Takano, E.; Breitling, R. Metabolomic Analysis of a Synthetic Metabolic Switch in *Streptomyces Coelicolor* A3 (2). *Proteomics* **2011**, *11* (24), 4622–4631.

(60) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; et al. Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25* (1), 25–29.

(61) Kersey, P. J.; Allen, J. E.; Allot, A.; Barba, M.; Boddu, S.; Bolt, B. J.; Carvalho-Silva, D.; Christensen, M.; Davis, P.; Grabmueller, C.; et al. Ensembl Genomes 2018: An Integrated Omics Infrastructure for Non-Vertebrate Species. *Nucleic Acids Res.* **2018**, *46* (D1), D802–D808.

(62) Pruitt, K. D.; Tatusova, T.; Maglott, D. R. NCBI Reference Sequences (RefSeq): A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins. *Nucleic Acids Res.* **2007**, *35* (Database issue), D61-5.

(63) The UniProt Consortium. UniProt: A Hub for Protein Information. *Nucleic Acids Res.* **2015**, *43* (D1), D204–D212.

(64) Bateman, A.; Coin, L.; Durbin, R.; Finn, R. D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E. L. L.; et al. The Pfam Protein Families Database. *Nucleic Acids Res.* **2004**, *32* (90001), 138D – 141.

(65) Mi, H.; Muruganujan, A.; Thomas, P. D. PANTHER in 2013: Modeling the Evolution of Gene Function, and Other Gene Attributes, in the Context of Phylogenetic Trees. *Nucleic Acids Res.* **2012**, *41* (D1), D377–D386.

(66) Bystrykh, L. V; Fernández-Moreno, M. A.; Herrema, J. K.; Malpartida, F.; Hopwood, D. A.; Dijkhuizen, L. Production of Actinorhodin-Related" Blue Pigments" by *Streptomyces Coelicolor* A3 (2). *J. Bacteriol.* **1996**, *178* (8), 2238–2244.

(67) Lahtvee, P.-J.; Sánchez, B. J.; Smialowska, A.; Kasvandik, S.; Elsemman, I. E.; Gatto, F.; Nielsen, J. Absolute Quantification of Protein and MRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. *Cell Syst.* **2017**, *4* (5), 495-504.e5.

(68) Diez, V.; Loznik, M.; Taylor, S.; Winn, M.; Rattray, N. J. W.; Podmore, H.; Micklefield, J.; Goodacre, R.; Medema, M. H.; Müller, U.; et al. Functional Exchangeability of Oxidase and Dehydrogenase Reactions in the Biosynthesis of Hydroxyphenylglycine, a Nonribosomal Peptide Building Block. *ACS Synth. Biol.* **2015**, *4* (7), 796–807.

(69) Gyan, S.; Shiohira, Y.; Sato, I.; Takeuchi, M.; Sato, T. Regulatory Loop between Redox Sensing of the NADH/NAD(+) Ratio by Rex (YdiH) and Oxidation of NADH by NADH Dehydrogenase Ndh in Bacillus Subtilis. *J. Bacteriol.* **2006**, *188* (20), 7062–7071.

(70) Brekasis, D.; Paget, M. S. B. A Novel Sensor of NADH/NAD+ Redox Poise in Streptomyces Coelicolor A3(2). *EMBO J.* **2003**, *22* (18), 4856–4865.

(71) Zaburannyi, N.; Rabyk, M.; Ostash, B.; Fedorenko, V.; Luzhetskyy, A. Insights into Naturally Minimised *Streptomyces Albus* J1074 Genome. *BMC Genomics* **2014**, *15* (1), 97.

(72) Phelan, R. M.; Sachs, D.; Petkiewicz, S. J.; Barajas, J. F.; Blake-Hedges, J. M.; Thompson, M. G.; Reider Apel, A.; Rasor, B. J.; Katz, L.; Keasling, J. D. Development of Next Generation Synthetic Biology Tools for Use in *Streptomyces Venezuelae*. *ACS Synth. Biol.* **2017**, *6* (1), 159–166.
(73) Monk, J. M.; Lloyd, C. J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; et al. IML1515, a Knowledgebase That Computes Escherichia Coli Traits. *Nat. Biotechnol.* **2017**, *35* (10), 904–908.
(74) Sánchez, B. J.; Zhang, C.; Nilsson, A.; Lahtvee, P.-J.; Kerkhoven, E. J.; Nielsen, J. Improving the Phenotype Predictions of a Yeast Genome-Scale Metabolic Model by Incorporating Enzymatic Constraints. *Mol. Syst. Biol.* **2017**, *13* (8), 935.

2.8 Supplementary Files

2.8.1 Summary of the main updates and new features present in the *i*AA1259 model compared to previous generations

Metabolic models	**iAA1259**
Number of genes included	1259 (18 genes replaced + 51 new genes)
Reactions	1912 (68 reactions modified + 53 new reactions)
Metabolites	1470 (34 metabolites added)
Additional gene databases identifiers included	Uniprot, Gene Ontology, RefSeq, BioCyc, Pfam, PANTHER (Already in the previous model: StrepDB)
Additional metabolites databases identifiers	Chebi, HMDB, IUPAC, CAS, SMILES, InChi, ChemSpider, BioCyc, Metlin, PubChem (Already in the previous model: BiGG, KEGG)
Additional data included	Metabolites monoisotopic mass and structures (SMILES), Protein sequences
Metabolic pathways added	Coelimycin Biosynthesis Butyrolactones Biosynthesis Xylan Degradation Cellulose Degradation
Metabolic pathways updated	Futalosine Pathway Chitin Degradation Oxidative Phosphorylation NADH dehydrogenases Cytochromes oxidases Menaquinone utilization Biomass GAM/NGAM Composition (menaquinones, demethylmenaquinones, polyphosphate)

Supplementary Table 2.1: Summary table of the updates and new features added to the *i*AA1259 model compared to the previous generations. This table shows the main updates done on the iAA1259 model compared to previous models, in particular the updates on metabolic pathways.

Reaction ID	Reaction Name	Reaction Equation	Genes Associated
CPKS1	CpkA initiation	ACPcpk[c] + 3 malcoa[c] + 3 h[c] <=> 3 coa[c] + hex24dACPcpk[c] + 3 co2[c] + 2 h2o[c]	SCO6275
CPKS2	CpkB polyketide elongation 1	hex24dACPcpk[c] + 2 malcoa[c] + 2 h[c] -> 2 coa[c] + h246dectACPcpk[c] + 2 co2[c] + h2o[c]	SCO6274
CPKS3	CpkC polyketide elongation 2 and thioester reductase	h246dectACPcpk[c] + malcoa[c] + nadh[c] -> h2o[c] + hdd2610t[c] + 2 h[c] + co2[c] + nad[c] +coa[c]+ ACPcpk[c]	SCO6273
CPKS4a	CpkG transaminase (with L-Ala)	hdd2610t[c] + ala-L[c] <=> 1add26810t5o[c] + pyr[c]	SCO6279
CPKS4b	CpkG transaminase (with L-Glu)	hdd2610t[c] + glu-L[c] <=> 1add26810t5o[c] + akg[c]	SCO6279
CPKS5	Flavin dependent epoxidases/dehydrogenases	add26810t5o[c] + fadh2[c] + 2 o2[c] -> cpkepox[c] + fad[c] 3 h2o[c]	(SCO6276 or SCO6281 or SCO6272)
CPKSt	Putative transmembrane efflux protein CpkF	cpkepox[c] -> cpkepox[e]	SCO6278
CPKS6	Extracellular spontaneous reaction	cpkepox[e] + accyst[e] -> ycpk[e] + h2o[e] + 2 h[e]	s0001
ACCS	N-acetylcysteine synthase (putative acyl-transferase)	accoa[c] + cys-L[c] -> accys[c] + coa[c] + h[c]	

ACCt	N-acetylcysteine transport	accys[c] <=> accys[e]	
ACPScpk	acyl-carrier protein synthase (cpk)	coa[c] + apoACPcpk[c] -> h[c] + ACPcpk[c] + pap[c]	SCO4744
ACPSpdscpk	acyl-carrier-protein phosphodiesterase (cpk)	ACPcpk[c] + h2o[c] -> apoACPcpk[c] + pan4p[c] + h[c]	SCO0046
XYLAN_DEGe	Xylan degradation (extracellular)	xylan[e] + 527 h2o[e] -> 528 xyl_D[e]	(SCO5931 or SCO2292 or SCO1883 or SCO0674 or SCO0105)
EX_xylan(e)	Xylan exchange	xylan[e] <=>	
CELLUL_DEGe	Cellulose degradation (extracellular)	cellul[e] + 499 h2o[e] -> 250 celb[e]	SCO6548
EX_cellul(e)	Cellulose exchange	cellul[e] <=>	
CHTNDG	Chitin degradation	chitin[e] + h2o[e] -> acgam[e] + h[e]	(SCO7263 or SCO0482 or SCO1429 or SCO1444 or SCO5376 or SCO5954 or SCO6345 or SCO5003 or SCO5673 or SCO2833 or SCO6012 or SCO2503 or SCO7225)
CHDHR	Chorismate dehydratase	chor[c] -> 3cvobz[c] + h2o[c]	SCO4506
ADXFUTSNT	Aminodeoxyfutalosine synthase	3cvobz[c] + amet[c] + h2o[c] -> 6adxfut[c] + met-L[c] + hco3[c]	SCO4494
ADXFUTDA	Aminodeoxyfutalosine deaminase	6adxfut[c] + h2o[c] -> fut[c] + nh4[c]	SCO5662
ADXFUTNS	Aminodeoxyfutalosine nucleosidase	6adxfut[c] + h2o[c] -> dhxfut[c] + ade[c] + h[c]	
FUTH	Futalosine hydrolase	fut[c] + h2o[c] -> dhxfut[c] + hxan[c]	SCO4327
DXFUTOR	Dehypoxanthine futalosine:S-	dhxfut[c] + amet[c] -> cdhxfut[c] + met-	SCO4550

	adenosyl-L-methionine oxidoreductase	L[c] + dad-5[c] + h[c]	
DH6NPHS	1,4-dihydroxy-6-naphthoate synthase	cdhxfut[c] -> dh6na[c] + glyald[c]	SCO4326
DHNAI	1,4-dihydroxynaphthoate isomerization (hypothetical reaction)	dh6na[c] -> dhna[c]	
DHNANT4	1,4-dihydroxy-2-naphthoate nonaprenyltransferase	dhna[c] + h[c] + nndp[c] -> 2dmmql9[c] + co2[c] + ppi[c]	(SCO4491 and/or SCO4556)
SCB1_1	Acyl-transferase (using beta keto-acyl ACP precursor)	dhap[c] + 3oiC9ACP[c] -> apoACP[c] + 2o3pp8m3onn[c]	SCO6266
SCB1_2	Aldol condensation SCB1 precursor (infered as spontaneous)	2o3pp8m3onn[c] -> h2o[c] + mh5odhf3mdp[c]	s0001
SCB1_3	Butenolide phosphatase reductase (SCB1 precursor)	mh5odhf3mdp[c] + nadph[c] + h[c] -> mh5othfmdp[c] + nadp[c]	SCO6267
SCB1_4	SCB1 precursor phosphatase	mh5othfmdp[c] + h2o[c] -> afactor[c] + pi[c]	
SCB1_5	Beta-keto-acid-CoA/ACP reductase	afactor[c] + nadph{c] + h[c] -> scb1[c] + nadp[c]	SCO6264
SCB1t	SCB1 transport (diffusible molecule)	scb1[c] <=> scb1[e]	s0001
SCB2_1a	Acyl-transferase (using beta keto-acyl CoA precursor)	dhap[c] + 3odcoa[c] -> coa[c] + 2o3pop3odn[c]	SCO6266
SCB2_1b	Acyl-transferase (using beta keto-acyl ACP precursor)	dhap[c] + 3oddecACP[c] -> apoACP[c] + 2o3pop3odn[c]	SCO6266

SCB2_2	Aldol condensation SCB2 precursor (infered as spontaneous)	2o3pop3odn[c]-> h2o[c] + 4o5odhmdhp[c]	s0001
SCB2_3	Butenolide phosphatase reductase (SCB2 precursor)	4o5odhmdhp[c] + nadph[c] + h[c] -> 4o5othf3mdp[c] + nadp[c]	SCO6267
SCB2_4	SCB2 precursor phosphatase	4o5othf3mdp[c] + h2o[c] -> 4hm3odhf2o[c] + pi[c]	
SCB2_5	Beta-keto-acid-CoA/ACP reductase	4hm3odhf2o[c] + nadph{c] + h[c] -> scb2[c] + nadp[c]	SCO6264
SCB2t	SCB2 transport (diffusible molecule)	scb2[c] <=> scb2[e]	s0001
SCB3_1	Acyl-transferase (using beta keto-acyl ACP precursor)	dhap[c] + 3oiC10ACP[c] -> apoACP[c] + 2o3pp8m3onn[c]	SCO6266
SCB3_2	Aldol condensation SCB3 precursor (infered as spontaneous)	2o3pp8m3onn[c] -> h2o[c] + 46mo5o25dhf3mdhp[c]	s0001
SCB3_3	Butenolide phosphatase reductase (SCB3 precursor)	46mo5o25dhf3mdhp[c] + nadph[c] + h[c] -> 465mo5othf3mdhp[c] + nadp[c]	SCO6267
SCB3_4	SCB3 precursor phosphatase	465mo5othf3mdhp[c] + h2o[c] -> 4hm36modhf2o[c] + pi[c]	
SCB3_5	Beta-keto-acid-CoA/ACP reductase	4hm36modhf2o[c]+ nadph{c] + h[c] -> scb3[c] + nadp[c]	SCO6264
SCB3t	SCB3 transport (diffusible molecule)	scb3[c] <=> scb3[e]	s0001
CYOO	cytochrome o oxidase (menaquinol-9: 2 protons)	4 h[c] + mql9[c] + 0.5 o2[c] -> h2o[c] +mqn9[c] + 2h[e]	(SCO7234 and SCO7235 and SCO7236 and SCO1934 and SCO7120)

NADH8	NADH dehydrogenase (demethylmenaquinone-9 & 3 protons)	2dmmq9[c] + 4 h[c] + nadh[c] -> 2dmmql9[c] + 3 h[e] + nad[c]	((SCO4562 or SCO4599) and (SCO4563 or SCO4600) and SCO4564 and (SCO3392 or SCO4565) and SCO4566 and (SCO4567 or SCO6560) and SCO4568 and (SCO4569 or SCO4602) and (SCO4570 or SCO4603) and (SCO4571 or SCO4604) and (SCO4572 or SCO4605) and (SCO4573 or SCO4606 or SCO6954) and (SCO4574 or SCO4607) and (SCO4575 or SCO4608 or SCO6956))
NADH9	NADH dehydrogenase (demethylmenaquinone-9 & 0 protons)	2dmmq9[c] + h[c] + nadh[c] -> 2dmmql9[c] + nad[c]	(SCO3092 or SCO7101 or SCO7319 or SCO6496 or SCO0158 or SCO4119)
NADPHQR4	NADPH Quinone Reductase (2-Demethylmenaquinone-8)	2dmmq9[c] + h[c] + nadph[c] -> 2dmmql9[c] + nadp[c]	SCO3823
AMMQT9r	S-adenosylmethione:2-demthylmenaquinone methyltransferase (menaquinone 9)	2dmmq9[c] + amet[c] <=> ahcys[c] + h[c] + mqn9[c]	(SCO4556 or SCO5940)
FRD3	fumarate reductase (2-demthylmenaquinone-9)	2dmmql9[c] + fum[c] -> 2dmmq9c] + succ[c]	(SCO0923 and SCO0922 and SCO0924)
G3PD7	glycerol-3-phosphate dehydrogenase (2-demthylmenaquinone-9)	2dmmq8[c] + glyc3p[c] -> 2dmmql8[c] + dhap[c]	(SCO0670 or SCO1661 or SCO4774 or (SCO7005 and SCO7006))

GLYCTO4	Glycolate oxidase	2dmmq8[c] + glyclt[c] -> 2dmmql8[c] + glx[c]	SCO2925

Supplementary Table 2.2: Table of the new reactions added to the iAA1259 model.

Metabolite name	Metabolite description	Neutral metabolite formula	Metabolite Compartment
2o3pop3odn[c]	2-oxo-3-(phosphonooxy)propyl 3-oxodecanoate	C13H23O8P	Cytosol
2o3popp8m3odn[c]	2-oxo-3-(phosphonooxy)propyl 8-methyl-3-oxodecanoate	C14H25O8P	Cytosol
2o3pp8m3onn[c]	2-oxo-3-(phosphonooxy)propyl 8-methyl-3-oxononanoate	C13H23O8P	Cytosol
3cvobz[c]	3-[(1-Carboxyvinyl)oxy]benzoate	C10H6O5	Cytosol
465mo5othf3mdhp[c]	(3S,4S)-4-(6-methyloctanoyl)-5-oxotetrahydrofuran-3-yl)methyl dihydrogen phosphate	C14H25O7P	Cytosol
46mo5o25dhf3mdhp[c]	(4-(6-methyloctanoyl)-5-oxo-2,5-dihydrofuran-3-yl)methyl dihydrogen phosphate	C14H23O7P	Cytosol
4hm36modhf2o[c]	(3S,4R)-4-(hydroxymethyl)-3-(6-methyloctanoyl)dihydrofuran-2(3H)-one	C14H24O4	Cytosol
4hm3odhf2o[c]	(3S,4R)-4-(hydroxymethyl)-3-octanoyldihydrofuran-2(3H)-one	C13H22O4	Cytosol
4o5odhmdhp[c]	(4-octanoyl-5-oxo-2,5-dihydrofuran-3-yl)methyl dihydrogen phosphate	C13H21O7P	Cytosol
4o5othf3mdp[c]	((3S,4S)-4-octanoyl-5-	C13H23O7P	Cytosol

	oxotetrahydrofuran-3-yl)methyl dihydrogen phosphate		
6adxfut[c]	6-Amino-6-deoxyfutalosine	C19H19N5O6	Cytosol
accys[c]	N-Acetyl-L-Cysteine	C5H9NO3S	Cytosol
accys[e]	N-Acetyl-L-Cysteine	C5H9NO3S	Extra-organism
ACPcpk[c]	Acyl carrier protein (specific to actinorhodin coelimycin)	C11H22N2O7PRS	Cytosol
add26810t5o[c]	1-Aminododeca-2,6,8,10-tetraen-5-ol	C12H19NO	Cytosol
afactor[c]	A-factor: (4R)-4-(hydroxymethyl)-3-(6-methylheptanoyl)oxolan-2-one	C13H22O4	Cytosol
apoACPcpk[c]	apoprotein [acyl carrier protein] (cpk)	RHO	Cytosol
cdhxfut[c]	Cyclic dehypoxanthine futalosine	C14H14O7	Cytosol
cellul[e]	Cellulose, chain length 500 glc_D assumed, DOI: 10.1002/anie.200460587	C3000H5002O2501	Extra-organism
chitin[e]	Chitin monomer	C8H13NO5	Extra-organism
cpkepox[c]	6-(3'-(prop-1-en-1-yl)-[2,2'-bioxiran]-3-yl)-2,3-dihydropyridine	C12H15NO2	Cytosol
cpkepox[e]	6-(3'-(prop-1-en-1-yl)-[2,2'-bioxiran]-3-yl)-2,3-dihydropyridine	C12H15NO2	Extra-organism
h246dectACPcpk[c]	3-Hydroxy-2,4,6-Decatetraenoyl-[acyl-carrier protein]	C21H34N2O9PRS	Cytosol
hdd2610t[c]	5-Hydroxydodeca-2,6,8,10-tetraenal	C12H16O2	Cytosol
hex24dACPcpk[c]	2,4-Hexadienoyl-ACP (n-	C17H30N2O8PRS	Cytosol

	C6:2ACP)		
mh5odhfmdp[c]	(4-(6-methylheptanoyl)-5-oxo-2,5-dihydrofuran-3-yl)methyl dihydrogen phosphate	C13H21O7P	Cytosol
mh5othfmdp[c]	((3S,4S)-4-(6-methylheptanoyl)-5-oxotetrahydrofuran-3-yl)methyl dihydrogen phosphate	C13H23O7P	Cytosol
scb1[c]	SCB1: (3S,4R)-3-((R)-1-hydroxy-6-methylheptyl)-4-(hydroxymethyl)dihydrofuran-2(3H)-one	C13H24O4	Cytosol
scb1[e]	SCB1: (3S,4R)-3-((R)-1-hydroxy-6-methylheptyl)-4-(hydroxymethyl)dihydrofuran-2(3H)-one	C13H24O4	Extra-organism
scb2[c]	SCB2: (3S,4R)-4-(hydroxymethyl)-3-((R)-1-hydroxyoctyl)dihydrofuran-2(3H)-one	C13H24O4	Cytosol
scb2[e]	SCB2: (3S,4R)-4-(hydroxymethyl)-3-((R)-1-hydroxyoctyl)dihydrofuran-2(3H)-one	C13H24O5	Extra-organism
scb3[c]	SCB3: (3S,4R)-3-((1R)-1-hydroxy-6-methyloctyl)-4-(hydroxymethyl)dihydrofuran-2(3H)-one	C14H26O4	Cytosol
scb3[e]	SCB3:(3S,4R)-3-((1R)-1-hydroxy-6-methyloctyl)-4-(hydroxymethyl)dihydrofuran-2(3H)-one	C14H26O4	Extra-organism
xylan[e]	Oat spelt xylan, MW 79200 (DOI: 10.1002/masy.200551405) = 528 xyl_D	C2640H4226O2113	Extra-organism

Supplementary Table 2.3: Table of the new metabolites added to the iAA1259 model

Gene	Type of Addition	Characterization Year	Reference
SCO0105	new content added	2016	Enkhbaatar, Bolormaa, et al. "Molecular characterization of xylobiose-and xylopentaose-producing β-1, 4-endoxylanase SCO5931 from Streptomyces coelicolor A3 (2)." Applied biochemistry and biotechnology 180.2 (2016): 349-360.
SCO0284	update of GPR		KEGG (alpha-galactosidase)
SCO0382	update of GPR		KEGG (UDPglucose dehydrogenase)
SCO0462	update of GPR		KEGG (2-dehydropantoate 2-reductase)
SCO0482	new content added		BioCyc (chitinase)
SCO0554	update of GPR		KEGG (endoglucanase)
SCO0674	new content added	2016	Enkhbaatar, Bolormaa, et al. "Molecular characterization of xylobiose-and xylopentaose-producing β-1, 4-endoxylanase SCO5931 from Streptomyces coelicolor A3 (2)." Applied biochemistry and biotechnology 180.2 (2016): 349-360.
SCO0765	update of GPR		KEGG (endoglucanase)
SCO0984	update of GPR		KEGG (3-hydroxyacyl-CoA dehydrogenase)
SCO1268	update of GPR		KEGG (Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex)- PDH
SCO1429	new content		KEGG (chitinase)

	added		
SCO1444	new content added	2013	KEGG (chitinase) + Nazari, Behnam, et al. "Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3 (2) grown in soil." Applied and environmental microbiology 79.2 (2013): 707-713
SCO1883	new content added	2016	Enkhbaatar, Bolormaa, et al. "Molecular characterization of xylobiose-and xylopentaose-producing β-1, 4-endoxylanase SCO5931 from Streptomyces coelicolor A3 (2)." Applied biochemistry and biotechnology 180.2 (2016): 349-360.
SCO2154	update of GPR	2017	Sawers, R. G., D. Falke, and M. Fischer. "Chapter One-Oxygen and Nitrate Respiration in Streptomyces coelicolor A3 (2)." *Advances in microbial physiology* 68 (2016): 1-40.
SCO2292	new content added	2016	Enkhbaatar, Bolormaa, et al. "Molecular characterization of xylobiose-and xylopentaose-producing β-1, 4-endoxylanase SCO5931 from Streptomyces coelicolor A3 (2)." Applied biochemistry and biotechnology 180.2 (2016): 349-360.
SCO2503	new content added	2013	Nazari, Behnam, et al. "Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3 (2) grown in soil." Applied and environmental microbiology 79.2 (2013): 707-717
SCO2833	new content added	2013	Nazari, Behnam, et al. "Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3 (2) grown in soil." Applied and environmental microbiology 79.2 (2013): 707-715
SCO3823	new content added		Blast against SCO tr|A0A024H6J2|A0A024H6J2_9MICC NAD(P)H quinone oxidoreductase, PIG3 family protein OS=Pseudarthrobacter siccitolerans GN=qor PE=4 SV=1
SCO3835	update of	2017	Millan-Oropeza, Aaron, et al. "Quantitative proteomic analysis confirmed oxidative

	GPR		metabolism predominates in Streptomyces coelicolor versus glycolytic metabolism in Streptomyces lividans." Journal of Proteome Research (2017).
SCO4326	new content added	2013	Mahanta, Nilkamal, et al. "Menaquinone biosynthesis: formation of aminofutalosine requires a unique radical SAM enzyme." *Journal of the American Chemical Society* 135.41 (2013).
SCO4327	new content added	2013	Mahanta, Nilkamal, et al. "Menaquinone biosynthesis: formation of aminofutalosine requires a unique radical SAM enzyme." *Journal of the American Chemical Society* 135.41 (2013).
SCO4491	update of GPR	2013	Cooper, Lisa E., et al. "In vitro reconstitution of the radical SAM enzyme MqnC involved in the biosynthesis of futalosine-derived menaquinone." Biochemistry 52.27 (2013)
SCO4494	new content added	2013	Mahanta, Nilkamal, et al. "Menaquinone biosynthesis: formation of aminofutalosine requires a unique radical SAM enzyme." *Journal of the American Chemical Society* 135.41 (2013).
SCO4506	new content added	2013	Mahanta, Nilkamal, et al. "Menaquinone biosynthesis: formation of aminofutalosine requires a unique radical SAM enzyme." *Journal of the American Chemical Society* 135.41 (2013).
SCO4550	new content added	2013	Mahanta, Nilkamal, et al. "Menaquinone biosynthesis: formation of aminofutalosine requires a unique radical SAM enzyme." *Journal of the American Chemical Society* 135.41 (2013).
SCO4556	update of GPR	2013	Cooper, Lisa E., et al. "In vitro reconstitution of the radical SAM enzyme MqnC involved in the biosynthesis of futalosine-derived menaquinone." Biochemistry 52.27 (2013)
SCO4655	update of GPR	2017	Millan-Oropeza, Aaron, et al. "Quantitative proteomic analysis confirmed oxidative metabolism predominates in Streptomyces

			coelicolor versus glycolytic metabolism in Streptomyces lividans." Journal of Proteome Research (2017).
SCO5003	new content added	2013	Nazari, Behnam, et al. "Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3 (2) grown in soil." Applied and environmental microbiology 79.2 (2013): 707-713
SCO5376	new content added	2013	KEGG (chitinase) + Nazari, Behnam, et al. "Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3 (2) grown in soil." Applied and environmental microbiology 79.2 (2013): 707-713
SCO5498	update of GPR		KEGG (Glutamyl-tRNA synthetase)
SCO5499	update of GPR		KEGG (Glutamyl-tRNA synthetase)
SCO5662	new content added	2013	Mahanta, Nilkamal, et al. "Menaquinone biosynthesis: formation of aminofutalosine requires a unique radical SAM enzyme." *Journal of the American Chemical Society* 135.41 (2013).
SCO5673	new content added	2013	Nazari, Behnam, et al. "Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3 (2) grown in soil." Applied and environmental microbiology 79.2 (2013): 707-714
SCO5931	new content added	2016	Enkhbaatar, Bolormaa, et al. "Molecular characterization of xylobiose-and xylopentaose-producing β-1, 4-endoxylanase SCO5931 from Streptomyces coelicolor A3 (2)." Applied biochemistry and biotechnology 180.2 (2016): 349-360.
SCO5954	new content added		KEGG (chitinase)
SCO6012	new content added	2013	Nazari, Behnam, et al. "Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3 (2) grown in

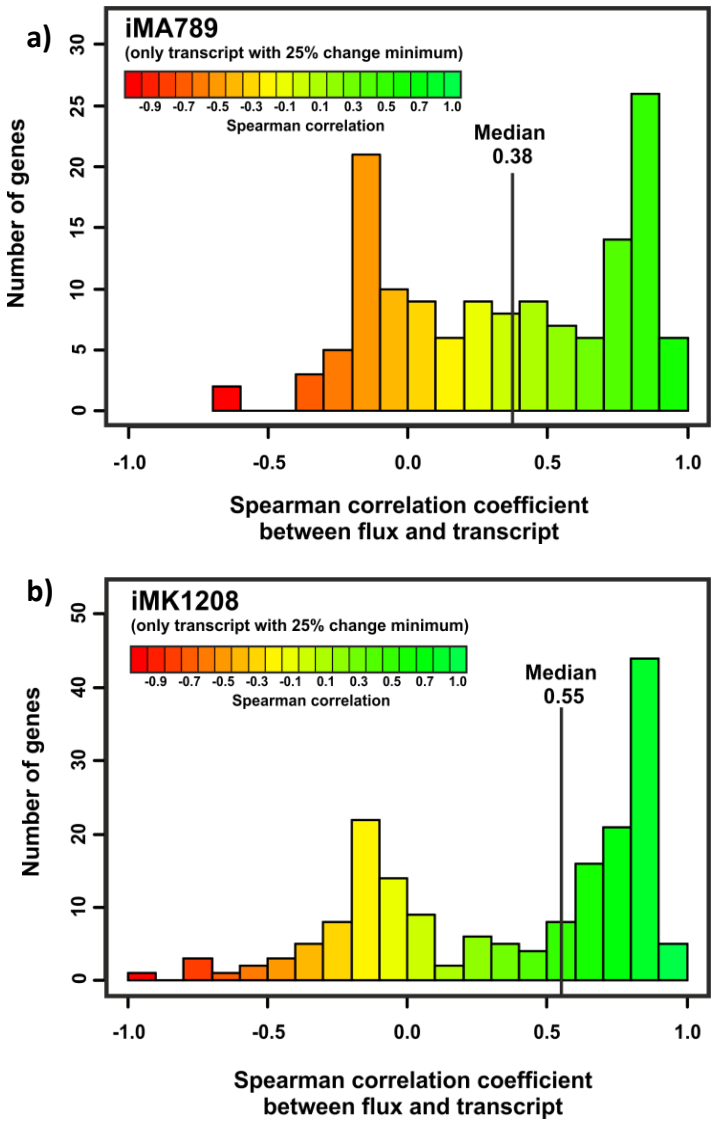
			soil." Applied and environmental microbiology 79.2 (2013): 707-716
SCO6272	new content added	2012	Gomez-Escribano JP, Song L, Fox DJ, Yeo V, Bibb MJ, Challis GL. Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of Streptomyces coelicolor M145. Chemical Science. 2012;3(9):2716-20.
SCO6273	new content added	2017	Awodi UR, Ronan JL, Masschelein J, de los Santos EL, Challis GL. Thioester reduction and aldehyde transamination are universal steps in actinobacterial polyketide alkaloid biosynthesis. Chemical Science. 2017;8(1):411-5. Pawlik K, Kotowska M, Chater KF, Kuczek K, Takano E. A cryptic type I polyketide synthase (cpk) gene cluster in Streptomyces coelicolor A3 (2). Archives of microbiology. 2007 Feb 1;187(2):87-99.
SCO6274	new content added	2007	Pawlik K, Kotowska M, Chater KF, Kuczek K, Takano E. A cryptic type I polyketide synthase (cpk) gene cluster in Streptomyces coelicolor A3 (2). Archives of microbiology. 2007 Feb 1;187(2):87-99.
SCO6275	new content added	2007	Pawlik K, Kotowska M, Chater KF, Kuczek K, Takano E. A cryptic type I polyketide synthase (cpk) gene cluster in Streptomyces coelicolor A3 (2). Archives of microbiology. 2007 Feb 1;187(2):87-99.
SCO6276	new content added	2012	Gomez-Escribano JP, Song L, Fox DJ, Yeo V, Bibb MJ, Challis GL. Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of Streptomyces coelicolor M145. Chemical Science. 2012;3(9):2716-20.
SCO6278	new content added	2012	Gomez-Escribano JP, Song L, Fox DJ, Yeo V, Bibb MJ, Challis GL. Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of Streptomyces coelicolor M145.

			Chemical Science. 2012;3(9):2716-20.
SCO6279	new content added	2017	Awodi UR, Ronan JL, Masschelein J, de los Santos EL, Challis GL. Thioester reduction and aldehyde transamination are universal steps in actinobacterial polyketide alkaloid biosynthesis. Chemical Science. 2017;8(1):411-5. Pawlik K, Kotowska M, Chater KF, Kuczek K, Takano E. A cryptic type I polyketide synthase (cpk) gene cluster in Streptomyces coelicolor A3 (2). Archives of microbiology. 2007 Feb 1;187(2):87-99.
SCO6279	new content added	2017	Awodi UR, Ronan JL, Masschelein J, de los Santos EL, Challis GL. Thioester reduction and aldehyde transamination are universal steps in actinobacterial polyketide alkaloid biosynthesis. Chemical Science. 2017;8(1):411-5. Pawlik K, Kotowska M, Chater KF, Kuczek K, Takano E. A cryptic type I polyketide synthase (cpk) gene cluster in Streptomyces coelicolor A3 (2). Archives of microbiology. 2007 Feb 1;187(2):87-99.
SCO6281	new content added	2012	Gomez-Escribano JP, Song L, Fox DJ, Yeo V, Bibb MJ, Challis GL. Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of Streptomyces coelicolor M145. Chemical Science. 2012;3(9):2716-20.
SCO6345	new content added		BioCyc (chitinase)
SCO6548	new content added	2016	Lim, Ju-Hyeon, et al. "Molecular characterization of Streptomyces coelicolor A (3) SCO6548 as a cellulose 1, 4-β-cellobiosidase." FEMS microbiology letters 363.3 (2016)
SCO6712	misannotation	2013	Sherif, Mohammed, et al. "Biochemical studies of the multicopper oxidase (small laccase) from Streptomyces coelicolor using bioactive phytochemicals and site-directed mutagenesis." *Microbial biotechnology* 6.5

			(2013): 588-597
SCO6956	new content added	2017	Sawers, R. G., D. Falke, and M. Fischer. "Chapter One-Oxygen and Nitrate Respiration in Streptomyces coelicolor A3 (2)." *Advances in microbial physiology* 68 (2016): 1-40.
SCO7225	new content added	2013	Nazari, Behnam, et al. "Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3 (2) grown in soil." Applied and environmental microbiology 79.2 (2013): 707-718
SCO7263	new content added	2013	BioCyc (chitinase) + Nazari, Behnam, et al. "Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3 (2) grown in soil." Applied and environmental microbiology 79.2 (2013): 707-713
SCO7266	update of GPR		KEGG (3-oxoacyl-[acyl-carrier-protein] reductase)- PDH
SCO7637	update of GPR		KEGG (endoglucanase)

Supplementary Table 2.4: Table of the new genes added to the iAA1259 model

2.8.2 Correlation analysis between gene expression and predicted fluxes for *i*MA789 and *i*MK1208 (gene expression showing a variation superior to 25%)



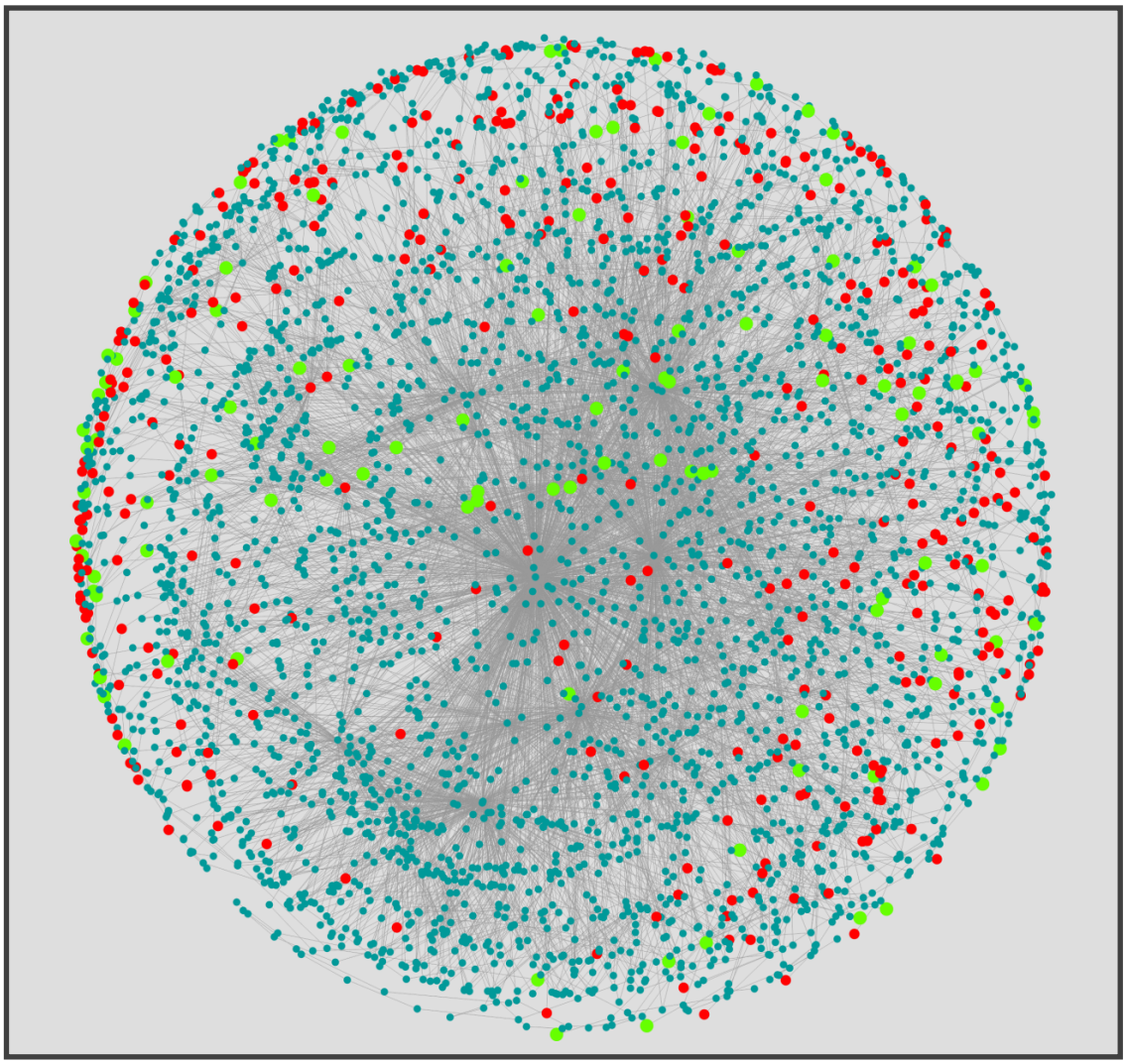
Supplementary Figure 2.1: Correlation analysis of gene expression to predicted fluxes of *i*MA789, and *i*MK1208 (for transcripts showing a change in level of expression superior to 25%).

a) Histogram of correlations for the *i*MA789 model, only taking in account genes with expression variation of more than 25% between the minimal and maximal transcript level.

b) Histogram of correlations for the *i*MK1208 model, only taking in account genes with expression variation of more than 25% between the minimal and maximal transcript level.

For both cases there is an increase of correlation compared to the whole expression dataset correlation, *i*MA789 goes from an overall Pearson correlation of 0.13 to 0.38, and *i*MK1208 from 0.18 to 0.55. This emphasis the improvements in predictions showed by *i*AA1259, with an increase in correlation from 0.56 to 0.78 when filtering for changing transcripts (Fig. 3c and 3d).

2.8.3 Automated mapping of an untargeted metabolomics dataset onto the metabolic network iAA1259, using standardized metabolite identifiers



Supplementary Figure 2.2: Mapping of observed metabolites in an untargeted metabolomics dataset onto the metabolic network.

To illustrate the easy integration of experimental data with the computational model, once standard metabolite identifiers are used, an LC-MS metabolomics dataset obtained from the literature (Jankevics et al., 2011) was automatically processed and annotated using MzMatch, then mapped onto the updated metabolic network *i*AA1259. The red dots correspond to the metabolites putatively annotated in the dataset (level 2 of confidence for metabolite identification [8]) , and the green dots represent the metabolites annotated with high confidence using chemical standards (level 1 of confidence for metabolite identification [8]). Figure generated with the software Cytoscape (Shannon et al., 2003). Once metabolites have been mapped, flux estimates inferred from the metabolomics experiment can, e.g., be used as additional flux constraints. A similar analysis based on traditional models using arbitrary metabolite identifiers would require manual curation for each metabolite and would rapidly become prohibitively time-consuming.

2.8.3.1 Supplementary Method for the metabolomics data analysis and mapping

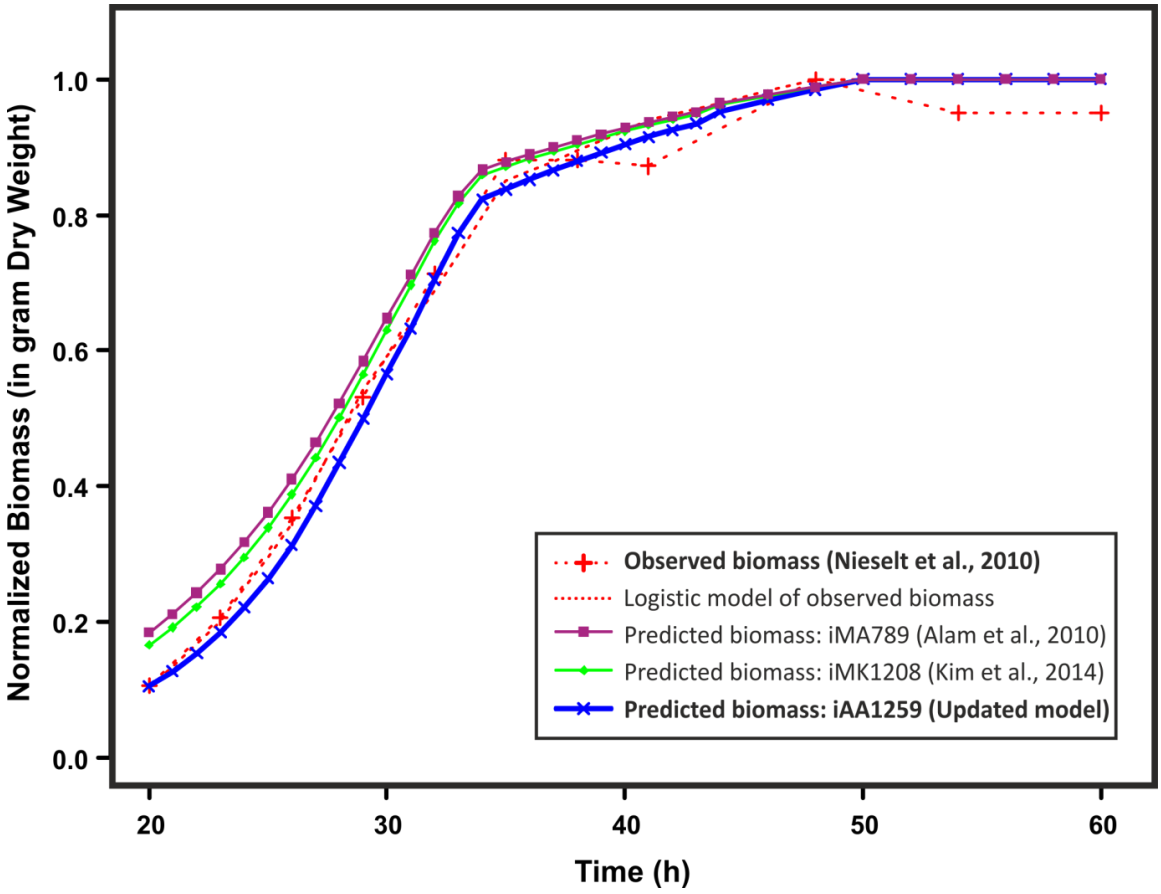
The untargeted metabolomics dataset used for illustration of the metabolite mapping enabled by the standard identifiers used in the updated model is derived from a previous *S. coelicolor* study using LC-MS (HILIC and C18 columns), with a large number of technical and biological replicates [1]. The data was processed and analysed using MzMatch [2], using a similar protocol as described in the initial study [1]; however, only metabolites putatively annotated in at least 80 samples were retained for the visualization. The annotated metabolites were automatically mapped to their corresponding annotation in the model, which was used as an input for highlighting the detected metabolites on the metabolic network visualisation in Cytoscape [3].

2.8.4 Constraints and *in-silico* growth rates predictions used for the Figure 1

Sample N°	**Glucose uptake (mmol/gDW.h)**	**O_2 (mmol/g.h)**	**CO_2 (mmol/g.h)**	**Actinorhodin (µ g/g.h)**	
1	0.5	1.8	1.9	2	
2	0.6	2	2	2	
3	0.8	2.4	2.5	415	
4	0.9	2.5	2.7	152	
5	1.1	3.1	3.1	60	
6	1.85	6.6	6.7	7	
7	2.1	7.2	7	5	
Sample N°	**Measured Growth Rate (gDW/h)**	**Predicted Growth Rate iIB711 (gDW/h)**	**Predicted Growth Rate iMA789 (gDW/h)**	**Predicted Growth Rate iMK1208 (gDW/h)**	**Predicted Growth Rate iAA1259 (gDW/h)**
1	0.035	0.0185	0.0272	0.0278	0.0284
2	0.045	0.0253	0.0396	0.0405	0.0412
3	0.06	0.0382	0.0539	0.0577	0.0588
4	0.072	0.0414	0.0657	0.0681	0.0694
5	0.092	0.0615	0.0862	0.0885	0.0901
6	0.115	0.1770	0.1088	0.1113	0.1134
7	0.128	0.1965	0.1385	0.1417	0.1443

Supplementary Table 2.5: Constraints used and predicted growth rates of the different models. This table contains the data used to constrain the genome-scale metabolic models (top table), these data were acquired by Melzoch et al. [4]. The table also contains the experimental specific growth rates, and the growth rates predicted by the different metabolic models of *S. coelicolor* (bottom part): *i*IB711 [5], *i*MA789 [6], *i*MK1209 [7], and *i*AA1259.

2.8.5 Metabolic models qualitative biomass predictions comparison



Supplementary Figure 2.3: Comparison of the normalized growth prediction of the metabolic models to the experimental data

When the dynamic growth data is normalized, the *i*AA1259 model shows a qualitative improvement in the prediction of the biomass. The *i*AA1259 model predictions (purple curve) are closer to the experimental data (red curve) than the previous models iMA789 (pink) and *i*MK1208 (green). The *i*AA1259 seems to enable better prediction of the growth curve from exponential to stationary phase.

2.9 Supplementary References

(1) Jankevics, A.; Merlo, M. E.; de Vries, M.; Vonk, R. J.; Takano, E.; Breitling, R. Metabolomic Analysis of a Synthetic Metabolic Switch in *Streptomyces Coelicolor* A3 (2). *Proteomics* **2011**, *11* (24), 4622–4631.

(2) Scheltema, R. A.; Jankevics, A.; Jansen, R. C.; Swertz, M. A.; Breitling, R. PeakML/MzMatch: A File Format, Java Library, R Library, and Tool-Chain for Mass Spectrometry Data Analysis. *Anal. Chem.* **2011**, *83* (7), 2786–2793.

(3) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504.

(4) Melzoch, K.; De Mattos, M. J. T.; Neijssel, O. M. Production of Actinorhodin by *Streptomyces Coelicolor* A3(2) Grown in Chemostat Culture. *Biotechnol. Bioeng.* **1997**, *54* (6), 577–582.

(5) Borodina, I.; Krabben, P.; Nielsen, J. Genome-Scale Analysis of *Streptomyces Coelicolor* A3(2) Metabolism. *Genome Res.* **2005**, *15* (6), 820–829.

(6) Alam, M. T.; Merlo, M. E.; Hodgson, D. A.; Wellington, E. M. H.; Takano, E.; Breitling, R. Metabolic Modeling and Analysis of the Metabolic Switch in *Streptomyces Coelicolor*. *BMC Genomics* **2010**, *11* (1), 1.

(7) Kim, M.; Yi, J. S.; Kim, J.; Kim, J. N.; Kim, M. W.; Kim, B. G. Reconstruction of a High-Quality Metabolic Model Enables the Identification of Gene Overexpression Targets for Enhanced Antibiotic Production in *Streptomyces Coelicolor* A3(2). *Biotechnol. J.* **2014**, *9* (9), 1185–1194.

(8) Sansone, S.-A.; Fan, T.; Goodacre, R.; Griffin, J. L.; Hardy, N. W.; Kaddurah-Daouk, R.; Kristal, B. S.; Lindon, J.; Mendes, P.; Morrison, N.; et al. The Metabolomics Standards Initiative. *Nat. Biotechnol.* **2007**, *25* (8), 846–848.

2.10 Acknowledgments

We thank Dr. Tauqeer Alam for his support with the *i*MA789 model analysis. We also thank Areti Tsigkinopoulou and Francesco Del Carratore for careful reading of the manuscript.

Funding

AA was supported by the School of Chemistry, Faculty of Science and Engineering, University of Manchester. This is a contribution from the Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM) and acknowledges the Biotechnology and Biological Sciences Research Council (BBSRC) and Engineering

and Physical Sciences Research Council (EPSRC) for financial support (Grant No. BB/M017702/1). This work received funding from the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 720793, H2020 TOPCAPI project).

Availability of data and materials

The *i*AA1259 genome-scale metabolic model is available in SBML and Excel format: *i*AA1259.xml and *i*AA1259.xls.

Chapter III

3. Comparative metabolic modelling of *Streptomyces coelicolor* and *Streptomyces lividans*: exploring the impact of primary metabolism variations on antibiotics production

Adam Amara, Rainer Breitling, and Eriko Takano*

Manchester Synthetic Biology Research Centre SYNBIOCHEM, Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, UK

***Corresponding author**: Eriko Takano, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK. Tel:+44 (0)161 306 4419. E-mail: eriko.takano@manchester.ac.uk

3.1 Preface

The work carried here was an *in silico* study, so all the experimental data used and discussed were from published or publicly available data. This work is presented as a chapter of this thesis and do not aim to be published.

3.2 Abstract

Streptomyces species are a rich source of secondary metabolites of medical and industrial interest, including antimicrobials that are increasingly sought after to tackle the spread of resistance. To facilitate the discovery of new antibiotics and increase the production of established ones, we need to better characterize and understand the metabolic, genetic, and environmental causes influencing antibiotics biosynthesis to both rationally trigger or increase production.

In this chapter, two very similar (99.7% amino-acid identity) *Streptomyces* strains, *Streptomyces coelicolor* A3(2) and *Streptomyces lividans* were analysed. These

two strains encode the same antibiotics biosynthesis pathways but trigger them differently and produce the antibiotics at different levels. The two strains were studied computationally in a detailed comparative metabolic analysis to characterize major metabolic differences; as well as trying to understand why *S. coelicolor* is a better antibiotic producer than *S. lividans*. Both strains possess biosynthetic genes clusters for four known antibiotics: actinorhodin, undecylprodigiosin, coelimycin P1, and calcium-dependant antibiotic. The differences in secondary metabolites production between these two strains are directly linked to their primary metabolism, and particularly the central carbon metabolism.

Despite being very similar at the genomic level, a few very specific metabolic differences have been shown to impact production such as presence or absence of enzymes involved in central metabolism. However, this comparison also suggests that some of the major regulatory phenomena observed are in response to metabolic differences in central metabolism.

3.3 Introduction

New antibiotics are increasingly needed as the threat of resistant bacteria has risen at an alarming rate [1]. The Actinobacteria, including the *Streptomyces* spp., have been a rich source of antibiotics for decades, providing about two thirds of the current antibiotics arsenal [2]. Although there are still many undiscovered secondary metabolites with potential antibiotic activity in this genus [3], their identification requires leveraging new strategies, such as a combination of genomics and computational methods[3,4]. Furthermore, once identified and expressed, the production titres of these compounds in their natural hosts are frequently too low to characterise the compounds, or in

other cases does not reach economically viable production levels. This can be overcome by strain engineering using computational modelling and synthetic biology, to optimize the production of the compounds either in the native strain or in a new host.

There have been significant advancements in the understanding of the complex physiology of *Streptomyces* species such as *Streptomyces coelicolor* and *Streptomyces lividans* [5]. However, a more detailed characterisation of the metabolic features associated with high production levels of secondary metabolites is still necessary to understand the genetic or external causes underlying production. This would open the possibility to rationally engineer better strains and conditions to produce and discover secondary metabolites of clinical or industrial interest using synthetic biology [6–8].

The *Streptomyces lividans* TK24 and *Streptomyces coelicolor* A3(2) genomes are highly similar (>99.7% Amino Acid identity), also both have the same biosynthetic gene clusters (*Figure 3.1*). Both strains contain some major biosynthetic gene clusters (BGCs), such as: Actinorhodin (ACT) — a type II polyketide synthase (PKS) product, Coelimycin P1 (CPK) — a type I PKS product, Calcium-Dependant Antibiotic (CDA) — a non-ribosomal peptide synthase (NRPS) product, and Undecylprodigiosin (RED) — a hybrid NRPS-PKS product. Actinorhodin is a pH dependent blue-pigmented antibiotic [9], and undecylprodigiosin is a tripyrrole red-pigmented antibiotics [10], as these compounds are easily measured by spectrophotometry, their production have been extensively analysed under different conditions in both *Streptomyces* strains. These two strains produce antibiotics very differently. *S. coelicolor* is known for producing higher levels of actinorhodin and undecylprodigiosin under standards laboratory conditions compared to *S. lividans*, which shows very low or no production under

similar conditions [11,12]. These differences in the awakening and levels of production of antibiotics make the comparative studies of these two strains attractive to understand the cause of these differences to design better *Streptomyces* antibiotic producing strains. Another reason to study these two strains is that *S. coelicolor* is a model strain for antibiotic production in the *Streptomyces* genus, and it produces diverse types of secondary metabolites. So having four different BGCs in the genome allows to study the metabolic impact and characteristics of four different types of antibiotics in the model strain.

The *S. coelicolor* strains have been subject to multiple experiments to attempt to understand and increase its antibiotics production. For example, the overexpression of an acetyl-CoA carboxylase (*acc*) increased ACT yield as it increased acetyl-CoA consuming fluxes toward malonyl-CoA (ACT main precursor) [13]. Also, the reduction of fluxes through the pentose-phosphate pathway (PPP) was tested by deleting the first enzyme of the PPP (a glucose-6-phosphate dehydrogenase (*zwf*)) led to a decreased yield of ACT [13]. The deletion of a citrate synthases (citA) and an aconitase (acoA) – enzymes of the TCA cycle (making *S. coelicolor* a glutamate auxotroph in minimal media) led to defective antibiotic biosynthesis [14,15]. The deletion of a phosphofructokinase isoenzyme (*pfkA2*) resulted in reduced glycolysis fluxes and increased fluxes toward the PPP; which pushed fluxes toward the phosphoketolase pathway increasing acetyl-CoA production, increasing production of ACT and RED [16]. The deletion of the two malic enzymes responsible for the conversion of L-malate into pyruvate (generating NADH and NADPH) led to a lower production of ACT and an accumulation of TAG [17]. The malic enzymes are involved in the anaplerotic metabolism, part of the PEP-pyruvate-oxaloacetate node involved in the switch of

carbon fluxes between catabolic and anabolic metabolism [18]. The deletion of the malic enzymes could trigger an accumulation of TCA cycle intermediates disturbing the carbon flux balance [18]. In contrast, increased fluxes through this enzyme could divert fluxes consuming TCA cycle intermediates to generate pyruvate and afterward acetyl-CoA.

In *S. lividans*, the deletion of two glucose-6-phosphate dehydrogenase isoenzymes (*zwf1* & *zwf2*) caused lower carbon fluxes through the PPP making glycolysis more efficient, and increased antibiotics production [19]. In the same study, the deletion of the 6-phosphogluconolactonase (*devB*) (second enzyme in the PPP) abolished the ACT and RED production, probably due to a reduced generation of NADPH (used for antibiotic biosynthesis) through the PPP [19]. A previous study pointed out that these results were in contradiction with the one observed in *S. coelicolor* [16]. However, the PPP and glycolysis is higher in *S. lividans* probably to generate sufficient amount of NADPH, glycerol-3-phosphate, and acetyl-CoA for a stronger TAG biosynthesis [20]. Whereas *S. coelicolor* has a lower glycolytic metabolism which might lead to reduced fluxes through the PPP. Another experimental study using metabolic flux analysis under phosphate limited conditions confirmed that lower fluxes through the PPP increased antibiotic production [21]. Otherwise, polyphosphate kinase (ppk) deletion in *S. lividans* led to higher ADP, lower polyphosphate, and TAG content than in the WT, which activated actinorhodin production [11,22,23].

Computational modelling methods such as genome-scale metabolic modelling (GSMMs) opened new opportunities to study microorganism metabolic phenotypes and physiology [24,25]. These developments are informing design strategies for synthetic biology applications [26,27]. One of the most popular modelling frameworks is the

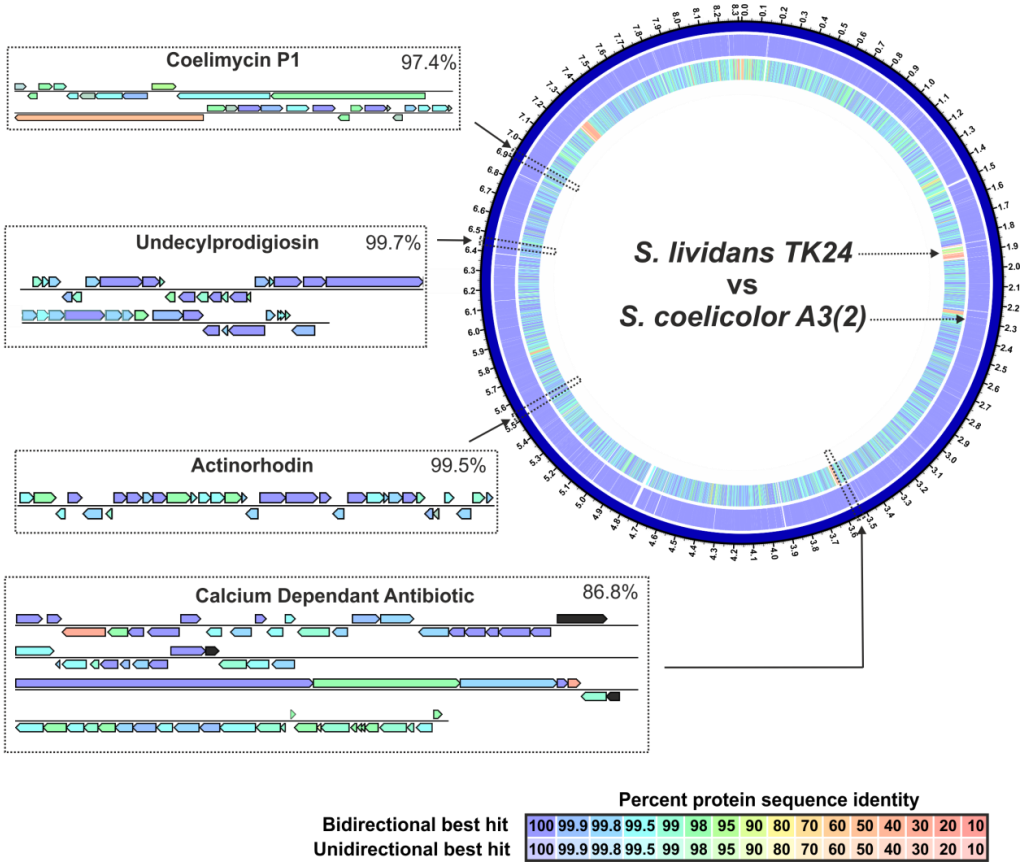


Figure 3.1: Circular representation of the comparison of the *S. coelicolor* and *S. lividans* genome and proteome

The comparison of the predicted proteome (mapped on the genome) of *S. coelicolor* (second outer ring) with the *S. lividans* (inner central ring) shows a very high protein sequence identity (second inner ring), with an overall protein identity >99.7%. Both strains contain the biosynthetic gene clusters (BGCs) for Coelimycin P1 (CPK), Undecylprodigiosin (RED), Actinorhodin (ACT), and Calcium-Dependant Antibiotic (CDA). As seen in the figure, the four antibiotic BGCs are found in the same locations on both chromosomes, as are all the genes constituting the BGCs (AA identity between 86.8 and 99.7%). The genes colours in the BGCs correspond to the protein sequence identity scale, genes in black are absent from *S. lividans*. . The few differences are mainly located in the genomic islands missing from *S. lividans* genome (white areas in the central ring), which were previously described [34,38]. Understanding the impact of the metabolic genes lost by *S. lividans* could help to identify the genes responsible for the difference in antibiotics production between the two strains, which we will identify using comparative metabolic modelling.

constraint-based reconstruction and analysis (COBRA) toolbox, which includes methods such as flux balance analysis (FBA) [28,29]. This method enables the prediction of metabolic states with a limited amount of data by using stoichiometric, environmental, thermodynamic, or flux constraints [28,29]. FBA is a method that uses linear programming

to predict quantitative growth phenotypes using a stoichiometric model, assuming that the metabolic network of a cell is at steady-state at a given time point [30]. FBA is efficient to predict relationships between the input (consumed nutrients) and the output (growth and/or metabolic by-product) [31]. Another similar method used is parsimonious FBA (pFBA), which is adding the assumption that the cell minimizes the enzymatic cost when other optimal routes exist; this method selects the optimal flux distributions while minimizing absolute fluxes [32]. A more detailed description of constraint-based modelling and genome-scale metabolic model can be found in the *Chapter I*.

Here, we applied metabolic modelling to explore and understand from a systems point of view, the metabolic, genetic, and environmental causes impacting production of antibiotics in *S. coelicolor* and *S. lividans*. For this purpose, a GSMM for *S. lividans* TK24 was generated from a previous reconstruction of *S. coelicolor* A3(2) [33]. As the two strains are so similar and *S. lividans* has mainly undergone loss of genetic material compared to *S. coelicolor* [34], *S. coelicolor* could be used as the suitable reference for the model construction. Multiple approaches were applied to explore the metabolic characteristics of the two strains. For the first time, the metabolic model was used to translate the differences between the two genomes into metabolic differences, and to identify metabolic functions lost and gained by *S. lividans*. The *S. lividans* GSMM was validated by comparing the growth predictions to observations from chemostat data to constrain nutrient uptakes. Then, the two strains' metabolic models were constrained with experimental data[35] to predict and compare their metabolic states, when *S. coelicolor* produced ACT and not *S. lividans* in a complex media (R2YE)[11]. A comparative proteomics dataset of the two strains was integrated in the two models

to identify metabolic differences that were not described in the original study of the sole proteomics data [20]. The models were then used to predict the impact of producing the four antibiotics – ACT, CDA, CPK, and RED – in the two strains. Finally, from collating the metabolic modelling results, and integrative published omics and experimental data analysis, we identified potential targets to increase antibiotics production (gene knock-outs, gene overexpression, or heterologous enzyme expression). Some important differences between the two strains were highlighted in the central and primary metabolism, such as the loss of the glyoxylate bypass metabolism and the presence of a pyruvate formate lyase in *S. lividans*, which are predicted to have a major impact on the central metabolism and production of secondary metabolites [36,37]. The metabolic modelling also helped to identify or confirm some key metabolic differences between the two strains, such as a lower glycolytic metabolism, a higher amino-acid catabolism, a higher gluconeogenesis, a higher oxidative phosphorylation in *S. coelicolor*. Also, the differences predicted by the models for antibiotics production are probably linked to the differences in metabolic regulation (particularly of central metabolism) rather than the metabolic network itself (e.g., a missing pathway limiting fluxes toward the antibiotics). Finally, some targets of interest include modifications of central metabolism that could force fluxes toward secondary metabolite precursors, such as the overexpression of acetyl-CoA carboxylase or the heterologous expression of a pyruvate formate lyase (with electron transporters).

3.4 Methods & Material

3.4.1 Comparative reconstruction of *Streptomyces lividans* genome-scale metabolic model

The genome-scale stoichiometric metabolic model of *S. lividans* TK24 (*i*SLT1240) was reconstructed using the metabolic model of the model *Streptomyces* strain *S. coelicolor A3*(2), *i*AA1259 [33], as template. The genome of *S. lividans* TK24 [39] was compared to the genome sequence of *S. coelicolor* A3(2) [40], using both the genomic sequence and the predicted protein sequences. The two predicted proteomes were compared using BlastX and PATRIC, to identify the genes missing from the *S. lividans* TK24 strain and identify the genes only present in *S. lividans*. This list of lost or gained genes was used to delete or add genes in the *i*AA1259 model to build the *i*SLT1240 model. The genes with a protein sequence percent identity <85% (401 genes out of 7021 genes matched between SCO and SLIV) were independently analysed with OrthoDB to find orthologues or homologues hit in the SLIV genome [41]. The genes with a protein sequence identity <90% and with no OrthoDB hit were deleted from the template model; if no other genes were associated to the reaction, the reaction was deleted; but if other genes were still associated the reaction was left in the model. The genes only present in SLIV were individually analysed to identify potential orthologues in *S. coelicolor* and identify the genes possible metabolic function. If the gene showed a metabolic function absent in *S. coelicolor*, then a new reaction associated (and new metabolites) was added. Before any modification on the template Gene-Protein-Reaction (GPR) associations, the genes to be deleted or added were compared to genome annotations in StrepDB (http://strepdb.streptomyces.org.uk), BioCyc [42], KEGG

[43], and UniProt [44]. A model of *S. lividans* TK24 automatically reconstructed with SEED [45] helped identify reactions and metabolites to add to the *i*SLT1240 model.

The biomass of the *S. lividans* TK24 model was modified compared to the *i*AA1259 biomass. As ATP/ADP ratios have been shown to be different and were quantitatively measured [11], the data was used to update to a lower ATP to ADP ratio in the *i*SLT1240 model. The amount of TAG was updated in the biomass based on the difference in triacylglycerol content observed experimentally, with a higher content in the *S. lividans* biomass. The modifications done on the *i*SLT1240 model compared to the *i*AA1259 model biomass are detailed in *Electronic Supplementary 3.1 - Additional File 1*.

The loss of some genes/reactions led to a *S. lividans* model unable to grow on some carbon sources, such as xylitol, rhamnose, and sorbitol. This has been tested by switching off carbon source uptake for selected compounds, such as glucose, and checking if predicted growth (with FBA) is achievable using this single carbon source. This was tested with both *i*AA1259 and *i*SLT1240.

Detailed modifications on the template model *i*AA1259 (gene, reactions, and metabolites deletion or additions), and the comparison of *S. lividans* TK24 genome to *S. coelicolor* A3(2) genome are available in the *Electronic Supplementary 3.1 - Additional File 1*.

The model is available in SBML format (*Electronic Supplementary 3.1 - Additional File 2*) to be fully compatible with the COBRA Toolbox [46] and COBRApy [47]. COBRApy was used to analyse, validate, and improve the model. The *in silico* simulations were performed using constraint-based modelling algorithms, such as parsimonious flux balance analysis (pFBA).

3.3.2 Model predictions validations and metabolic comparison

First, the predictions of *i*SLT1240 were validated using chemostat data from a *S. lividans* TK24 experiment in a glucose-limited media used for a previous metabolic model [35]. The data was used to constrain glucose and ammonium consumption and predict *S. lividans* TK24 biomass across time. The *i*SLT1240 metabolic model growth predictions were in good agreement with the observed growth in the chemostat.

The metabolic models *i*AA1259 and *i*SLT1240 of the two strains were constrained with uptake and export data for phosphate, glucose, and gamma-actinorhodin from a published experiment with phosphate-limited solid R2YE medium [11]. The first set of constraints were built based on the R2YE medium nutrients content [48]; as this is a complex media with multiple sources of nutrients, the maximal concentrations available were estimated and used as maximum possible uptake constraints (*Electronic Supplementary 3.1 - Additional File 3*). The uptake and export concentrations over time were used to calculate the exchange rates into $mmol.gDW^{-1}.h^{-1}$ to constrain uptakes and exports in the model (*Electronic Supplementary 3.1 - Additional File 3*). Biomass was set as the objective function and optimised using pFBA in COBRApy [47], the growth rate predicted were used to calculate the biomass generated over time and compared to the experimental data. The growth predicted by *i*AA1259 and *i*SLT1240 were in good agreement with the experimental growth curve (*Figure 3.3*), so the next step was to compare the metabolic fluxes across time between both strains. The fluxes predicted by pFBA for all the reactions at each time points were matched for the two strains using the unique BiGG reactions IDs [49]. Then, fluxes for each reaction were compared by calculating the fold difference of fluxes between both strains. Afterwards, to focus on large metabolic differences

between the two strains, we filtered the reactions with a flux fold difference > 1.5 times in one of the strains and the reactions only active in one strain (*Electronic Supplementary 3.1 - Additional File 4*). Then, the key differences identified were included on a metabolic map representing primary metabolic differences (*Figure 3.7*).

3.3.3 Proteomics data constraint and analysis

The two models *i*AA1259 and *i*SLT1240 were constrained with data from a comparative proteomics study of *S. lividans* TK24 and *S. coelicolor* M145 [20]; then, proteomics data was integrated and analysed with the models. The first set of constraints corresponded to the R2YE medium nutrients contents, as described above. The second set corresponded to the different metabolites fluxes measured in the media across the three time points 36, 48, and 72H. The exchange fluxes were calculated from the concentrations measured for glucose, proline, NO_3/NO_2, phosphate, ammonium (details available in *Electronic Supplementary 3.1 - Additional File 5*). The third set of constraints came from the proteomics data, by switching off genes (in the models) if their corresponding proteins had a median quantification of zero across the three time points. The genes were switched off by knocking-out (KO) the genes using the gene KO function in COBRAPy [47]. However, some genes were essential for growth, so these genes were not switched off and were analysed individually; see *Electronic Supplementary 3.1 - Additional File 6* for the list of genes and their analysis. There were 28 genes knocked-out in *i*SLT1240 and 11 genes knocked-out in *i*AA1259. The fluxes were predicted by pFBA with COBRApy [47] at 36, 48, and 72H; then, for both strains, the predicted fluxes were matched based on the reactions IDs. The proteomics data was matched to the predicted fluxes based on the SCO gene IDs [40] associated with the

reactions. Then, the proteomics data was compared to the predicted fluxes by comparing the fold changes and the Pearson correlation between proteomics and predicted fluxes were calculated in *R*, see *Electronic Supplementary 3.1 - Additional File 7* for the predicted flux data matched to the proteomics.

3.3.4 Analysis of the antibiotic production capabilities

First, we evaluated the antibiotic production capabilities of the metabolic networks by estimating the optimal production of ACT, CDA, CPK, and RED in the *i*AA1259 and *i*SLT1240 models under minimal growth conditions. The minimal media was composed of glucose (sole carbon source), phosphate, NH_3 (sole nitrogen source), trace elements and minerals (e.g., Ca^{2+}, Cl^{-}, K^{+}, Fe^{2+}, H_20), and O_2. The uptake of these nutrients were constrained at −1000 $mmol.gDW^{-1}.h^{-1}$. Then, the individual antibiotic productions were set as objective one by one for the model and optimized using pFBA in COBRApy [47]; the optimum value was used as the maximum possible antibiotic flux. The antibiotic fluxes were constrained one by one between 0 and the maximum possible antibiotic flux (using a continuous range), at each antibiotic flux value in the range the biomass was optimized by pFBA in COBRApy [47]. The constrained antibiotic fluxes were plotted as X and the optimized biomass as Y, this showed a curve of the trade-off between the antibiotics production and maximum biomass possible (see *Figure 8* and *Supplementary Figure 3.5*). This trade-off curve represents the competition of resources between antibiotic production and growth. As the pyruvate formate lyase (PFL) was identified as the single reaction with the most impact on production in the *i*SLT1240 model, the analysis has been applied to *i*AA1259, *i*SLT1240, *i*AA1259 with the addition of the PFL reaction, and *i*SLT1240 with the PFL reaction deleted.

Under minimal media conditions, the model was constrained to a minimum growth rate of 0.1 $gDW.h^{-1}$, which is similar to growth rates in transition phase. Then, the predicted exchange fluxes were box plotted to compare the metabolites imported/exported from the media. The predicted fluxes for the different strains were plotted against each other to compare the fluxes between the WT strains (*i*AA1259 & *i*SLT1240) and the PFL mutants (*i*AA1259+PFL, *i*SLT1240-PFL). This aimed at having an overview of the presence or absence of the PFL impacted the overall metabolism, as PFL modifies the central metabolism of the two strains.

The next step was to apply the trade-off analysis as described above but with constraints closer to industrial conditions than a minimal media. So, the constraints used came from the R2YE medium nutrients constraints and to the exchange fluxes of glucose, ammonium, and gamma-actinorhodin [11]. For the R2YE medium and exchange constraints see *Electronic Supplementary 3.1 - Additional File 3*. The constraints were applied to *i*AA1259, *i*SLT1240, *i*SLT1240 without the PFL reaction, and *i*AA1259 with the PFL reaction. The trade-off curves were plotted for all four scenarios (see *Figure 11*, and *Supplementary Figure 3.10*).

3.5 Results & Discussion

3.5.1 Comparative reconstruction of a S. lividans TK24 metabolic model

S. coelicolor A3(2) and *S. lividans* TK24 are two phylogenetically closely related strains. Their predicted protein complements are very similar (>99.7% AA identity). Consequently, the genome-scale metabolic model of *S. lividans* TK24 could be reconstructed using our published *S. coelicolor* model *i*AA1259 as template [33]. In order to reconstruct the *S. lividans* model the two genomes and the predicted proteomes

were compared, as described in *Methods*, to identify metabolic genes that were absent or only present in *S. lividans*. The reconstructed metabolic model of *S. lividans* – *i*SLT1240 – has fewer genes than *i*AA1259 but a similar number of reactions and metabolites (*Table 3.1*). The details of the reconstructed *S. lividans* model are included in the *Electronic Supplementary 3.1 - Additional File 1*.

Model Characteristics	***S. coelicolor* A3(2) iAA1259**	***S. lividans* TK24 iSLT1240**	**Added**	**Deleted**
Genes	1259	1240	10	29
Reactions	1912	1912	6	6
Metabolites	1471	1470	3	4

Table 3.1: Number of genes, reactions, and metabolites, in the reference model *i*AA1259 and the reconstructed metabolic model *i*SLT1240 for *S. lividans* TK24

3.5.2 Metabolic functions lost and gained by *S. lividans* TK24

The metabolic genes deleted or added from the reference model *i*AA1259 to generate the *i*SLT1240 model resulted in the predicted loss or gain of metabolic functions in *S. lividans* TK24 compared to *S. coelicolor* (*Table 3.2*). The *S. lividans* model lacks some important metabolic reactions involved in primary metabolism, including important reaction within gluconeogenesis and glyoxylate metabolism. Also, the presence of some genes that are absent in the *S. coelicolor* genome resulted in the putative gain of new interesting metabolic reactions, such as the PFL activity involved in anaerobic glycolysis. In many other cases, metabolic genes missing in the *S. lividans* genome do not involve a loss of metabolic reactions in the *i*SLT1240 model, as isoenzymes are also associated with the same reactions. However, this does not mean that these deletions would not have any physiological effect, as multiple genes

Missing in *S. lividans* metabolic model			
Genes	**Enzymes**	**Reaction ID in iAA1259**	**Metabolic functions**
SCO0046	Acyl-carrier-protein phosphodiesterase	ACPpds	ACP prosthetic group turnover
SCO0548	Malonyl-ACP decarboxylase	MACPD	Generation of precursor for fatty acids synthesis
SCO0982 SCO0983	Isocitrate lyase Malate Synthase	ICL MALS	Glyoxylate bypass metabolism
SCO0985	5-methyltetrahydropteroyltri glutamate-homocysteine S-methyltransferase	MHPGLUT	De novo L-methionine biosynthesis
SCO3439 SCO3440	Cuprous Oxidase (Cu+1) Ferroxidase	CU1O FERO	Oxidation of Cu^{+} to Cu^{2+} Oxidation of Fe^{2+} to Fe^{3+}
SCO3486	Lactaldehyde Dehydrogenase Glycoladehyde Dehydrogenase	LCADi GCALDD	Final steps of rhamnose and fucose degradation Glycoladehyde degradation
SCO6147	Sorbitol oxidase Xylitol oxidase	SBTO XYLTO	Assimilation of sorbitol and xylitol
Added in *S. lividans* metabolic model			
Genes	**Enzymes**	**Reaction ID in iSLT1240**	**Metabolic functions**
SLIV_RS36830	Pyruvate formate lyase	PFL	Anaerobic acetyl-CoA production
SLIV_RS36830	2-Oxobutanoate formate lyase	OBTFL	Anaerobic amino-acid metabolism
SLIV_RS36190 SSPG_01897	2-methylcitrate dehydratase 2-methylcitrate synthase	MCITD MCITS	Survival to isocitrate lyase loss. Compensation to glyoxylate bypass loss.
SLIV_RS08550	Glucose-1-phosphate guanylyltransferase	GDP1PT	GDP-glucose synthesis

Table 3.2: List of genes deleted or added to the *S. coelicolor* iAA1259 model that are abolishing or adding a metabolic function in the reconstructed *S. lividans* TK24 metabolic model

The upper part of this table (in red) shows the genes and associated reactions deleted from the *i*AA1259 template model to generate the *S. lividans* TK24 model *i*SLT1240. The lower part of this table (in blue) shows the genes and associated reactions added to the *i*SLT1240 model. These are only the genes deleting or adding a metabolic function in the model, other genes deleted or added but not modifying the model metabolic functions are in *Electronic Supplementary 3.1 - Additional File 1*. These genes directly impact the metabolic network and potentially *S. lividans* metabolism.

associated with the same reaction can be differentially expressed under different conditions. This could result in a loss of function under specific condition, but experimental data would be required to test this.

A previous comparative genomic study between the two strains identified the absence of large *S. coelicolor* genomic islands in *S. lividans* [38]; as expected, multiple metabolic genes deleted from the model were localized in these islands (*Supplementary Table 3.1*).

The gene SCO0046 (*acpD*/*azoR*) is a putative acyl-carrier-protein (ACP) phosphodiesterase involved in the ACP prosthetic group turnover, by hydrolysis of a holo-ACP to an apo-ACP. The knock-out of this gene in *Escherichia coli* has shown to stop the ACP turnover but it was non-essential to growth [50]. In the *S. coelicolor* model this enzyme is involved in the production of the secondary metabolites CDA, RED, ACT, and CPK, together with the holo-ACP generation for important biological functions such as fatty-acids biosynthesis; however, it is not predicted as essential for growth by the model as other reactions are able to produce holo-ACP. *S. lividans* has also lost a malonyl-ACP decarboxylase (encoded by SCO0548) responsible for the decarboxylation of malonyl-ACP in acetyl-ACP, which is then used for straight-chain fatty acids (SCFAs) biosynthesis [51]. However, the acetyl-ACP is still produced, as an acetyl-CoA ACP transacyclase produces it by using an acetyl-CoA and a holo-ACP. Furthermore, it has been shown that multiple enzymes primarily involved in other reactions (e.g., acetoacetyl-ACP synthases) also have a malonyl-ACP decarboxylase activity [51–53]. Still, the loss of this enzyme is likely to change the balance between SCFAs and branched-chain fatty acids (BCFAs). It might also impact polyketides (PKS) biosynthesis as malonyl-ACP is a precursor of aromatic PKSs such as ACT and CPK [54]. This change could

promote acetyl-CoA consumption to produce acetyl-ACP instead of using malonyl-ACP, which might impact the available pool of acetyl-CoA (and consequently malonyl-CoA), another key precursor to produce secondary metabolites such as ACT, CDA, yCPK, or RED. It is likely that the absence of both SCO0046 and SCO0548 increases BCFAs biosynthesis and decreases secondary metabolites production in *S. lividans*. The ACP prosthetic group turnover is abolished, and the main precursor of SCFAs, acetyl-ACP, is generated from acetyl-CoA instead of malonyl-ACP. However, this needs to be investigated experimentally in *S. coelicolor* (e.g., by gene knock-out of SCO0046 and SCO0548).

The *S. lividans* genome does not contain *aceA* (SCO0982) and *aceB* (SCO0983) genes, encoding for an isocitrate lyase and a malate synthase, respectively, both part of the glyoxylate bypass metabolism. The glyoxylate bypass metabolism loss in *S. lividans* was previously described in a comparative genomic study of *S. coelicolor* and *S. lividans* [34]. This pathway bypasses the decarboxylation steps in the tricarboxylic acid (TCA) cycle to generate succinate from acetyl-CoA. The pathway is important for the utilisation of alternative carbon sources such as C_2 fatty acids or acetate during gluconeogenesis. As alternative pathways were previously identified in *Streptomyces* spp. the glyoxylate bypass absence in *S. lividans* does not necessarily block the utilisation of acetate and other C_2 fatty acids [55]. However, the absence of this mechanism should reduce the cellular redox potential in *S. lividans* compared to *S. coelicolor* because it must systematically use the full TCA cycle, generating more reduced electron transporters (i.e., NADH).

Bypassing the last part of the TCA cycle conserves some carbons for gluconeogenesis, which will reduce electron flux into respiration as it does not

generate reduced electron carriers (i.e., NAD(P)H and $FADH_2$) [36,56]. A recent study showed that *S. coelicolor* had a stronger oxidative phosphorylation and a higher gluconeogenesis than *S. lividans* [11,20]. The loss of the glyoxylate bypass in *S. lividans* might cause this difference in redox balance between the two strains. In *S. coelicolor*, when the oxidative stress is high in the respiration chain, the strain could redirect the fluxes toward the glyoxylate bypass to generate less reduced electron transporters. A bioinformatics study of bacterial genomes indicated that only microorganisms with an aerobic metabolism possess a glyoxylate bypass pathway [36]. The glyoxylate metabolism has been identified as an important pathway to tolerate oxidative stress [36,57], so its absence in *S. lividans* could increase its oxidative stress. This will be explored in more detail in the later sections of this chapter.

The gene *metE* (SCO0985), missing from *S. lividans,* encodes a cobalamin-independent methionine synthase catalysing the transfer of a methyl from a 5-methyltetrahydrofolate-glutamate to a homocysteine. This is the last step of the *de novo* L-methionine biosynthesis. The *metE*-encoded methionine synthase enables methionine production in the absence of exogenous cobalamin [58], and if this pathway is disrupted it would make *S. lividans* a methionine auxotroph [59]. Despite this gene missing, the *de novo* synthesis of methionine is not completely abolished, as a *metH* methionine synthase (encoded by SCO1657) and the endogenous cobalamin biosynthesis pathway – cobalamin is necessary for the *metH* methionine synthase activity [60] – are conserved in the *S. lividans* genome. Among other genes lost, SCO3439 and SCO3440 are encoding for two putative multi-copper oxidases with a ferroxidase activity. These enzymes are involved in the turnover of Fe(III) to Fe(II) to support iron uptake, as well as oxidising Cu(I) to Cu(II) contributing to copper resistance [61,62]. The

loss of these oxidases could result in a disturbance in copper and iron homeostasis. Otherwise, the gene SCO3486, absent from *S. lividans*, corresponds to an aldehyde dehydrogenase with a lactaldehyde and glycolaldehyde dehydrogenase activity. The lactaldehyde dehydrogenase is involved in the oxidative pathway of methyl-pentose such as rhamnose, which is degraded to an L-lactaldehyde and a dihydroxyacetone phosphate (DHAP) [63]. The lactaldehyde can be oxidized to lactate (using NAD^+ as a cofactor), then converted into pyruvate by a lactate dehydrogenase to enter central metabolism [64]. The other enzymatic activity of glycolaldehyde dehydrogenase oxidizes glycolaldehyde in glycolate, which can enter the glyoxylate metabolism [65]. Due to the absence of this lactaldehyde dehydrogenase, *S. lividans* should be unable to use rhamnose as a carbon source, which is predicted by the metabolic model: *S. lividans* cannot grow if it uses rhamnose as a sole carbon source, whereas *S. coelicolor* grows on rhamnose (*Figure 3.2*). Another enzyme missing from *S. lividans* is the alditol oxydase encoded by SCO6147, it uses xylitol or sorbitol as substrate [66]. This enzyme oxidises xylitol in xylulose which can be phosphorylated to enter the pentose phosphate pathway. The same enzyme can also oxidise sorbitol in fructose that can then enter central metabolism through fructolysis. This loss is possibly impacting the ability of *S. lividans* to use xylitol and sorbitol as carbon sources. The model predicts that *S. lividans* cannot grow if xylitol and/or sorbitol are the sole carbon sources (*Figure 3.2*). This is a very strong prediction that could easily be tested experimentally.

Among the genes unique to *S. lividans,* the strain seems to have acquired a few metabolic genes contributing to the difference in metabolic capabilities. For example, *S. lividans* gained a putative formate lyase (PFL), with a pyruvate formate and 2-oxobutanoate lyase activity [67]. During anaerobic glycolysis, a pyruvate formate lyase

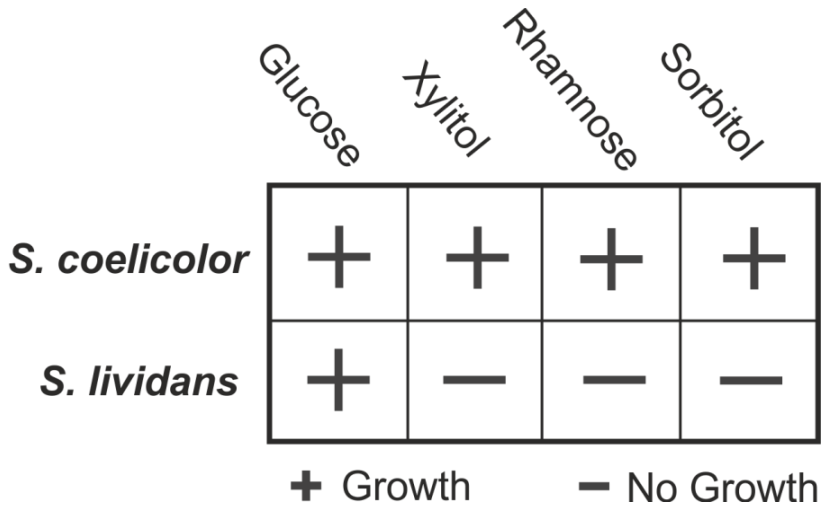


Figure 3.2: Growth predictions for *S. coelicolor* and *S. lividans* with different carbon sources

As the absence of genes in *S. lividans* suggested that the strain could not grow in some specific carbon sources. The growth was predicted using different unique carbon sources with the iAA1259 (*S. coelicolor*) and iSLT24 (*S. lividans*) genome-scale metabolic models. While *S. coelicolor* was predicted to be able to grow on glucose, xylitol, rhamnose, or sorbitol as sole carbon sources, this was only possible on glucose for *S. lividans*. The plus represents predicted growth and the minus represents the absence of predicted growth.

converts a pyruvate and a CoA in a formate and an acetyl-CoA to replenish the TCA cycle. This enzyme is a central enzyme for the anaerobic glucose metabolism in facultative aerobes such as *E. coli* [37]. The 2-oxobutanoate lyase activity is transforming 2-oxobutanoate (or α-ketobutyrate) and a CoA into a formate and a propionyl-CoA, which can be converted in methylmalonyl-CoA, then converted in a succinyl-CoA to enter the TCA cycle. The 2-ketobutyrate is associated with amino-acids metabolisms, such as degradation of threonine and aspartate, isoleucine biosynthesis, as well as cysteine and methionine metabolism [68]. A bioinformatics analysis determined that PFL and the PFL activating enzymes were only present in *S. lividans* genome. The gene and protein sequences of the PFL and its activating enzyme (from *E. coli* and *Mycobacterium tuberculosis*) did align to two genes in *S. lividans* but not in *S. coelicolor*

(using 3 different approaches: NCBI Blast, OrthoDB, and Ensembl), this data is available and described in *Supplementary Document 1 (see 3.8.1)*.

S. lividans seems to have acquired a 2-methylcitrate synthase that transforms an oxaloacetate and a propionyl-CoA in a 2-methylcitrate and a CoA [69]. As well as a 2-methylcitrate dehydrogenase transforming the 2-methylcitrate in a 2-methylaconitate. This pathway is involved in anaerobic propionate degradation and has been identified in multiple prokaryote organisms before [69,70]. It is likely that other citrate synthases (found in *S. coelicolor* and *S. lividans*) also present a low 2-methylcitrate synthase activity [69], which would suggest that , to a lesser extent, the function could also be present in *S. coelicolor*.

Finally, *S. lividans* seems to have a glucose-1-phosphate guanylyltransferase; this enzyme catalyses the biosynthesis of GDP-glucose from glucose-1-phosphate and GTP (releasing an inositol-diphosphate). Another route exists to produce GDP-glucose and glucose-6-phosphate from trehalose-6-phosphate and GDP by a GDP-glucose-glucose-phosphate glucosyltransferase [24].

Another interesting feature is the possible duplication of the phosphate transport operon (SCO6814-16) in *S. coelicolor* (*Supplementary Table 3.1*). As well as the presence of 4 nitrate reductases operons in *S. coelicolor* while *S. lividans* only have 3 operons (SCO4947-50 is absent), see *Supplementary Table 3.1*. Theses operons duplications in *S. coelicolor* correlate with a higher oxidative metabolism in this strain compared to *S. lividans* [11]; as *S. coelicolor* may consume more phosphate for oxidative phosphorylation and have a more active nitrate oxidative respiration.

The first step of comparative reconstruction of the *S. lividans* metabolic model, *i*SLT1240, from the *S. coelicolor* model *i*AA1259, already highlighted some interesting

metabolic differences. For example, *S. lividans* seems to have lost the ability to use sorbitol, xylitol, and rhamnose as carbon sources, which was validated by the model predictions (*Figure 3.2*). Also, *S. lividans* might be more adapted to anaerobic conditions while *S. coelicolor* might be more adapted to aerobic conditions. For example, the absence of the glyoxylate bypass metabolism (SCO0982 and SCO0983) in *S. lividans* makes the strain less resistant to high oxidative stress (i.e., aerobic respiration). Furthermore, *S. lividans* acquired a pyruvate formate lyase, which is regulated by an activating enzyme sensitive to O_2 levels and only active under anaerobic condition [67]. Multiple studies showed that secondary metabolism is directly impacted by oxidative metabolism in *Streptomyces* [c,19,20,71], so these differences are likely to change antibiotic production. Additional interesting features were highlighted in *S. lividans* fatty acid metabolism, with the loss of an ACP phosphodiesterase (SCO0046) and a malonyl-ACP decarboxylase (SCO0548), which might reduce the availability of ACP and acetyl-CoA for secondary metabolites such as ACT and RED. The metabolic variations potentially responsible for the lower production of antibiotics in *S. lividans* will be explored in the next sections by combining metabolic modelling with omics data integration and analysis.

3.5.3 Initial validation of the *Streptomyces lividans* metabolic model

Before using predictions of the *S. lividans* GSMM (*i*SLT1240), it was necessary to verify the validity of the predictions. We compared the predicted bacterial growth to experimental data, as previously done for the validation of the *S. coelicolor* model *i*AA1259 [33]. The model was constrained using ammonium and glucose uptake rates measured in a chemostat experiment using *S. lividans* TK24 [35] (*Figure 3.3*). The biomass

predicted with the *i*SLT1240 model is in good agreement with the observed biomass (*Figure 3.3*). Glucose is the main source of carbon in this experiment, but it is also a limiting nutrient, the depletion of which seems to trigger slower growth and entry in transition phase (purple curve in *Figure 3.3*). Once this validation was done, for each new prediction using a given dataset under a specific condition, the biomass was predicted and compared to the experimental observations to verify the validity of these predictions.

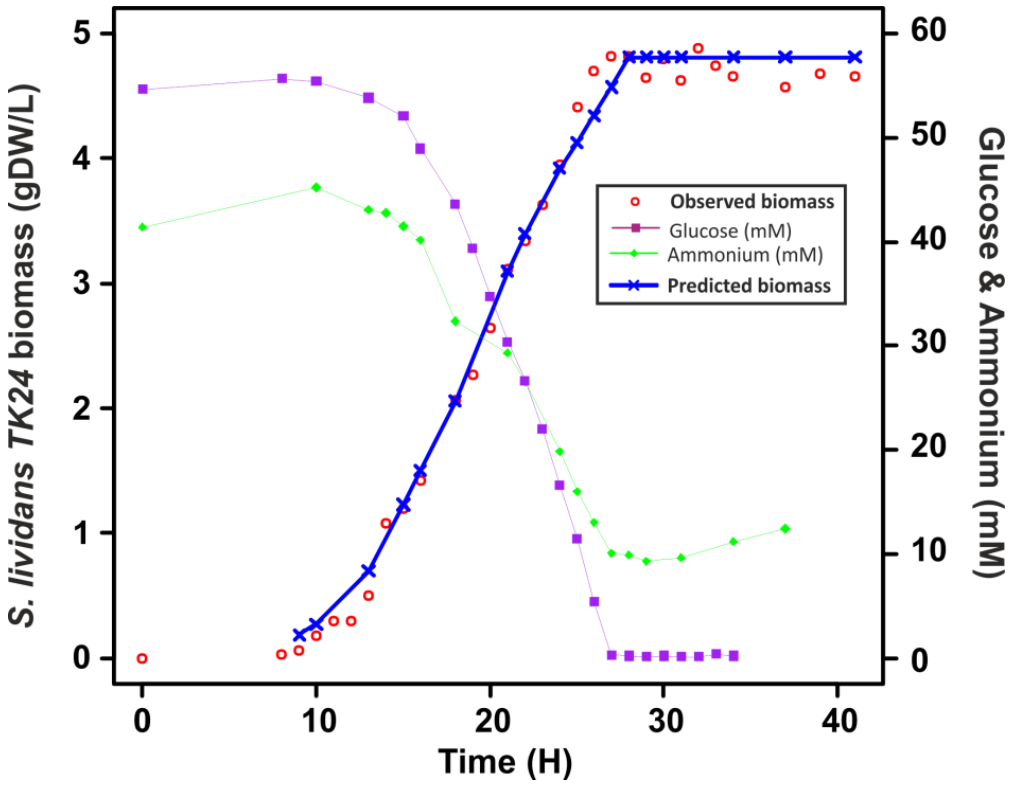


Figure 3.3: Validation of the growth predictions of the i*SLT1240* metabolic model

The *S. lividans* metabolic model - *i*SLT1240 - was constrained based on the minimal medium data and using glucose (purple curve) and ammonium (green curve) uptake fluxes from a chemostat experiment at each time points [35]. The *i*SLT1240 metabolic model predictions are valid, as the predicted growth (blue curve) is close to the experimental values (red points). The organism enters stationary phase at 29h due to glucose depletion at 28h. The model correctly predicts the growth switch around 29h, matching the experimental observations.

3.5.4 Comparative analysis of *S. coelicolor* and *S. lividans* metabolism

The metabolic differences between the two strains were predicted by comparing the predicted flux distribution across the different metabolic pathways using constraint-based metabolic modelling. The model was constrained using experimental data from a study carried out in R2YE solid media under Pi-limited conditions for both *S. coelicolor* M145 and *S. lividans* TK24 [11]. At first, the possible exchanges were constrained based on the R2YE media composition [11,48]. Then, the metabolite uptake (e.g., glucose and phosphate) and the export flux (e.g., actinorhodin) were also constrained (*Figure 3.4*). Finally, the biomass was constrained using the levels of triacylglycerol (TAG) and ATP to ADP ratios as these were measured in *S. lividans* compared to *S. coelicolor*[11] (details of the constraints in *Electronic Supplementary – Additional File 3*). The exchange fluxe constraints were applied at different time points across the growth curve of the organisms, with the two first points in exponential phases and the remaining ones in transition and stationary phase. The predicted growth across time was verified by the experimental growth under these conditions for both the *S. coelicolor* model *i*AA1259 (*Figure 3.4a*) and the *S. lividans* model *i*SLT1240 (*Figure 3.4b*).

The metabolic prediction comparison for both strains shows that about 1175 reactions have zero fluxes in both strains; this represents 61% of the metabolic network reactions. A lot of reactions are switched-off in the network as those were probably not used under this specific condition, but it could be used under a different condition (e.g., different media). In the 39% of non-zero fluxe reactions in either strains (745 reactions), some reactions were excluded from the detailed analysis: 49

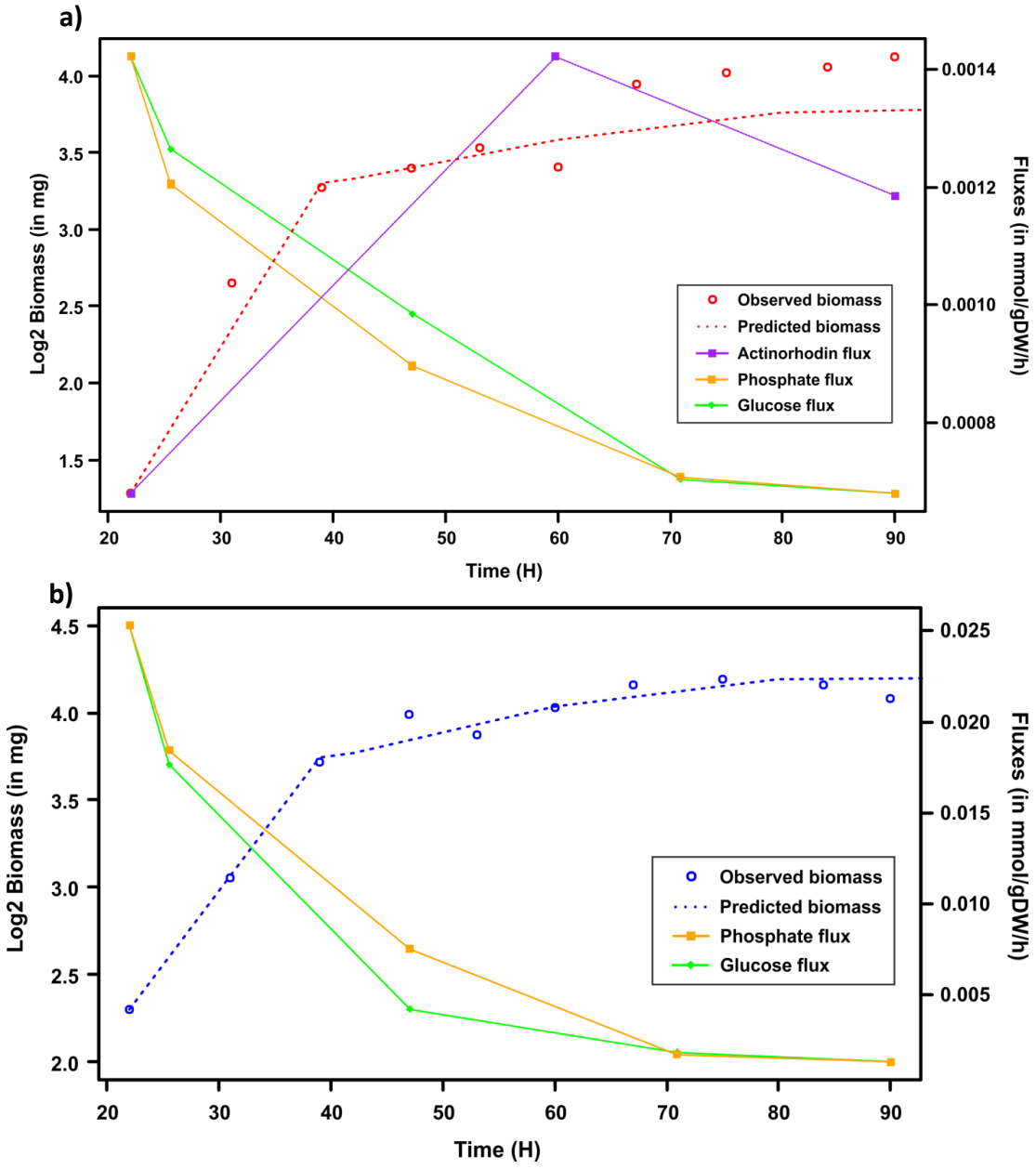


Figure 3.4: Validation of the *S. coelicolor* and the *S. lividans* metabolic model's growth predictions with constraints from the dataset used

a) Biomass predictions and dynamic constraints used for *S. coelicolor*

The *S. coelicolor i*AA1259 metabolic model was constrained across multiple time points with the glucose and phosphate (orange and green lines) uptake, and the actinorhodin export (purple line), in the same minimal media conditions used for the experiment [11]. The biomass predictions (red dashed line) are in good agreement with the experimental biomass data (red dots).

b) Biomass predictions and dynamic constraints used for *S.lividans*

The *S. lividans i*SLT1240 metabolic model had the same method applied, unless actinorhodin constraint as it was not producing it. The biomass predicted (blue dashed line) here as well is in good agreement with the experimental data (blue dots).

active exchange reactions, 40 active transporters, 40 reactions with almost zero fluxes (<0.00051 fmol/gDW/h), and the reaction corresponding to ATP maintenance.

Finally, 612 intracellular reactions were compared and analysed in detail (*Electronic Supplementary – Additional File 4*). Twenty-four reactions had fluxes only active in *S. coelicolor* with 19 involved in actinorhodin biosynthesis (only this strain produced ACT), and 30 reactions with fluxes at least 1.5 times higher in *S. coelicolor*. Seven reactions were active only in *S. lividans*, and 204 reactions carried fluxes at least 1.5 times higher in *S. lividans* (including 182 reactions involved in 9 different straight and branched chain fatty acid biosynthesis pathways). From these 204 reactions, 22 reactions were left with 9 reactions involved in lipids precursors biosynthesis and 6 reactions involved in isoleucine and valine degradation for branched-chain fatty acid biosynthesis. The predicted fluxes only present or higher in *S. coelicolor* (*Figure 3.5*) and those specific or higher for *S. lividans* (*Figure 3.6*) were analysed in more detail to identify and understand the main metabolic differences between the two strains.

A large part of the fatty acids biosynthesis pathways was higher in *S. lividans*, with 182 reactions involved in the direct production of fatty acids, 9 reactions involved in precursors generation, and 6 reactions involved in the degradation of isoleucine and valine to produce branched-chain fatty acids. This prediction was probably due to the higher production of triacylglycerol in *S. lividans* that was included as a constraint in the *i*SLT1240 biomass reaction. The *i*SLT1240 model also predicted a higher glycolytic metabolism in *S. lividans,* with 7 reactions involved in glycolysis higher than in *S. coelicolor* (*Figure 3.6*). This was experimentally observed under the same conditions, indicated by the higher accumulation of triacylglycerol and higher glucose uptake in

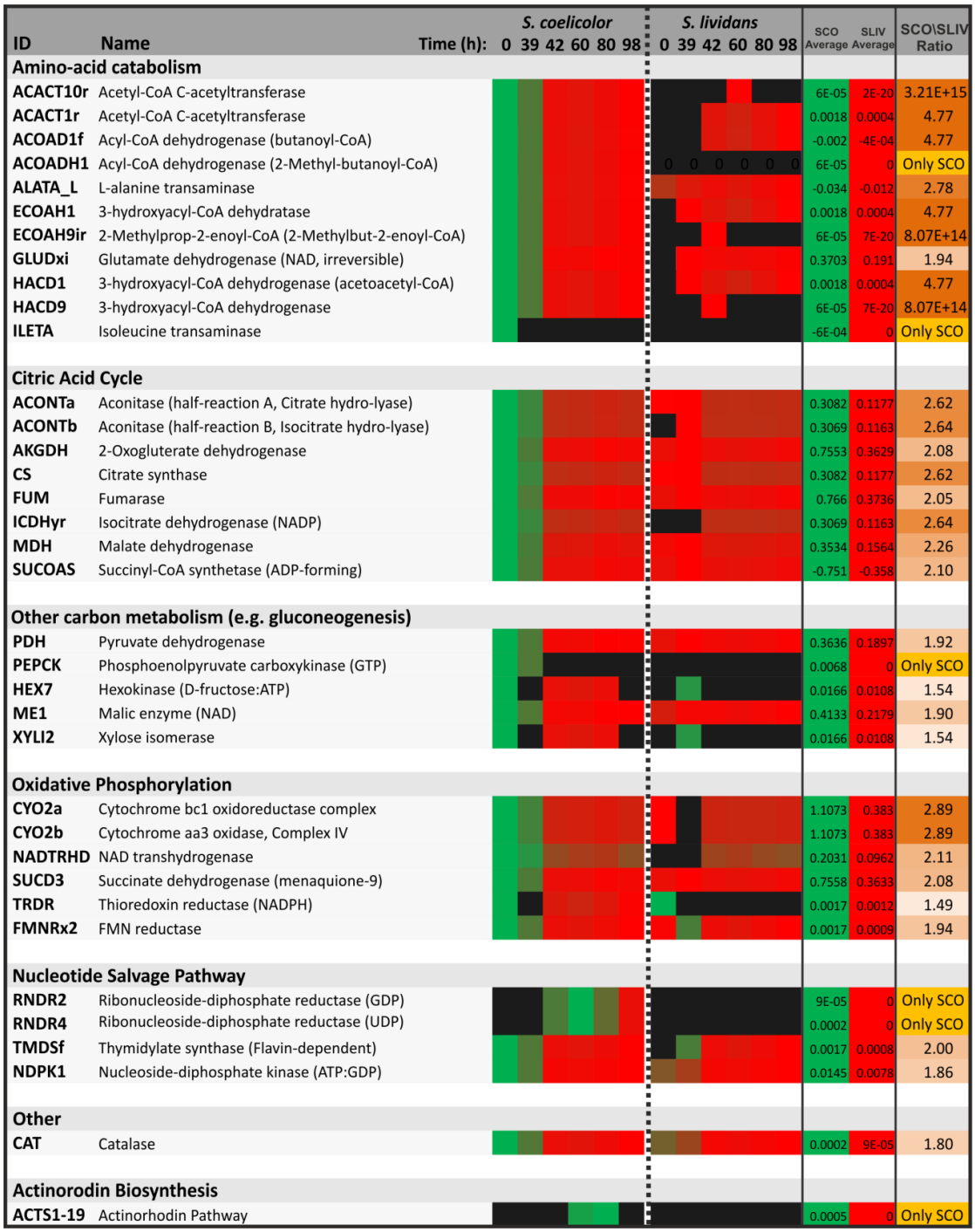


Figure 3.5: Predicted fluxes with a higher value in *S. coelicolor*

The metabolic fluxes predicted as higher or only active in *S. coelicolor* are represented as a heatmap (relative high flux in green, low flux in red, and null flux in black). 24 reactions are predicted as only active in *S. coelicolor* (i.e., only SCO), and 30 reactions with a predicted flux at least 1.5 times higher in *S. coelicolor* than in *S. lividans*. The reactions are ordered by pathways or functions associated. The heatmap of *S. coelicolor* to *S. lividans* flux ratio (i.e., SCO:SLIV ratio) shows higher ratio in dark orange and lower ratio in light orange. The full dataset is available in *Electronic Supplementary – Additional File 4*.

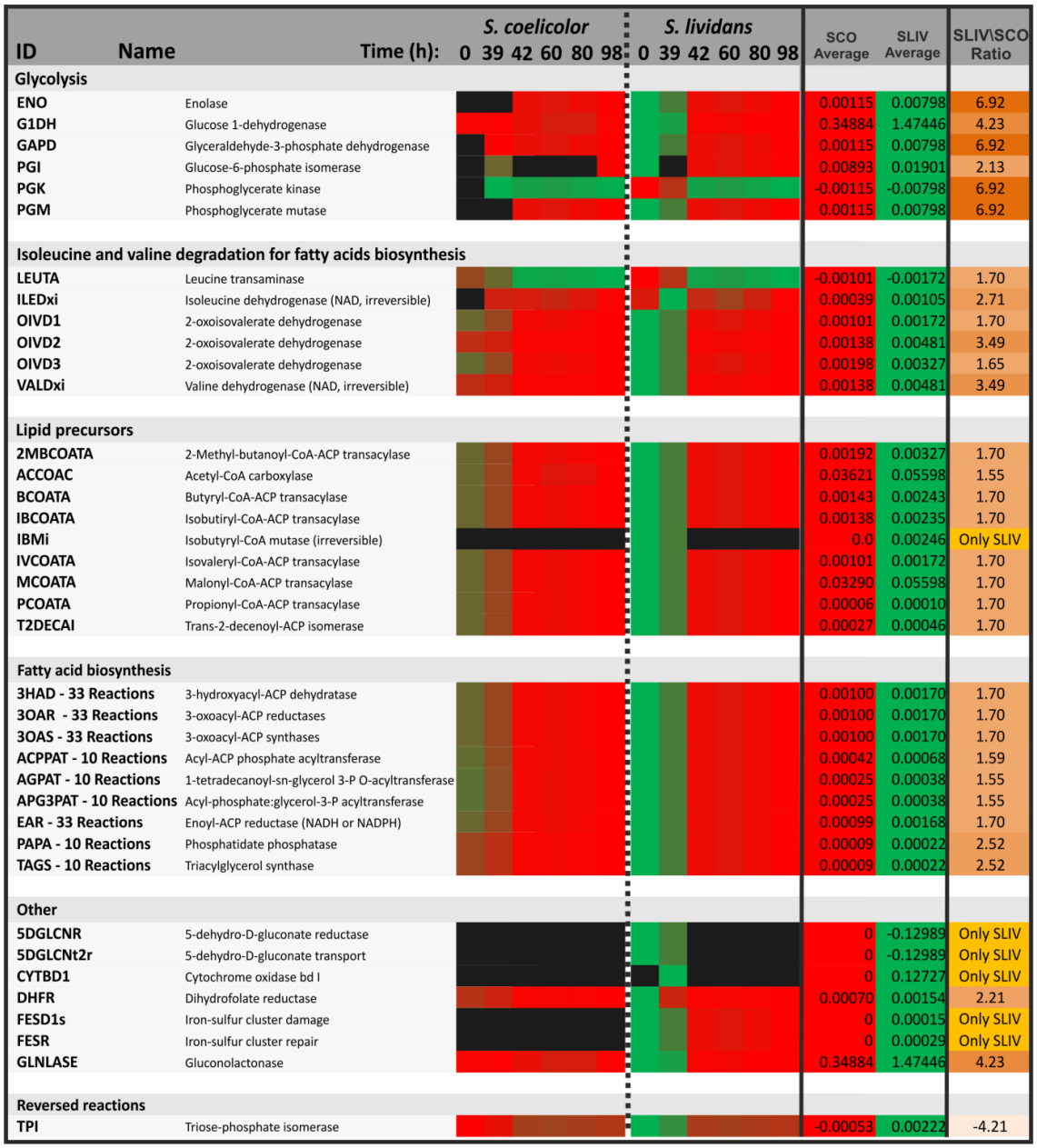


Figure 3.6: Predicted fluxes with a higher value in *S. lividans*

The metabolic fluxes predicted as only active or higher in *S. lividans* are represented as a heatmap (relative high flux in green, low flux in red, and null flux in black). 7 reactions are only active in *S. lividans* (i.e., only SLIV), 204 reactions have predicted fluxes at least 1.5 times higher in *S. lividans*, The heatmap of the *S. coelicolor* to *S. lividans* flux ratio (i.e., SCO:SLIV ratio) shows higher ratio in dark orange and lower ratio in light orange. These results are analysed in detail in the manuscript, and the full dataset is available in *Electronic Supplementary – Additional File 4*.

S. lividans [11,20]. The higher glycolysis prediction was apparently caused by the higher glucose uptake constrain in the *i*SLT1240 model.

Predictions suggest that *S. coelicolor* has a stronger amino-acid catabolism, with 11 reactions involved in amino-acid degradation. This includes multiple reactions in entry points of the central metabolism, such as the acetyl-CoA C-acetyltransferase (EC 2.3.1.9) that catalyses an intermediate metabolite of the valine, leucine, and isoleucine degradation, the 2-methyl-3-acetoacetyl-CoA into a propanoyl-CoA and an acetyl-CoA (*Figure 3.5*) [72]. This was observed in proteomics data where *S. coelicolor* had a higher relative level of proteins associated with amino-acid catabolism compared to *S. lividans* [20], while another study showed that *S. coelicolor* is more efficient at degrading glutamate than glucose [73].

The model metabolic predictions show higher average fluxes through the TCA cycle in *S. coelicolor,* with 8 reactions at least two times higher. However, in stationary phase (60 to 98h) the TCA cycle fluxes are predicted as higher in *S. lividans* than in *S. coelicolor*, this corresponds to the start of actinorhodin production in *S. coelicolor* (data in *Electronic Supplementary – Additional File 4*). Acetyl-CoA is the main building block for the ACT pathway and an entry point of glycolysis toward the TCA cycle as well as TAG biosynthesis, so in *S. coelicolor* the fluxes using this precursor are split between ACT production, the TCA cycle, and TAG biosynthesis. Whereas *S. lividans* does not produce ACT, so acetyl-CoA fluxes go toward the TCA cycle more than in *S. coelicolor*. For *S. coelicolor*, the metabolic modelling predicted higher fluxes in pathways used instead of glycolysis to generate TCA cycle precursors, such as the amino-acids catabolism and the anaplerotic metabolism (PEPCK and ME1) (*Figure 3.5*). The anaplerotic reactions node involves phosphoenolpyruvate, pyruvate, and oxaloacetate

as substrate and plays a role of connection between the glycolysis, the gluconeogenesis, and the TCA cycle. *S. coelicolor* also has a higher gluconeogenesis [20], which consumes precursors coming from the anaplerotic pathway that could explain the increased fluxes for the anaplerotic reactions. The phosphoenolpyruvate carboxykinase (PEPCK in *Figure 3.5*) carries higher predicted fluxes in *S. coelicolor*; however, gluconeogenesis does not seem fully active in the *i*AA1259 model predictions as some of the key enzymes (i.e., glucose 6-phosphatase) do not carry fluxes (*Electronic Supplementary – Additional File 4*). An interesting prediction is the triose-phosphate isomerase (TPI) reaction with fluxes predicted in the direction of the dihydroxyacetone phosphate (DHAP) conversion to glyceraldehyde-3-phosphate (G3P) in *S. lividans* and in the opposite direction in *S. coelicolor (Figure 3.6)*. However, this difference is observed only during the exponential phase (from 0 to 42h), once in transition and stationary phase *S. coelicolor* fluxes are only in the forward direction. The TPI is a major point of regulation for central metabolism as it is a convergence point for the glycolysis, the gluconeogenesis, and the pentose phosphate pathway [74]. This enzyme regulation is also coordinated by oxidative energy [75], but the metabolic model does not take into account enzymatic regulations so the flux direction difference has to be caused by a metabolic difference. Here, during the exponential phase, the TPI fluxes go toward glycolysis in *S. lividans* and toward gluconeogenesis in *S. coelicolor*, which agrees with experimental observations [11,20]. Nevertheless, the model is unlikely to predict fluxes toward (full) gluconeogenesis, as producing glucose does not lead to optimal biomass (the objective of the model). So, it seems like the reactions involved in the other carbon metabolism (*Figure 3.5*), such as the phosphoenolpyruvate carboxykinase discussed earlier, generates intermediates for the

central metabolism (e.g., TCA cycle), electron transporters (e.g., NADH), and ADP for the oxidative respiration. It seems as if the higher need of ATP by *S. coelicolor* is constrained by the ATP needed in the biomass reaction, which forces the production of more ADP (to generate more ATP) and reduced electron transporters (e.g., NADH). This would explain the higher fluxes toward gluconeogenesis and other carbon metabolism (*Figure 3.5*) as many of these reactions generate ADP and reduced electron transporters needed by the respiration chain to produce more ATP. The higher need of ADP for oxidative phosphorylation also enforces higher fluxes in the nucleotide salvage pathway, which is used in the turnover of cofactors such as GDP, AMP, UDP (*Figure 3.5*). As expected, the *S. coelicolor* model predicts higher fluxes for oxidative metabolism in *S. coelicolor* (with 5 reactions being significantly higher, *Figure 3.5*). The differences in biomass constraints (e.g., ATP) impacted the whole metabolism of the two strains, as the differences in ATP and nutrients uptake modifies the balance of fluxes generating the reduced and oxidized cofactors necessary to reach an optimal growth. Thus, the modelling predicts that the effect of the difference in oxidative phosphorylation (input as ATP ratios) between the two strains impacts the whole metabolism. The balance of oxidised and reduced electron transporters forces fluxes in different metabolic pathways in the two strains, which could be linked to the regulation of these pathways.

Overall these predictions are in very good agreement with experimental observations: *S. coelicolor* has a lower glycolysis, higher gluconeogenesis, higher amino-acid catabolism, lower triacylglycerol anabolism, and a higher oxidative metabolism than *S. lividans* [11,20] (*Figure 3.7*). The impact of the differences in the oxidative balance (constrained) between the two strains supports that the redox

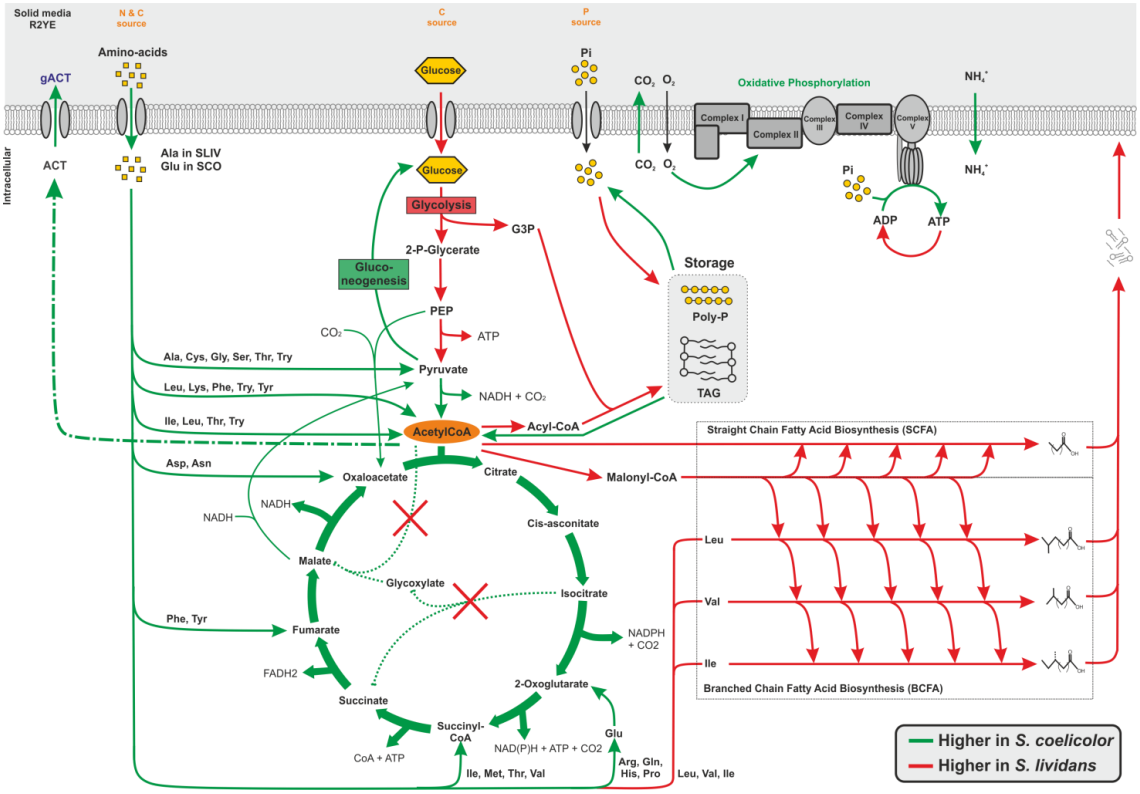


Figure 3.7: Summary metabolic map of the main metabolic differences predicted between *S. coelicolor* and *S. lividans*

This metabolic map summarizes the major predicted metabolic differences in *S. coelicolor* compared to *S. lividans* under the conditions studied. The fluxes higher in S. coelicolor are in green and the ones lower in *S. coelicolor* are in red. The reactions crossed in red correspond to the glyoxylate pathway which is absent in *S. lividans*. The dotted lines are the reactions carrying fluxes only in *S. coelicolor*. *S. lividans* is predicted to have a higher glycolytic metabolism, whereas *S. coelicolor* seems to have a higher gluconeogenesis. The amino-acid catabolism is overall higher in *S. coelicolor*, while the anabolism of branched and straight-chain fatty acids biosynthesis from amino-acid and malonyl/acetyl-CoA is higher in *S. lividans*. The oxidative phosphorylation seems to be overall higher in *S. coelicolor*, which is probably due to the higher ATP/ADP ratio constraint. These agree with experimental observations [11,20].

balance underlies the metabolic differences. This could be caused by the loss of the glyoxylate bypass metabolism that plays an important role in the resistance to oxidative stress.

3.5.5 Integrative analysis of comparative proteomics data

To better understand the metabolic differences between the two strains, and to validate the metabolic predictions, a comparative proteomics dataset [20] was integrated

and analysed with the two metabolic models. The first step was to constrain the metabolic models for both strains with available flux exchanges with the media used for: glucose, proline, ammonia (NH_4), nitrate (NO_3), nitrite (NO_2), and phosphate. Then, the biomass was predicted for both strains and compared with experimental data. In both cases, the biomass predictions were in agreement with the experimental data (*Supplementary Figure 3.3*).

The metabolic models *i*AA1259 and *i*SLT1240 were also constrained using the proteomics data at different time points: 36, 48, and 72H. The proteins with a null concentration were switched OFF in the models (as described in *Methods*). There were 611 annotated proteins with proteomics data for the *i*SLT1240 model and 615 for the *i*AA1259 model (fewer than 50% of the metabolic genes annotated). In the *i*SLT1240 model 53 genes were switched OFF, and in *i*AA1259 model 26 genes were switched OFF. The *S. lividans* model predicted that 10 genes out of the 53 were essential for growth and 2 genes reduced biomass. While, in *S. coelicolor* 3 genes out of 26 were essential and 2 genes reduced biomass.

In multiple cases the proteomics did not influence the model, when switching off proteins not produced – but with isoenzymes – in the model, will not switch off the reaction (*Supplementary Figure 3.2a*). Some of the proteins switched off in the model correspond to zero flux reactions, which will not impact the metabolic network (*Supplementary Figure 3.2b*). In the end, the few proteomic constraints that forced flux redirection resulted in a reduced predicted growth rate (*Supplementary Figure 3.2c*), which did not "break" the model in contrast to the effect of switching off the essential reactions (*Supplementary Figure 3.2d*). The essentiality of the proteins with a null concentration could have multiple explanations; isoenzymes encoded in the genome

but not annotated in model, or compensatory metabolic pathways absent from the models. Also, as the media is complex, not all the metabolites present are known, so there might be metabolites present in the media but not included in the constraints (for details on the constraints see *Method*). From the proteome constraints, excluding essential genes only 2 genes in *S. coelicolor* (SCO2151 and SCO4564, respectively encoding a cytochrome aa3 oxidase and an NADH dehydrogenase) and 3 genes in *S. lividans* (SCO2151 encoding a cytochrome aa3 oxidase, SCO2828 and SCO2831 both encoding an arginine ABC transporter) had a significant impact on the model (see *Electronic Supplementary – Additional File 6*). After applying the proteomics, the media constraints, and the metabolites uptake or export constraints (see *Electronic Supplementary – Additional File 5*) the flux predictions did not correlate with the proteomics data (r=-0.05, *Supplementary Figures 3.3 and 3.4*). The samples used for the proteomics data and the constraints grew in O_2 limited conditions (1% dO_2), and the data points were at stationary phase [20] (see *Electronic Supplementary – Additional File 7*). As the two *Streptomyces* strains grew in a complex media under O_2 and phosphate limited conditions, it is likely to have triggered stress responses directly influencing their metabolism (e.g., nitrogen respiration) [5,76]. The prediction accuracies were inaccurate as regulations changed the metabolism, and the models are probably lacking alternative metabolic pathways used under these conditions (e.g., nitrogen respiration). Instead, the model was used as a framework to analyse the proteomics data in the metabolic context and to identify key metabolic differences between the two strains.

The proteomics data constraints and integrated analysis helped to interpret the change in proteome related to the metabolic change. In the *S. coelicolor* proteomics

data constraints, 3 proteins were predicted as essential in the *i*AA1259 model: a myo-inositol phosphatase (SCO5860) involved in glycerophospholipids precursor biosynthesis, and 2 proteins associated to adenosylcobalamide biosynthesis (SCO1852 and SCO2173). In *S. lividans*, 10 proteins were identified as essential in the *i*SLT1240 model, including SCO1852 involved in cobalamin biosynthesis. Among the predicted essential genes switched off in *S. lividans* we can find: a cardiolipin synthase (SCO1389) essential for membrane biosynthesis, an alanine racemase (SCO4745) involved in cell wall biosynthesis (e.g., murein). As well as proteins involved in essential cofactors and vitamins biosynthesis, such as a pyridoxal 5-phosphate synthase (SCO1522), a kinase for menaquinones biosynthesis (SCO3148), a chorismate synthase (SCO3851) from the hydrofolates metabolism, a synthase (SCO4204) for the mycothiol biosynthesis, a threalose synthase (SCO4290), and an aerobic protoporphyrinogen oxidase (SCO6041) essential for protoheme biosynthesis (see *Electronic Supplementary – Additional File 6*). The proteomics measurements under hypoxic conditions and at stationary growth show many proteins involved in aerobic processes or in biomass production (e.g., cell wall biosynthesis) as not translated in *S. coelicolor* or *S. lividans*.

In both strains, a cytochrome aa3 oxidase (SCO2151) when switched OFF, decreased the predicted biomass. This cytochrome is favourably expressed under high oxygen concentration, so it may be switched off because of the experimental hypoxic conditions [77]. The *S. coelicolor* predicted biomass decreased due to the deletion of an NADH dehydrogenase (SCO4564), which is involved in oxidative metabolism. In *S. lividans* a different set of redundant NADH dehydrogenases genes (SCO4605 and SCO7319) were switched OFF but did not impact growth. In *S. lividans* other enzymes involved in oxidative metabolism were not produced such as a thioredoxin reductase

(SCO7298) and a pyruvate oxidase (SCO7412). Hypoxic conditions upregulate and downregulate key genes from the respiratory chain in *S. coelicolor* [78], but there are different and more genes downregulated in *S. lividans*. This is in accordance with the proteomics data interpretation by Millan-Oropez et al., concluding that oxidative metabolism was higher in *S .coelicolor* [20].

Other proteins switched OFF were involved in Coelimycin P1 biosynthesis and export (SCO6273-75, and SCO6278), indicating that Coelimycin P1 is not produced under these conditions in both strains (see *Electronic Supplementary – Additional File 6*). Also, S. lividans did not produce biosynthetic enzymes associated with the CDA (SCO3230-31, SCO3236, and SCO3249) but *S. coelicolor* did; this indicates that the production of CDA would be abolished in *S.lividans* but not in *S. coelicolor* under these conditions. See *Electronic Supplementary – Additional File 6* to find the different genes switched OFF in the two strains metabolic models based on the proteomics data.

In the end, the models did not predict accurately the differences due to the limited O_2 conditions, the low growth (in stationary phase), and in a complex media. Despite inaccurate flux predictions, using the model as a framework to analyse proteomics data helped to identify a few metabolic differences between the two strains under these conditions. The genome-scale metabolic models did not represent regulatory mechanism so it could not predict metabolic changes during stress conditions (i.e., hypoxic growth). This highlights the need of new *Streptomyces* models that integrate kinetic and enzymatic constraints as done in *E. coli* [79] and yeast [80]. As well, applying new methods to predict metabolic fluxes from proteins level would be useful [81].

3.5.6 Metabolic differences and impact on secondary metabolism

The productions of the four antibiotics -ACT, CDA, CPK, and RED- were compared *in-silico* between *S. coelicolor* and *S. lividans*, to identify the impact of the strains' metabolic differences on secondary metabolite production. The antibiotic production was predicted *in-silico* with an increasing complexity of constraints, because of the limited experimental data available for antibiotics yield in the two strains. The simulations were started with a minimum growth media (containing glucose, NH_4, Pi, O_2, and trace elements). Then, to identify the trade-off between the production of secondary metabolites and growth, the fluxes were fixed for the antibiotics ACT, CDA, CPK, and RED at different ranges of values from null to the maximum possible *in-silico* production with optimal growth (see details in *Methods*). This helps to see the impact of antibiotic production on the growth, and see if the two functions compete for resources. The results were counter-intuitive, as the production of antibiotic to growth rate was worse in *S. coelicolor* than in *S. lividans* (*Supplementary Figure 3.5*), differing from the usual experimental observations in *S. lividans* with lower production rates of ACT and RED. The slightly higher growth rates in *S. lividans* could come from its lower ATP/ADP ratio compared to *S. coelicolor* [11], as the model needs fewer nutrients to generate ATP for the growth. However, it did not explain the higher optimal fluxes of antibiotics supported at lower growth rates (<1 $gDW.h^{-1}$), especially that low growth rates are closer to real growth conditions associated to the antibiotic production at transition and stationary phase. To identify the causes for these differences, the same method was applied by adding back the reactions deleted, and deleting the reactions added in the *i*SLT1240 model. The addition or deletion of the PFL was the single reaction with the biggest impact on the *S. lividans* model. For every single antibiotic

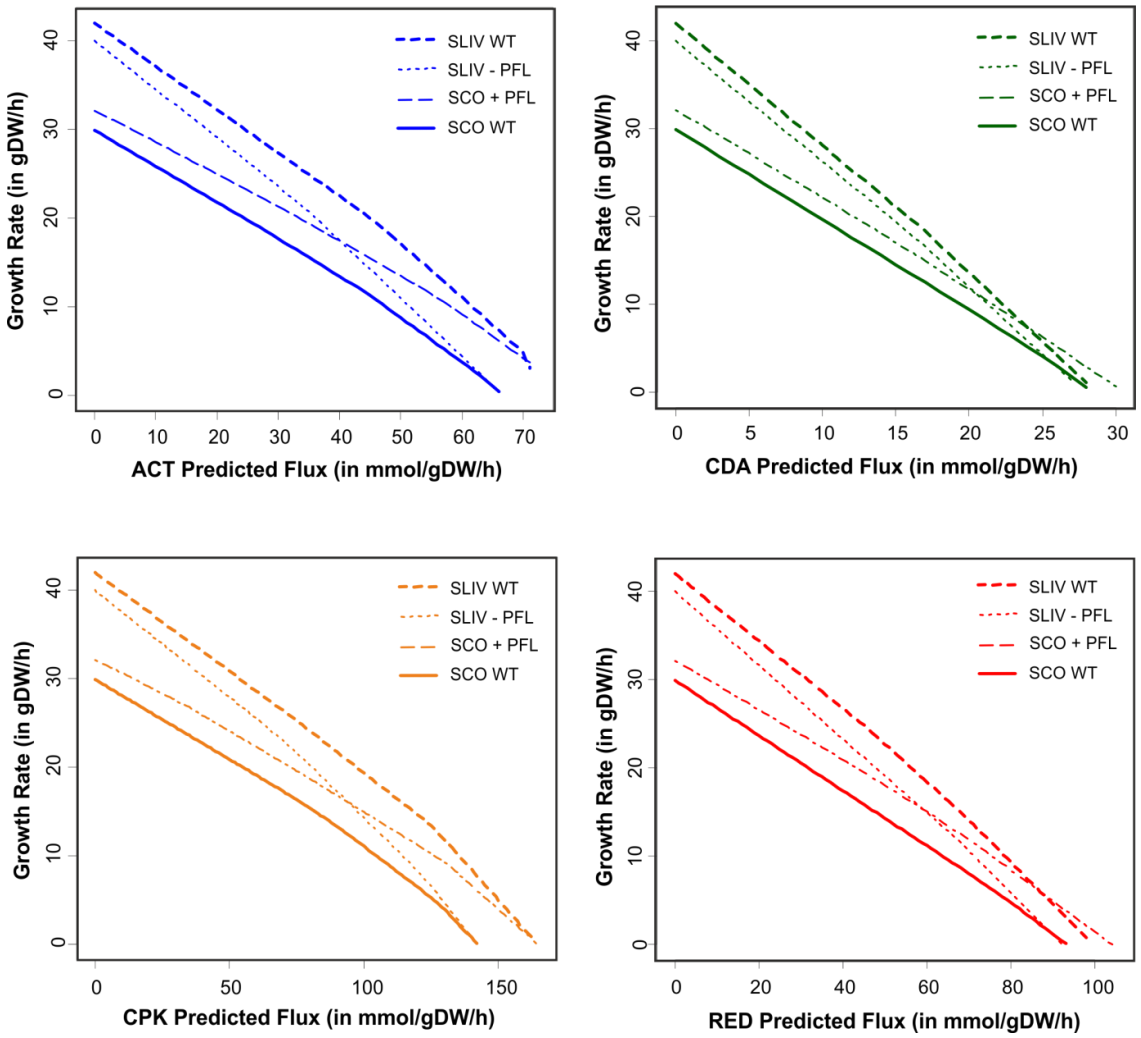


Figure 3.8: Optimal antibiotics production versus optimal growth in *S. coelicolor* and in *S. lividans*, with or without pyruvate formate lyase under minimal conditions

The optimal growth rates to antibiotics production were predicted with ACT (blue), CDA (green), CPK (orange), and RED (red). This was compared with *S. coelicolor* (SCO WT - solid lines), *S. coelicolor* + PFL (SCO + PFL - dotdash lines), *S. lividans* (SLIV WT - dashed lines), and *S.lividans* - PFL (SLIV - PFL - dotted lines) metabolic models. In every cases growth is in direct competition with the antibiotics production, when antibiotics production increases the optimal growth reduces. The presence of PFL seems to enable higher optimal growth (e.g., SCO vs SCO + PFL and SLIV vs SLIV - PFL) when antibiotics production is high. But PFL does not account for all of the growth difference between the two strains, at least in high growth optimum (low antibiotics production).

produced the deletion of PFL led to a drop of antibiotic production, and in the low growth rates (<1 $gDW.h^{-1}$) production reached the same levels or lower than *S. coelicolor* (*Figure 3.8*). In the opposite scenario, when PFL was added to the *i*AA1259 model the growth rate to antibiotic production increased by 10 to 60% depending on

the growth rate. In low growth rates (<1 $gDW.h^{-1}$) to antibiotics production fluxes reached equal or higher level than *S. lividans* (*Figure 3.8*).

The PFL is an enzyme present in multiple facultative and obligate aerobes [37,82,83], which is important for the use of pyruvate under anaerobic conditions. PFL converts pyruvate (+ CoA) into acetyl-CoA and formate. PFL is activated by a pyruvate formate-lyase activating enzyme. PFL has the same function (but under anaerobic conditions) as the Pyruvate Dehydrogenase (PDH) in the metabolism, which converts pyruvate into acetyl-CoA generating NADH and CO_2. The presence of PFL introduces a second path in the model to generate acetyl-CoA with a lower burden on the oxidative balance (as it does not generate NADH). The higher production of acetyl-CoA by PFL explains the higher predicted fluxes toward ACT, CDA, CPK, and RED as acetyl-CoA is either directly used or converted in malonyl-CoA to produce these antibiotics (*Figure 3.9*). The model estimates higher predicted fluxes consuming acetyl-CoA and malonyl-CoA for antibiotics production. In fact, the effect of the deletion/addition of PFL is more dramatic on ACT and CPK fluxes (*Figure 3.8*), as these use malonyl-CoA and acetyl-CoA as building blocks (respectively type II and type I PKS), whereas RED and CDA consume a diverse mix of precursors (e.g., AA). The models were optimized to produce the individual antibiotics while maintaining a low growth rate (< 0.1 $gDW.h^{-1}$) close to the ones experimentally observed at transition phase. Then, to study the impact of PFL on the overall metabolism, the predicted exchanges were compared between the strains with or without PFL, as major intracellular metabolic changes impact the exchange fluxes between the media and the intracellular milieu (*Supplementary Figures 3.6, 3.7, 3.8, and 3.9*). For example, the introduction of the PFL in the *S. coelicolor* model

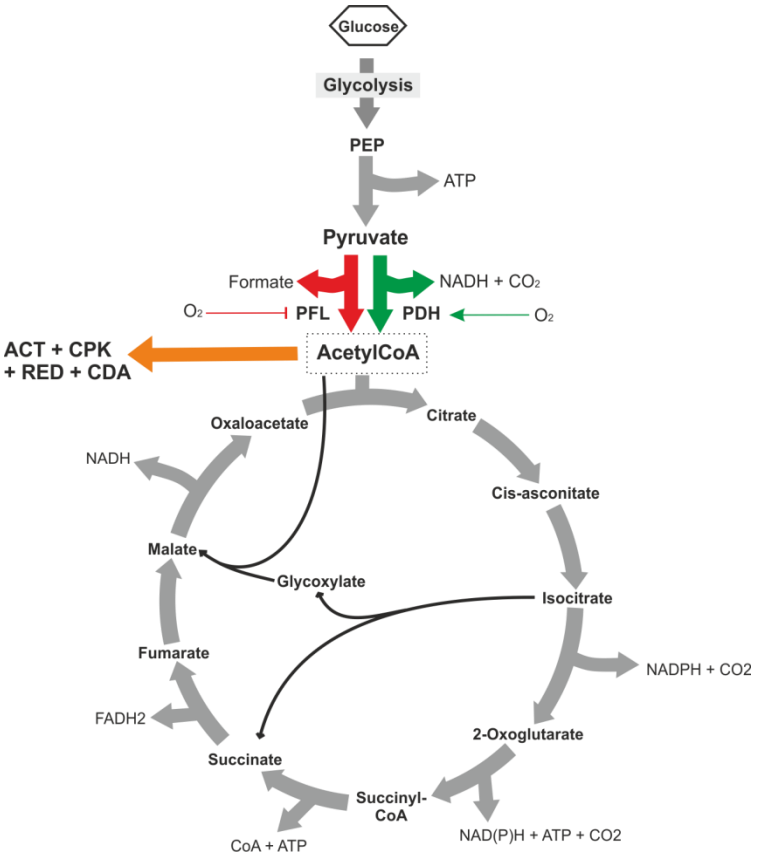


Figure 3.9: Introduction of the Pyruvate Formate Lyase in the central metabolism

The presence of the PFL has a major impact on the metabolism because it is placed in a key point in the metabolic network, the PFL carry fluxes from pyruvate produced by glycolysis toward acetyl-CoA that will be used for both the TCA cycle and the production of the four main antibiotics. Another enzyme the pyruvate dehydrogenase (PDH) has the same function but is used under aerobic condition by opposition to PFL used in anaerobic condition. The impact of PFL on production of antibiotics can be explained by its connection to the acetyl-CoA, which by increasing fluxes toward it will increase the fluxes available to go toward antibiotics production. Furthermore, the PFL is not affected by oxidative balance as it does not need NAD like PDH.

resulted in an increased export of formate and a decreased export of CO_2, this is due to the model favouring PFL that produces formate instead of PDH that produces CO_2; the opposite phenomenon was observed in *S. lividans* when PFL was deleted.

Under the same conditions, the overall predicted fluxes were compared for the different strains with or without PFL and for all four antibiotics produced (*Figure 3.10*). In the *S. coelicolor* model with PFL added, multiple fluxes significantly increase or

decrease showing a significant impact of PFL on multiple predicted fluxes *(Figure 3.10a* and *3.10b)*. The overall predicted fluxes in *S. coelicolor* with PFL compared to *S. lividans* WT, show similar fluxes distribution with the exception of a few dozen fluxes, as well as lower averages for the Euclidian distances between the fluxes (*Figure 3.10c*). The opposite case *S. lividans* without PFL compared to *S. coelicolor* shows a similar distribution of predicted fluxes with the exception of twenty fluxes; the Euclidian distances are also lower in this case (*Figure 3.10d*). The addition of only PFL considerably impacts the predicted differences in fluxes and optimal antibiotic production. However, the exchange reactions are not constrained with experimental data, so the conditions are not close enough to reality to conclude the impact of PFL.

The models were constrained using exchange fluxes in the R2YE media to simulate the experimental conditions from Esnault et al., [11]. These conditions predicted significantly different results than with the reduced constraints in minimal media. Under minimum constraints, for every antibiotic the trade-off between production levels and biomass shows a direct competition for resources, with a linear decrease of biomass when antibiotic productions increase (*Figure 3.8*). However, the trade-off is more complex with experimental uptake and complex media constraints, showing different type of trade-off depending on the strain or the antibiotic. For CDA the trade-off is linear and very similar for both strains, whereas for ACT, RED, and CPK the growth rate decreases at different rates around similar values for both strains (*Figure 3.11*). For ACT, CPK, and RED production does not seem to directly impact growth until a "breaking point", where the growth rate abruptly decreases (*Figure 3.11 a, b, & c*). *S. coelicolor* seems to withstand slightly higher fluxes of ACT, RED, and CPK than *S. lividans*, as the growth rate fall later (*Figure 3.11 a, b, & c*). In all the cases the

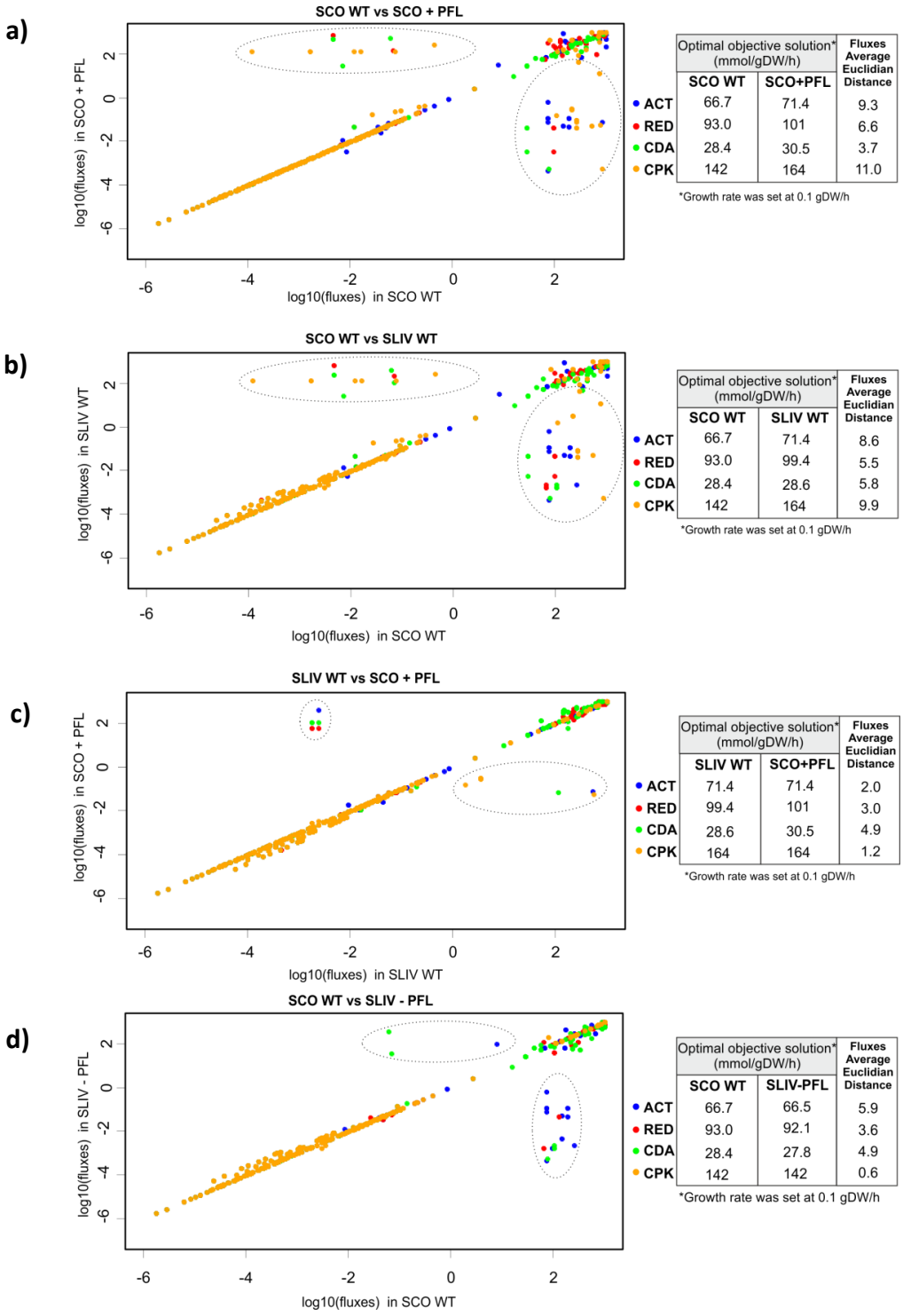


Figure 3.10: Overall flux comparison of *S. coelicolor* and *S. lividans* with or without the pyruvate formate lyase

The fluxes were compared between the two strains and in presence or absence of PFL. The fluxes values show that by adding PFL to *S. coelicolor* the predicted flux values get very close to the ones in *S. lividans* (SCO + PFL vs SLIV WT) and same for the opposite case (SLIV – PFL vs SCO WT). The flux values between SCO WT vs SCO + PFL are as distant as for SCO WT vs SLIV WT. The four antibiotics were tested ACT (blue), RED (red), CDA (green), and CPK (orange), the fluxes were predicted at a fixed growth rate (0.1 $gDW.h^{-1}$) for all the strains and antibiotics. The average Euclidian distance between fluxes shows the increase of difference between the fluxes in the different cases.

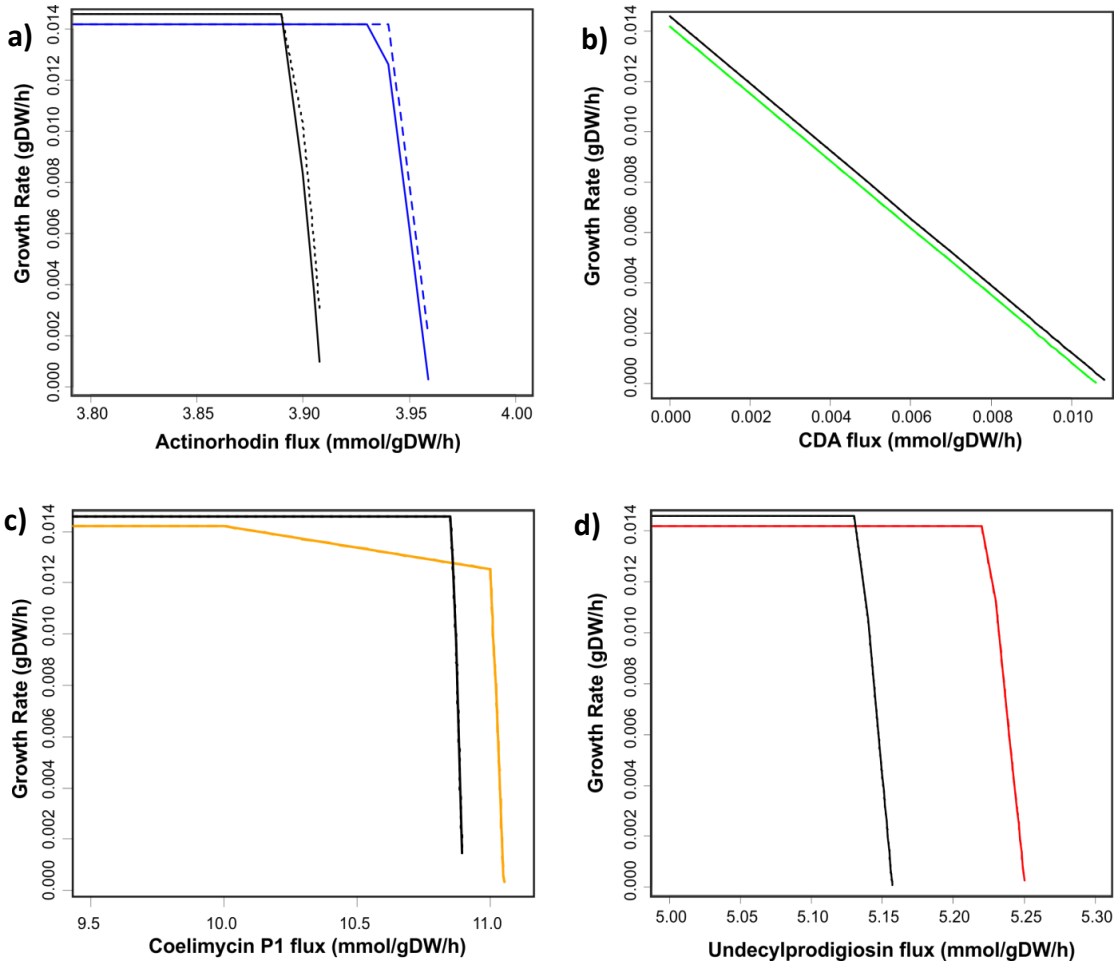


Figure 3.11: Optimal antibiotics production versus optimal growth in *S. coelicolor* and in *S. lividans*, with or without pyruvate formate lyase in a complex media

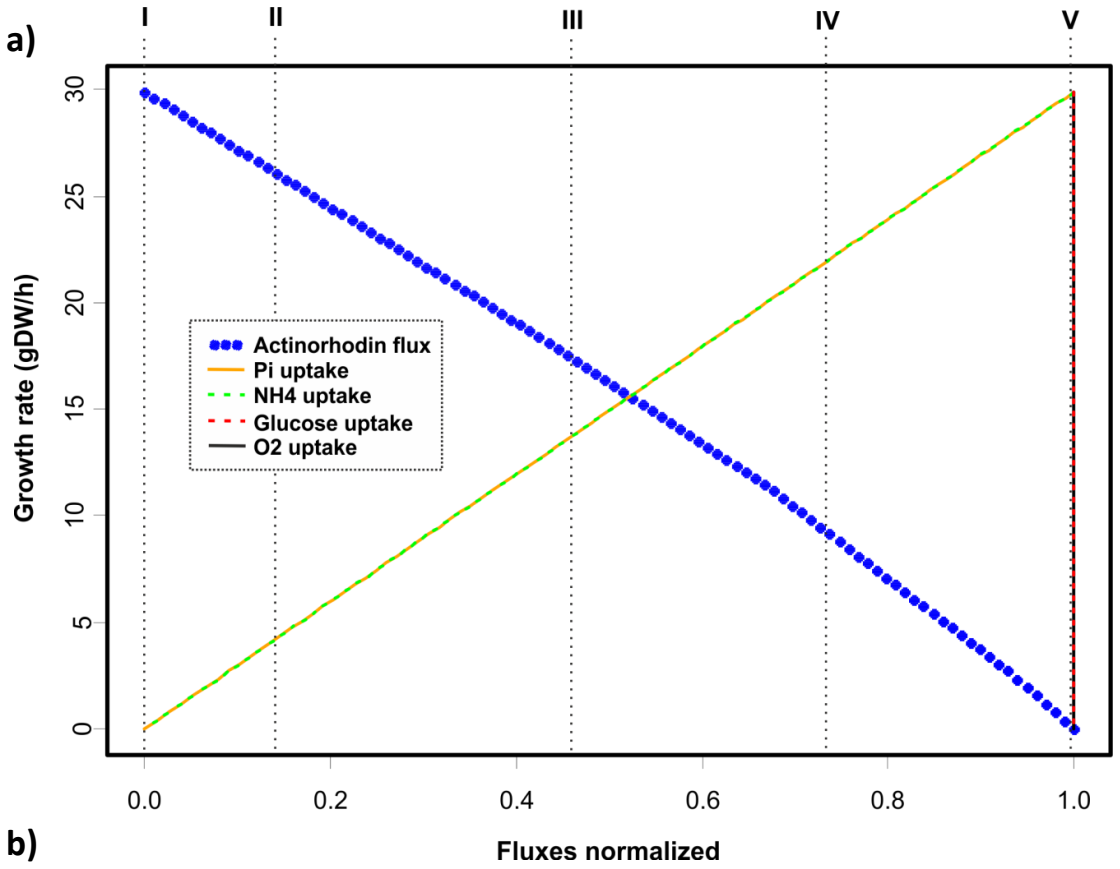
The optimal growth rates to antibiotics production were predicted with ACT (blue), CDA (green), CPK (orange), and RED (red) for *S. coelicolor* and black for *S. lividans*. This was compared with S. coelicolor (SCO WT - solid lines), *S. coelicolor* + PFL (SCO + PFL - dotdash lines), *S. lividans* (SLIV WT - dashed lines), and *S. lividans* - PFL (SLIV - PFL - dotted lines) metabolic models. The models were constrained for a growth in in Pi-limited R2YE media using experimental uptakes [11]. Here, the presence of PFL does not seem to significantly impact either antibiotic production or growth (slightly with ACT). Only CDA shows a direct competition between growth and CDA production, with very similar values for both. For ACT, RED, and CPK the biomass does not seem to be in competition with production until a "breaking" point where biomass production dramatically decreases. This shows the optimum point under these conditions where the strains can sustain optimal growth and antibiotic production (for ACT, RED, & CPK). This graph is a version with a truncated x-axis to focus on the differences; the non-truncated version is shown in *Supplementary Figure 3.9*.

presence or absence of PFL does not seem to have as much impact on the antibiotics production as in minimal media (*Figure 3.11*). In the case of ACT, CPK, and RED (to a lesser extent), the trade-off trend is slightly different between *S. coelicolor* and *S. lividans* as a producer, in *S. coelicolor* the growth rate decreases at different paces creating two "breaking points" (*Figure 3.11a, 3.11b,* and *3.11c*). This trade-off with two inflection points could come from the use of different resources limiting the antibiotics production while maintaining the growth rate. For example, CDA uses amino-acids, acetyl-CoA, malonyl-CoA, and prephenate (amino-acids biosynthesis intermediate) as precursors [84], these metabolites are interconnected with the central metabolism limiting the opportunities for the metabolic network to optimize growth while producing CDA. Whereas, ACT, RED, and CPK have a major building block in common acetyl-CoA (or other related metabolites like malonyl-CoA), which is a central metabolite connected to multiple pathways (*Figure 3.9*), that offers more solutions to optimize biomass while keeping optimal production until it reaches a breakpoint. The difference in trade-off between biomass and antibiotic production and biomass seem to come from the difference in carbon sources available. In the minimal media, the model has fewer opportunities to compensate the strong use of acetyl-CoA and cofactors for antibiotic production and biomass.

A detailed robustness analysis was carried out to understand the impact of environmental conditions (i.e., different media) on antibiotic production. Different tests were carried out: estimation of the sensitivity of the network optimal state to production of ACT and nutrients uptake, as well as estimating the shadow prices of the nutrients uptakes. The shadow price represents the amount by which the objective function (i.e., biomass) would increase if adding a given nutrient [85]. In *S. coelicolor* in

minimal media, the optimal growth to ACT production decreases linearly at the same time as phosphate and NH_4 available, while O_2 and glucose already reached the maximum constraint possible (*Figure 3.12a*). The shadow prices for the nutrients O_2, glucose, NH_4, and Pi do not show significant values compared to the biomass values (*Figure 3.12b*). So, the biomass trend is likely associated to the direct competition for internal resources between ACT and biomass as a single nutrient does not limit directly the growth here. See *Electronic Supplementary – Additional File 8* for the robustness analysis of biomass against nutrient exchange fluxes in *S. coelicolor* WT or with PFL, and *S. lividans* WT or without PFL, with production of ACT, CDA, CPK, or RED. In the complex media, the glucose and Pi reached the maximum uptake constrained constantly, whereas other nutrients like amino acids reached it later (*Figure 3.13a*). The amino acids reach a maximum uptake one after another until the last one, isoleucine, where the biomass starts dropping (*Figure 3.13a*). The shadow prices of the different nutrients gives some insight into the way the constraints explain the robustness analysis of the biomass against the ACT production as these showed the nutrients responsible for limiting growth and ACT production. During the plateau phase where biomass to antibiotic is maintained to the optimal value, the only significant shadow price is for Pi (*Figure 3.13b*). So, the Pi uptake constraints are responsible for blocking biomass production to this level. The shadow values for glucose, histidine, proline, and O_2 once the biomass drops, while Pi shadow value is null (*Figure 3.13b*). This means that the biomass drop is triggered by a lack of new carbon sources available to maintain biomass and production of ACT, and at this stage more phosphate would not increase biomass to antibiotics at this point. The NH_4 shadow values are null confirming that the amino acids are used for their carbon, not

a)



b)

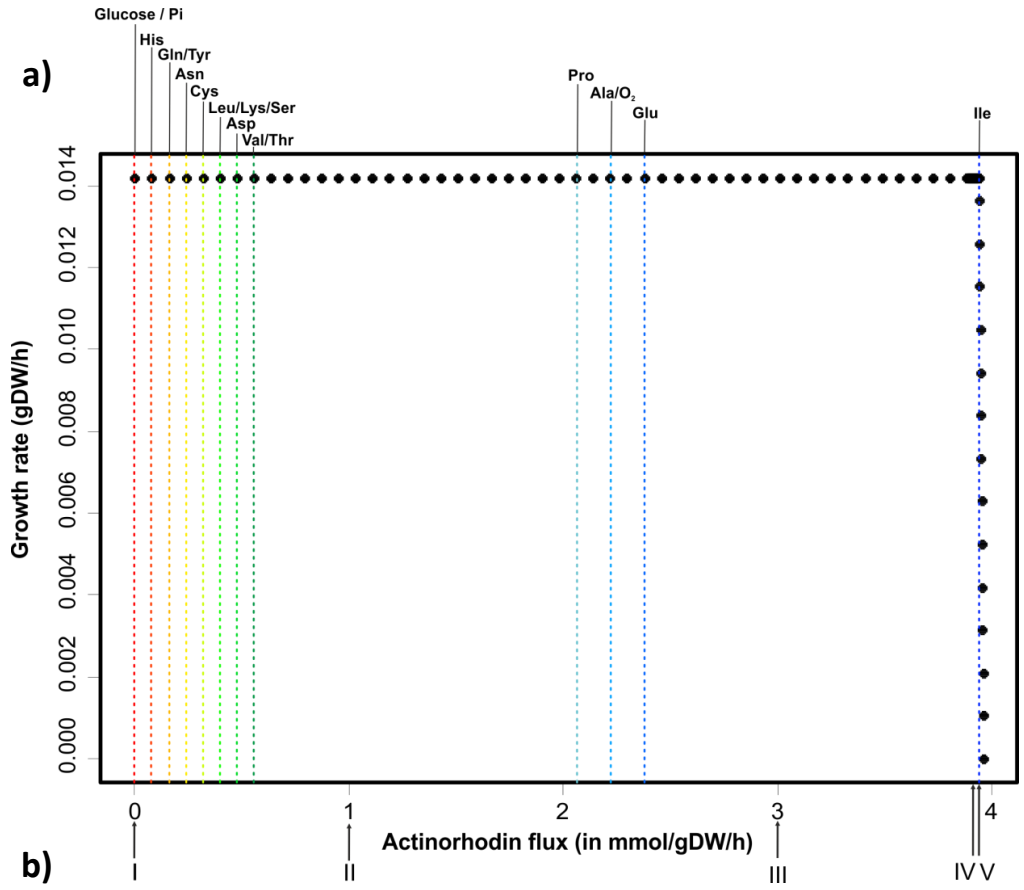
Phase	ACT	Biomass	Pi	Glucose	O_2	NH4
I	0	29.8994	0	-2.56E-03	-0.007927	0
II	10	25.8496	0	0	-0.007927	0
III	30	17.7170	0	0	-8.54E-03	0
IV	50	8.7921	0	0	-0.0116	0
V	66.8	0.0496	0	-0.0033	-0.0132	0

Figure 3.12: The effects of varying Actinorhodin production on the ability of the *S. coelicolor* WT metabolic network to support growth in minimal media.

The uptake rates of the different nutrients were unconstrained (using default 1000 mmol/gDW/h), only the ACT production was constrained.

a) The biomass production (gDW/h) as a function of ACT production, phosphate, NH_4, glucose, and O_2 uptake. The glucose (red dashed line) and O_2 (grey line) uptake reaches the maximum allowed on the whole curve, while the Pi and NH_4 uptake decrease linearly with the biomass decrease associated to the production of ACT.

b) Table of the shadow prices for the main nutrients of the media at different phase on the biomass to ACT production. The shadow cost of Pi and NH_4 are null as these two nutrients do not result in growth decrease. Glucose shadow price is negligible, and the O_2 shadow price is low compared to the biomass.



Phase	ACT	Biomass	Pi	Glucose	O_2	HIS	PRO	ILE	NH4
I	0	0.0142	-1.339	4.44E-16	0	0	0	0	0
II	1	0.0142	-1.339	-1.04E-14	0	-1.55E-15	0	4.44E-16	0
III	3	0.0142	-1.339	-2.47E-16	5.27E-16	1.73E-16	1.48E-16	0	0
IV	3.95	0.0061	0.000	-0.0628	-0.0389	-0.0400	-0.0165	0	0
V	3.959	0.0003	0.000	-0.0628	-0.0389	-0.0400	-0.0165	0	0

Figure 3.13: The effects of varying Actinorhodin production on the ability of the *S. coelicolor* WT metabolic network to support growth in complex media.

The model was constrained with the complex media R2YE nutrient uptakes, with ACT production constrained at different levels and the biomass optimized.

a) The biomass production (gDW/h) as a function of ACT production. The dashed coloured lines corresponds to the amino-acids reaching their maximum uptake at different points of ACT production, isoleucine is the last one before the biomass drops.

b) Table of the shadow prices for some key nutrients consumed by the model. The NH_4 shadow price is null, so nitrogen is not limiting here. The phosphate shadow price is high during the plateau phase (at 0.014 gDW/h of biomass); this means that the phosphate is limiting higher growth there. Glucose, O_2, histidine, and proline have null or near null shadow prices, so the carbon or oxygen sources are not limiting the growth in the plateau phase. However, when the biomass drops the shadow price for the phosphate becomes null, while the shadow prices for the glucose, histidine, proline, and O_2 increased. So if more glucose, or histidine, or proline, or O_2 constraints allowed higher uptakes the biomass could reach higher values. Isoleucine has null shadow prices, so despite being the last amino-acids uptake before the biomass drop, more of it would not increase biomass.

for the nitrogen. However, this pattern of growth to antibiotics is not observed in CDA, where the competition is direct between biomass and CDA production. In this case, the phosphate and glucose are limiting, despite the amino acids in the media the growth decrease linearly while CDA production increases. The asparagine, threonine, tryptophan, and tyrosine consumption increase linearly with CDA production, as these are CDA precursors (see CDA heat maps in *Electronic Supplementary – Additional File 8*). See *Electronic Supplementary – Additional File 8* for the other robustness analysis of biomass against nutrient exchange fluxes in *S. coelicolor* WT or with PFL, and *S. lividans* WT or without PFL, producing ACT, CDA, CPK, or RED.

Production may be optimized without major growth loss (until a critical point) for compounds such as ACT, CPK, and RED using a limited number of major building blocks such as acetyl-CoA. The differences in antibiotics production between *S. coelicolor* and *S. lividans* may not come from a difference in their metabolic network, but are likely caused by regulatory phenomena in the central metabolism and the biosynthetic pathways. This analysis helped to identify the potential impact of PFL expression on antibiotics production in *S. coelicolor* and in *S. lividans*. So far, it appears that PFL has never been studied in *Streptomyces* species and this enzyme has the potential to redirect fluxes toward acetyl-CoA for antibiotics production. Interestingly the presence or absence of PFL seems to mainly impact production under minimal media conditions, an effect that is still not fully understood. It is possible that PFL makes glucose use by glycolysis more efficient, which is the main carbon source in minimal media. In contrast, the effect of the PFL could be less dramatic in the presence of diverse carbon sources as other alternate pathways can generate precursors for biomass and antibiotics production.

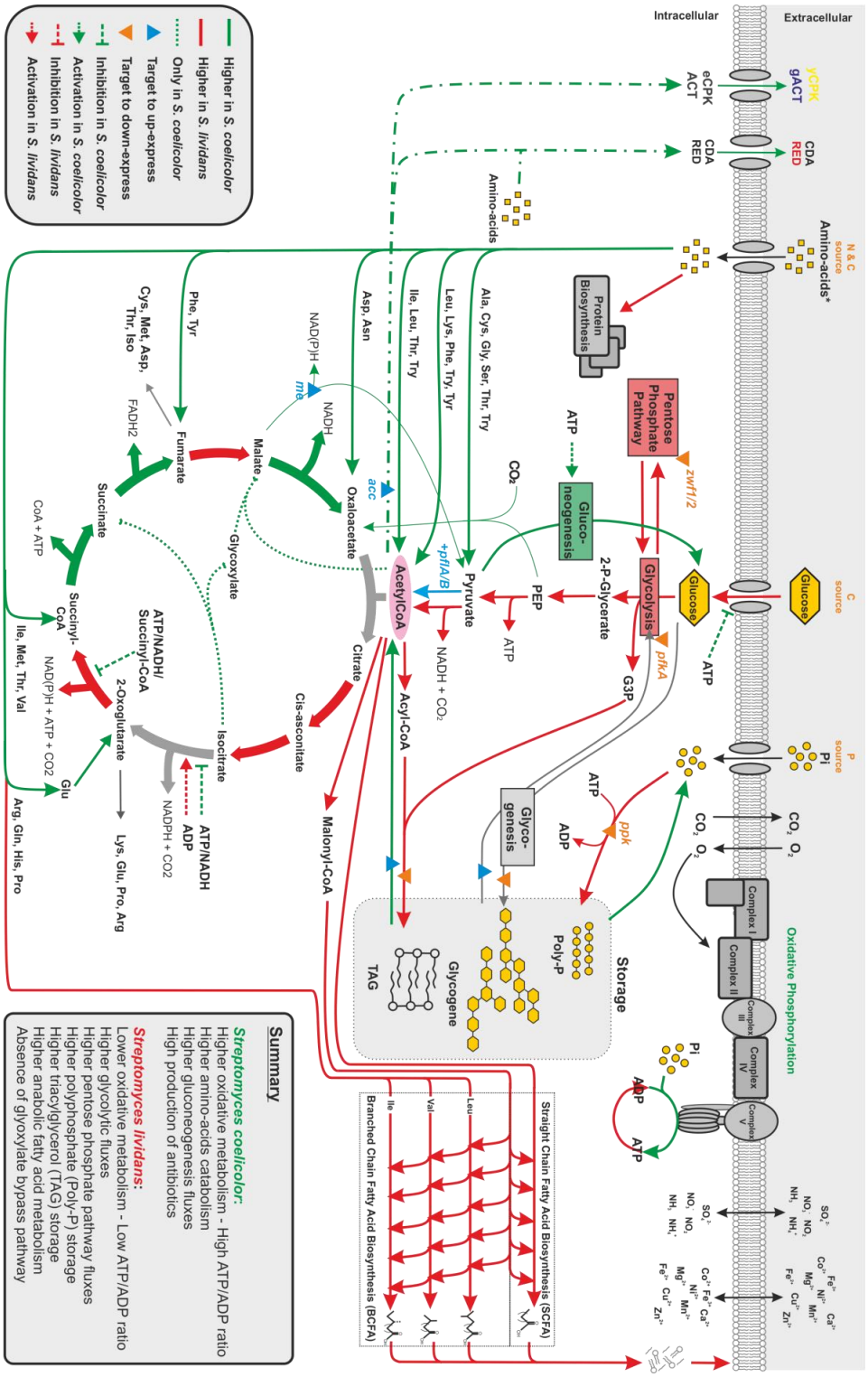


Figure 3.14: Summary metabolic map of metabolic differences and targets of interests to further study and increase antibiotics production in *S. coelicolor* and *S. lividans*

The main differences between *S. coelicolor* and *S. lividans* metabolism were summarized based on metabolic predictions, proteomics data analysis, and literature. A few metabolic or genetic targets are highlighted, which could help to understand the primary metabolic mechanisms impacting antibiotics differences, and potentially help increase antibiotics production.

The expression of PFL in *S. coelicolor* could increase the production of antibiotics by increasing fluxes toward acetyl-CoA while saving up carbons, compared to the PDH that generates CO_2 as a waste product. However, the PFL is only expressed under anaerobic conditions, and repressed by the presence of O_2 under the control of an activating enzyme [86,87]. A PFL from *E. coli* (*pflB*) was expressed with its activator (*pflA*) in a *Saccharomyces cerevisiae* strain lacking a PDH, this required the expression of electron donors used by the activating enzyme as cofactor (flavodoxin and ferredoxin) to get an active PFL in aerobic conditions [88]. Thus, the *in-vivo* expression of PFL under aerobic condition could work also in Streptomyces and potentially increase antibiotics production.

3.6 Conclusion

The aim of this study was to identify major metabolic differences associated with the difference in antibiotic production between two very closely related *Streptomyces* strains, *S. coelicolor* and *S. lividans*. The metabolic models of both strains were used to predict the major metabolic differences, apply an integrative proteomics data analysis, and explore the production capabilities of four different antibiotics (ACT, RED, CPK, and CDA). The models predicted some major metabolic differences and helped as frameworks to analyse from a formal metabolic perspective experimental studies done on the two *Streptomyces* strains. Finally, this study helped to have an overview of the metabolism of both strains and its impact on antibiotics production, as well as highlighting potential targets to increase antibiotics production. A comprehensive metabolic map was built by including predictions, proteomics data

analysis, metabolites exchanges, and potential targets to increase antibiotics production (*Figure 3.14*).

A new genome-scale metabolic model for *S. lividans* TK24, *i*SLT1240, was reconstructed from a *S. coelicolor* model, *i*AA1259, that we previously reconstructed [33]. Initially, the comparison of the metabolic genes between the two models revealed some interesting features in the *S. lividans* model. This included the absence of the glyoxylate pathway that plays a key role in central metabolism, and the addition of a pyruvate formate lyase involved in anaerobic central carbon metabolism. Then, the *S. lividans* model was validated using multiple experimental datasets. The two models were constrained, and the metabolic predictions were analysed and compared to experimental data [11]. Under these conditions, the *S. coelicolor* model predicted a higher amino-acid catabolism, a partially higher gluconeogenesis, and higher oxidative phosphorylation. In contrast, the *S. lividans* model predicted a higher glycolysis and a higher fatty acid biosynthesis (due to higher TAG storage). These predictions were also broadly validated by proteomics data acquired under similar conditions [20]. Some of the metabolic differences predicted seem to come from the differences in oxidative balance between the two strains; the oxidative phosphorylation was partially constrained using the experimental data as the *S. coelicolor* model has a higher ATP/ADP ratio than *S. lividans*. Overall, *S. lividans* has a more glycolytic metabolism, a lower oxidative phosphorylation, and a higher triacylglycerol and polyphosphate storage (*Figure 3.14*). In contrast, *S. coelicolor* has a strong amino acid catabolism replenishing the TCA cycle, a higher oxidative phosphorylation, and acetyl-CoA consumption derived toward ACT rather than the TCA cycle (*Figure 3.14*). The high oxidative phosphorylation is critical to the redox balance, which modifies the cofactor

and electron transporter balance (i.e., ATP, NADH, NADPH, $FADH_2$, FMNH, etc), impacting the entire metabolism including the production of secondary metabolites. The differences in oxidative balance multiple reactions fluxes involved in the primary metabolism, as these reactions consume or regenerate reduced cofactors directly involved in oxidative phosphorylation. Afterwards, the proteomics data was used to constrain the model, although the proteomics data did not show significant correlation with the individual flux predictions because of the experimental conditions (measurements done in hypoxic conditions at stationary growth phase). The proteomics information was finally used for an integrative analysis in the models. The proteomics integration showed that some proteins switched off differently between the two strains; these proteins were associated to the oxidative respiration adaptation to hypoxic conditions, and to the metabolic switch occurring at stationary phase with enzymes involved in membrane lipids biosynthesis, cofactors and vitamins metabolism (e.g., adenosylcobalamine, cobalamine). Following that, the antibiotics production capabilities were studied by predicting biomass and antibiotics production in different scenarios. Unexpectedly, under minimal conditions *S. lividans* was predicted as a better producer of all four antibiotics. However, this was largely caused by one enzyme only acquired by *S. lividans*, the pyruvate formate lyase, which is only used in the central metabolism under anaerobic condition. However, when growing in a complex media with multiple carbon sources and constrained with experimental fluxes, *S. coelicolor* was a slightly better producer than *S. lividans*. The absence of major differences under these conditions suggests that major changes in antibiotic production are not caused by the differences of topology between the two strains' metabolic network (i.e., the absence / presence of enzyme-coding genes). Concretely,

the fluxes are not directed toward antibiotic production in *S. coelicolor* compared to *S. lividans* because of the fundamental structure of the metabolic network. However, under minimal conditions that are frequently used in the lab, this would be different, as the network is less "robust" under these condition, and only a few main pathways seem to influence the production of antibiotics. The two strains have very similar metabolic networks but in different states, as indicated by the differences in nutrient uptake and antibiotic production.

To further understand the impact of the metabolic differences between *S. coelicolor* and *S. lividans*, it could help to express some key genes in *S. lividans* or delete them in *S. coelicolor*. For instance, the gene SCO0046 involved in fatty-acid biosynthesis and regenerate holo-ACP used to for antibiotics biosynthesis. The gene SCO0548 encodes a malonyl-ACP decarboxylase absent from *S. lividans* genome, which may force the use of acetyl-CoA instead of malonyl-ACP to generate fatty acids precursors. Finally the most interesting ones are SCO0982 and SCO983 missing from *S. lividans* genome, leading to a loss of the glyoxylate bypass metabolism. This is likely to have a major impact on the oxidative balance of the strain, and modify the accumulation of TCA cycle intermediates that is known to change the carbon control of primary metabolism and impact antibiotic production [17,89]. The reintroduction of these primary metabolic enzymes lost in *S. lividans* could help trigger antibiotic production and help understand their role for secondary metabolism. Understanding the primary metabolism and particularly the central metabolism would help to rationally engineer overproduction of antibiotics in *S. coelicolor*.

Finally, the prediction results were combined with available data to have an overview of the two strains metabolism and identify potential targets to increase

antibiotics production. Many genes deleted or overexpressed to increase antibiotics production were in central metabolism. For example, the reduction of fluxes in glycolysis and the pentose phosphate pathway by gene knockout increased actinorhodin production. Or the overexpression of some enzymes such as the acetyl-CoA carboxylase or the malic enzymes increased fluxes toward antibiotics production. These were summarized in a comprehensive metabolic map of the two strains' major metabolic differences and potential targets to increase antibiotic production (*Figure 3.14*). Discovering new enzymes to introduce in the antibiotic producers could increase production by creating new pathways toward antibiotic precursors. An example would be the pyruvate formate lyase, which could modify central metabolism by redirecting fluxes to produce more acetyl-CoA. Other metabolic differences, such as the glyoxylate pathway absent from *S. lividans,* should be studied to understand their effect on antibiotic production, for example, by expressing it in *S. lividans* to observe if antibiotic production is recovered or by knocking it out in *S. coelicolor*. Finally, the modelling method used here did not take into account any regulatory phenomena, but with appropriate experimental constraints it helped to assess the metabolic state of the strains and provided a metabolic framework to analyse omics data. In the future, the development of genome-scale metabolic models that include enzyme and regulatory constraints – as have already been presented for *E. coli* [79,90] and *S. cerevisiae* [80] – would help to understand and engineer *Streptomyces* metabolism with greater confidence. For example, the high oxidative phosphorylation in *S. coelicolor* results in higher levels of ATP, ATP has an inhibitory effect on glucose import and degradation through glycolysis, which could explain the lower glycolytic metabolism in *S. coelicolor*. In *S. lividans*, glucose has a repressing effect on the

production of ACT, which was relieved when glycerol was used as the sole carbon source [91]. As glucose is easier to degrade the strain probably prefers to use a glycolytic metabolism, which does not increase oxidative phosphorylation, consequently ACT production was not triggered. As these regulatory phenomena have a major impact on the primary and secondary metabolism, it makes it more complex to predict the effect of mutations without taking into account these regulations. For example, the deletion of the malic enzymes in *S. coelicolor* led to lower actinorhodin production, which was due to a lower expression of the transcriptional activator ActII-ORF4 that controls the ACT gene cluster [17]. If a BGC is under control of its native promoters or repressors, modifying the primary metabolism will probably impact the expression of the BGC rather than the production of the compound itself. Another example, if one wants to decrease storage of biomolecules (e.g., TAG or glycogen) to increase precursors availability for secondary metabolites biosynthesis, one would need to take in account the interplay between central carbon metabolism and morphological development [13]. So, increasing degradation when antibiotic production starts could avoid interfering with morphological differentiation such as sporulation [92,93].

In the end, the carbon central metabolism has a major influence on antibiotic production, so it is critical to understand it, and learn to control it to push fluxes toward specific secondary metabolites. In particular, with polyketides biosynthesis that largely uses acetyl-CoA and malonyl-CoA as building blocks. However, the control of the central metabolism is a major challenge, as it is highly interconnected to the rest of metabolism with many regulation points; these characteristics confer to biological systems their adaptability to diverse environment and conditions [94,95]. The rewiring of central metabolism has been a major focus in synthetic biology and metabolic

engineering, as this is the metabolic platform distributing resources for carbon-based molecules of organisms [96,97]. For *Streptomyces*, the complexity of the metabolic connection between primary and secondary metabolism comes from the evolutionary constraints associated with the stress of surviving in a competitive and changing environment such as soils. So, to enable rational engineering of this species it is important to understand the interplay between nutritive stress, metabolic regulation of central metabolism, and secondary metabolism. A promising path is the study of the influence of overflow in central metabolism (linked to oxidative phosphorylation) on secondary metabolites production. Thus, one of the main theories for the evolutionary origin of secondary metabolism considers that overflow metabolism led to the evolution of secondary pathways to keep the metabolism "ticking-over", then these compounds were selected because their functions gave an advantage (e.g., competitive advantage of antibiotics, or nutritive advantage with chelators) [98]. This would explain many of the metabolic regulations and functions found in the soil-dwelling bacteria. For instance, when nitrogen or phosphate sources are limiting, to keep generating cofactors for respiration, the organism could use the rich and diverse carbon sources available in the soil, which could lead to an overflow in carbon metabolism; this would be kept running by diverting fluxes toward secondary pathways. Consequently, it would make sense that the secondary metabolic pathways are under direct or indirect regulation by central carbon metabolism, as it would switch fluxes toward these compounds when needed. Then, the pathways with compounds giving an evolutionary advantage were selected. For example, antibiotics giving a competitive advantage, chelators capturing trace elements, or compounds with electron transport abilities needed for the oxidative phosphorylation. Under

starvation the organism could not generate enough primary electron transporters (e.g., menaquinones), so it would need alternative electron transporter, selecting secondary metabolites with that function (e.g., antibiotics and electron transporters). This could explain the correlation between oxidative phosphorylation and ACT production between the two strains[11], as ACT in addition to being an antibiotic [71] can catalyse oxidative reactions [99] and is a redox active compound impacting oxidative regulation in *S. coelicolor* [100]. As previously pointed out, CDA and RED are likely to also have functions related to energy dissipation [11]. In conclusion, understanding these complex metabolic interplays and the origins of secondary metabolism would help to move away from brute force methods (e.g., random mutagenesis) to a rational engineering of natural products overproduction, and discovery.

3.7 References

(1) Organization, W. H. *Antimicrobial Resistance: Global Report on Surveillance*; World Health Organization, 2014.

(2) Watve, M. G.; Tickoo, R.; Jog, M. M.; Bhole, B. D. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **2001**, *176* (5), 386–390.

(3) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Brown, L. C. W.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014**, *158* (2), 412–421.

(4) Medema, M. H.; Blin, K.; Cimermancic, P.; Jager, V. de; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. **2011**.

(5) Chater, K. F. Recent advances in understanding *Streptomyces*. *F1000Research* **2016**, *5*.

(6) Medema, M. H.; Breitling, R.; Bovenberg, R.; Takano, E. Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat. Rev. Microbiol.* **2011**, *9* (2), 131–137.

(7) Wohlleben, W.; Mast, Y.; Muth, G.; Röttgen, M.; Stegmann, E.; Weber, T. Synthetic biology of secondary metabolite biosynthesis in actinomycetes: engineering precursor supply as a way to optimize antibiotic production. *FEBS Lett.* **2012**, *586* (15), 2171–2176.

(8) Kim, E.; Moore, B. S.; Yoon, Y. J. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* **2015**, *11* (9), 649–659.

(9) Bystrykh, L. V; Fernández-Moreno, M. A.; Herrema, J. K.; Malpartida, F.; Hopwood, D. A.; Dijkhuizen, L. Production of actinorhodin-related" blue pigments" by *Streptomyces coelicolor* A3 (2). *J. Bacteriol.* **1996**, *178* (8), 2238–2244.

(10) Stankovic, N.; Senerovic, L.; Ilic-Tomic, T.; Vasiljevic, B.; Nikodinovic-Runic, J. Properties and applications of undecylprodigiosin and other bacterial prodigiosins. *Appl. Microbiol. Biotechnol.* **2014**, *98* (9), 3841–3858.

(11) Esnault, C.; Dulermo, T.; Smirnov, A.; Askora, A.; David, M.; Deniset-Besseau, A.; Holland, I.-B.; Virolle, M.-J. Strong antibiotic production is correlated with highly active oxidative metabolism in *Streptomyces coelicolor* M145. *Sci. Rep.* **2017**, *7* (1), 200.

(12) Lai, C.; Xu, J.; Tozawa, Y.; Okamoto-Hosoya, Y.; Yao, X.; Ochi, K. Genetic and physiological characterization of rpoB mutations that activate antibiotic production in *Streptomyces lividans*. *Microbiology* **2002**, *148* (11), 3365–3373.

(13) Ryu, Y.-G.; Butler, M. J.; Chater, K. F.; Lee, K. J. Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*. *Appl. Environ. Microbiol.* **2006**, *72* (11), 7132–7139.

(14) Viollier, P. H.; Minas, W.; Dale, G. E.; Folcher, M.; Thompson, C. J. Role of Acid Metabolism in *Streptomyces coelicolor* Morphological Differentiation and Antibiotic Biosynthesis. *J. Bacteriol.* **2001**, *183* (10), 3184 LP – 3192.

(15) Viollier, P. H.; Nguyen, K. T.; Minas, W.; Folcher, M.; Dale, G. E.; Thompson, C. J. Roles of Aconitase in Growth, Metabolism, and Morphological Differentiation of *Streptomyces coelicolor*. *J. Bacteriol.* **2001**, *183* (10), 3193 LP – 3203.

(16) Borodina, I.; Siebring, J.; Zhang, J.; Smith, C. P.; van Keulen, G.; Dijkhuizen, L.; Nielsen, J. Antibiotic overproduction in Streptomyces coelicolor A3 (2) mediated by phosphofructokinase deletion. *J. Biol. Chem.* **2008**, *283* (37), 25186–25199.

(17) Rodriguez, E.; Navone, L.; Casati, P.; Gramajo, H. Impact of Malic Enzymes on Antibiotic and Triacylglycerol Production in *Streptomyces coelicolor*. *Appl. Environ. Microbiol.* **2012**, *78* (13), 4571 LP – 4579.

(18) Sauer, U.; Eikmanns, B. J. The PEP—pyruvate—oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol. Rev.* **2005**, *29* (4), 765–794.

(19) Butler, M. J.; Bruheim, P.; Jovetic, S.; Marinelli, F.; Postma, P. W.; Bibb, M. J. Engineering of Primary Carbon Metabolism for Improved Antibiotic Production in *Streptomyces lividans*. *Appl. Environ. Microbiol.* **2002**, *68* (10), 4731 LP – 4739.

(20) Millan-Oropeza, A.; Henry, C.; Blein-Nicolas, M.; Aubert-Frambourg, A.; Moussa, F.; Bleton, J.; Virolle, M.-J. Quantitative Proteomics Analysis Confirmed Oxidative Metabolism Predominates in *Streptomyces coelicolor* versus Glycolytic Metabolism in *Streptomyces lividans*. *J. Proteome Res.* **2017**, *16* (7), 2597–2613.

(21) Rossa, C. A.; White, J.; Kuiper, A.; Postma, P. W.; Bibb, M.; Teixeira de Mattos, M. J. Carbon Flux Distribution in Antibiotic-Producing Chemostat Cultures of Streptomyces lividans. *Metab. Eng.* **2002**, *4* (2), 138–150.

(22) Ghorbel, S.; Smirnov, A.; Chouayekh, H.; Sperandio, B.; Esnault, C.; Kormanec, J.; Virolle, M.-J. Regulation of *ppk* Expression and In Vivo Function of Ppk in *Streptomyces lividans* TK24. *J. Bacteriol.* **2006**, *188* (17), 6269 LP – 6276.

(23) Chouayekh, H.; Virolle, M.-J. The polyphosphate kinase plays a negative role in the control of antibiotic production in *Streptomyces lividans*. *Mol. Microbiol.*

2002, *43* (4), 919–930.
(24) Liu, L.; Agren, R.; Bordel, S.; Nielsen, J. Use of genome-scale metabolic models for understanding microbial physiology. *FEBS Lett.* **2010**, *584* (12), 2556–2564.
(25) Shoaie, S.; Nielsen, J. Elucidating the interactions between the human gut microbiota and its host through metabolic modeling. *Front. Genet.* **2014**, *5*, 86.
(26) Simeonidis, E.; Price, N. D. Genome-scale modeling for metabolic engineering. *J. Ind. Microbiol. Biotechnol.* **2015**, *42* (3), 327–338.
(27) Kim, W. J.; Kim, H. U.; Lee, S. Y. Current state and applications of microbial genome-scale metabolic models. *Curr. Opin. Syst. Biol.* **2017**, *2*, 10–18.
(28) Schellenberger, J.; Que, R.; Fleming, R. M. T.; Thiele, I.; Orth, J. D.; Feist, A. M.; Zielinski, D. C.; Bordbar, A.; Lewis, N. E.; Rahmanian, S.; et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6* (9), 1290–1307.
(29) Palsson, B. *Systems Biology : Constraint-Based Reconstruction and Analysis*.
(30) Varma, A.; Palsson, B. Ø. Parametric sensitivity of stoichiometric flux balance models applied to wild-type *Escherichia coli* metabolism. *Biotechnol. Bioeng.* **1995**, *45* (1), 69–79.
(31) Cakir, T.; Efe, C.; Dikicioglu, D.; Hortacsu, A.; Kirdar, B.; Oliver, S. G. Flux balance analysis of a genome-scale yeast model constrained by exometabolomic data allows metabolic system identification of genetically different strains. *Biotechnol Prog* **2007**, *23* (2), 320–326.
(32) Lewis, N. E.; Hixson, K. K.; Conrad, T. M.; Lerman, J. A.; Charusanti, P.; Polpitiya, A. D.; Adkins, J. N.; Schramm, G.; Purvine, S. O.; Lopez-Ferrer, D.; et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **2010**, *6*, 390.
(33) Amara, A.; Takano, E.; Breitling, R. Development and validation of an updated computational model of *Streptomyces coelicolor* primary and secondary metabolism. *BMC Genomics* **2018**, *19* (1), 519.
(34) Lewis, R. A.; Laing, E.; Allenby, N.; Bucca, G.; Brenner, V.; Harrison, M.; Kierzek, A. M.; Smith, C. P. Metabolic and evolutionary insights into the closely-related species *Streptomyces coelicolor* and *Streptomyces lividans* deduced from high-resolution comparative genomic hybridization. *BMC Genomics* **2010**, *11* (1), 682.
(35) D'Huys, P.-J.; Lule, I.; Vercammen, D.; Anné, J.; Van Impe, J. F.; Bernaerts, K. Genome-scale metabolic flux analysis of *Streptomyces lividans* growing on a complex medium. *J. Biotechnol.* **2012**, *161* (1), 1–13.
(36) Ahn, S.; Jung, J.; Jang, I.-A.; Madsen, E. L.; Park, W. Role of Glyoxylate Shunt in Oxidative Stress Response. *J. Biol. Chem.* **2016**, *291* (22), 11928–11938.
(37) Sawers, G.; Böck, A. Anaerobic regulation of pyruvate formate-lyase from *Escherichia coli* K-12. *J. Bacteriol.* **1988**, *170* (11), 5330–5336.
(38) Jayapal, K. P.; Lian, W.; Glod, F.; Sherman, D. H.; Hu, W.-S. Comparative genomic hybridizations reveal absence of large *Streptomyces coelicolor* genomic islands in *Streptomyces lividans*. *BMC Genomics* **2007**, *8* (1), 229.
(39) Rückert, C.; Albersmeier, A.; Busche, T.; Jaenicke, S.; Winkler, A.; Friðjónsson, Ó. H.; Hreggviðsson, G. Ó.; Lambert, C.; Badcock, D.; Bernaerts, K.; et al. Complete genome sequence of *Streptomyces lividans* TK24. *J. Biotechnol.* **2015**, *199*, 21–22.
(40) Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A.-M.; Challis, G. L.; Thomson, N.

R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **2002**, *417* (6885), 141–147.

(41) Zdobnov, E. M.; Tegenfeldt, F.; Kuznetsov, D.; Waterhouse, R. M.; Simão, F. A.; Ioannidis, P.; Seppey, M.; Loetscher, A.; Kriventseva, E. V. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **2017**, *45* (D1), D744–D749.

(42) Caspi, R.; Billington, R.; Ferrer, L.; Foerster, H.; Fulcher, C. A.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L. A.; et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **2016**, *44* (D1), D471–D480.

(43) Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27–30.

(44) The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **2015**, *43* (D1), D204–D212.

(45) Devoid, S.; Overbeek, R.; DeJongh, M.; Vonstein, V.; Best, A. A.; Henry, C. Automated Genome Annotation and Metabolic Model Reconstruction in the SEED and Model SEED; Humana Press, Totowa, NJ, 2013; pp 17–45.

(46) Becker, S. A.; Feist, A. M.; Mo, M. L.; Hannum, G.; Palsson, B. Ø.; Herrgard, M. J. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.* **2007**, *2* (3), 727–738.

(47) Ebrahim, A.; Lerman, J. A.; Palsson, B. Ø.; Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7* (1), 74.

(48) Kieser, T. *Practical Streptomyces Genetics*; John Innes Foundation, 2000.

(49) King, Z. A.; Lu, J.; Drager, A.; Miller, P.; Federowicz, S.; Lerman, J. A.; Ebrahim, A.; Palsson, B. Ø.; Lewis, N. E.; J., H. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **2016**, *44* (D1), D515–D522.

(50) Thomas, J.; Cronan, J. E. The enigmatic acyl carrier protein phosphodiesterase of *Escherichia coli*: genetic and enzymological characterization. *J. Biol. Chem.* **2005**, *280* (41), 34675–34683.

(51) Li, Y.; Florova, G.; Reynolds, K. A. Alteration of the fatty acid profile of *Streptomyces coelicolor* by replacement of the initiation enzyme 3-ketoacyl acyl carrier protein synthase III (FabH). *J. Bacteriol.* **2005**, *187* (11), 3795–3799.

(52) Jackowski, S.; Rock, C. O. Acetoacetyl-acyl carrier protein synthase, a potential regulator of fatty acid biosynthesis in bacteria. *J. Biol. Chem.* **1987**, *262* (16), 7927–7931.

(53) Smirnova, N.; Reynolds, K. A. Engineered fatty acid biosynthesis in *Streptomyces* by altered catalytic function of beta-ketoacyl-acyl carrier protein synthase III. *J. Bacteriol.* **2001**, *183* (7), 2335–2342.

(54) Long, P. F.; Bisang, C.; Corte´s, J.; Westcott, J.; Crosby, J.; Matharu, A.-L.; Cox, R. J.; Simpson, T. J.; Staunton, J.; Leadlay, P. F. A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **1999**, *401* (6752), 502–505.

(55) Akopiants, K.; Florova, G.; Li, C.; Reynolds, K. A. Multiple pathways for acetate

assimilation in *Streptomyces cinnamonensis*. *J. Ind. Microbiol. Biotechnol.* **2006**, *33* (2), 141–150.
(56) White, D.; Drummond, J.; Fuqua, C. *The Physiology and Biochemistry of Prokaryotes*; 2007.
(57) Kornberg, H. L. The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochem. J.* **1966**, *99* (1), 1–11.
(58) Ferla, M. P.; Patrick, W. M. Bacterial methionine biosynthesis. *Microbiology* **2014**, *160* (Pt_8), 1571–1584.
(59) Blanco, J.; Coque, J. J. R.; Martin, J. F. The Folate Branch of the Methionine Biosynthesis Pathway in *Streptomyces lividans*: Disruption of the 5,10-Methylenetetrahydrofolate Reductase Gene Leads to Methionine Auxotrophy. *J. Bacteriol.* **1998**, *180* (6), 1586–1591.
(60) Takano, H.; Hagiwara, K.; Ueda, K. Fundamental role of cobalamin biosynthesis in the developmental growth of *Streptomyces coelicolor* A3 (2). *Appl. Microbiol. Biotechnol.* **2015**, *99* (5), 2329–2337.
(61) Stoj, C. S.; Augustine, A. J.; Solomon, E. I.; Kosman, D. J. Structure-Function Analysis of the Cuprous Oxidase Activity in Fet3p from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **2007**, *282* (11), 7862–7868.
(62) Singh, S. K.; Grass, G.; Rensing, C.; Montfort, W. R. Cuprous Oxidase Activity of CueO from *Escherichia coli*. *J. Bacteriol.* **2004**, *186* (22), 7815–7817.
(63) Baldomà, L.; Aguilar, J. Involvement of lactaldehyde dehydrogenase in several metabolic pathways of *Escherichia coli* K12. *J. Biol. Chem.* **1987**, *262* (29), 13991–13996.
(64) Di Costanzo, L.; Gomez, G. A.; Christianson, D. W. Crystal structure of lactaldehyde dehydrogenase from *Escherichia coli* and inferences regarding substrate and cofactor specificity. *J. Mol. Biol.* **2007**, *366* (2), 481–493.
(65) Caballero, E.; Baldomá, L.; Ros, J.; Boronat, A.; Aguilar, J. Identification of lactaldehyde dehydrogenase and glycolaldehyde dehydrogenase as functions of the same protein in *Escherichia coli*. *J. Biol. Chem.* **1983**, *258* (12), 7788–7792.
(66) Heuts, D. P. H. M.; van Hellemond, E. W.; Janssen, D. B.; Fraaije, M. W. Discovery, Characterization, and Kinetic Analysis of an Alditol Oxidase from *Streptomyces coelicolor*. *J. Biol. Chem.* **2007**, *282* (28), 20283–20291.
(67) Hesslinger, C.; Fairhurst, S. A.; Sawers, G. Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in *Escherichia coli* that degrades L-threonine to propionate. *Mol. Microbiol.* **1998**, *27* (2), 477–492.
(68) Risso, C.; Van Dien, S. J.; Orloff, A.; Lovley, D. R.; Coppi, M. V. Elucidation of an Alternate Isoleucine Biosynthesis Pathway in *Geobacter sulfurreducens* . *J. Bacteriol.* **2008**, *190* (7), 2266–2274.
(69) Gerike, U.; Hough, D. W.; Russell, N. J.; Dyall-Smith, M. L.; Danson, M. J. Citrate synthase and 2-methylcitrate synthase: structural, functional and evolutionary relationships. *Microbiology* **1998**, *144* (4), 929–935.
(70) Upton, A. M.; McKinney, J. D. Role of the methylcitrate cycle in propionate metabolism and detoxification in *Mycobacterium smegmatis*. *Microbiology* **2007**, *153* (12), 3973–3982.
(71) Mak, S.; Nodwell, J. R. Actinorhodin is a redox-active antibiotic with a complex mode of action against Gram-positive cells. *Mol. Microbiol.* **2017**, *106* (4), 597–

613.
(72) Lynen, F.; Ochoa, S. Enzymes of fatty acid metabolism. *Biochim. Biophys. Acta* **1953**, *12* (1), 299–314.
(73) Wentzel, A.; Bruheim, P.; Øverby, A.; Jakobsen, Ø. M.; Sletta, H.; Omara, W. A. M.; Hodgson, D. A.; Ellingsen, T. E. Optimized submerged batch fermentation strategy for systems scale studies of metabolic switching in *Streptomyces coelicolor* A3(2). *BMC Syst. Biol.* **2012**, *6* (1), 59.
(74) Trujillo, C.; Blumenthal, A.; Marrero, J.; Rhee, K. Y.; Schnappinger, D.; Ehrt, S. Triosephosphate isomerase is dispensable in vitro yet essential for *Mycobacterium tuberculosis* to establish infection. *MBio* **2014**, *5* (2), e00085.
(75) Grüning, N.-M.; Du, D.; Keller, M. A.; Luisi, B. F.; Ralser, M. Inhibition of triosephosphate isomerase by phosphoenolpyruvate in the feedback-regulation of glycolysis. *Open Biol.* **2014**, *4* (3), 130232.
(76) Falke, D.; Fischer, M.; Sawers, R. G. Phosphate and oxygen limitation induce respiratory nitrate reductase 3 synthesis in stationary-phase mycelium of *Streptomyces coelicolor* A3(2). *Microbiology* **2016**, *162* (9), 1689–1697.
(77) Richardson, D. J. Bacterial respiration: a flexible process for a changing environment. *Microbiology* **2000**, *146* (3), 551–571.
(78) Sawers, R. G.; Falke, D.; Fischer, M. Oxygen and Nitrate Respiration in *Streptomyces coelicolor* A3(2); 2016; pp 1–40.
(79) Khodayari, A.; Maranas, C. D. A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nat. Commun.* **2016**, *7*, 13806.
(80) Sánchez, B. J.; Zhang, C.; Nilsson, A.; Lahtvee, P.-J.; Kerkhoven, E. J.; Nielsen, J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* **2017**, *13* (8), 935.
(81) Zelezniak, A.; Vowinckel, J.; Capuano, F.; Messner, C. B.; Demichev, V.; Polowsky, N.; Mülleder, M.; Kamrad, S.; Klaus, B.; Keller, M. A.; et al. Machine Learning Predicts the Yeast Metabolome from the Quantitative Proteome of Kinase Knockouts. *Cell Syst.* **2018**, *0* (0).
(82) Knappe, J.; Blaschkowski, H. P.; Grobner, P.; Schmitt, T. Pyruvate Formate-Lyase of *Escherichia coli*: the Acetyl-Enzyme Intermediate. *Eur. J. Biochem.* **2018**, *50* (1), 253–263.
(83) Sawers, G.; Watson, G. A glycyl radical solution: oxygen-dependent interconversion of pyruvate formate-lyase. *Mol. Microbiol.* **2002**, *29* (4), 945–954.
(84) Bum Kim, H.; Smith, C. P.; Micklefield, J.; Mavituna, F. Metabolic flux analysis for calcium dependent antibiotic (CDA) production in *Streptomyces coelicolor*. *Metab. Eng.* **2004**, *6* (4), 313–325.
(85) Edwards, J. S.; Palsson, B. O. Robustness Analysis of the *Escherichia coli* Metabolic Network. *Biotechnol. Prog.* **2000**, *16* (6), 927–939.
(86) Zhang, W.; Wong, K. K.; Magliozzo, R. S.; Kozarich, J. W. Inactivation of Pyruvate Formate-Lyase by Dioxygen: Defining the Mechanistic Interplay of Glycine 734 and Cysteine 419 by Rapid Freeze-Quench EPR. *Biochemistry* **2001**, *40* (13), 4123–4130.
(87) Crain, A. V; Broderick, J. B. Pyruvate Formate-lyase and Its Activation by Pyruvate Formate-lyase Activating Enzyme. *J. Biol. Chem.* **2014**, *289* (9), 5723–

5729.
(88) Zhang, Y.; Dai, Z.; Krivoruchko, A.; Chen, Y.; Siewers, V.; Nielsen, J. Functional pyruvate formate lyase pathway expressed with two different electron donors in *Saccharomyces cerevisiae* at aerobic growth. *FEMS Yeast Res.* **2015**, *15* (4).
(89) Gubbens, J.; Janus, M.; Florea, B. I.; Overkleeft, H. S.; van Wezel, G. P. Identification of glucose kinase-dependent and -independent pathways for carbon control of primary metabolism, development and antibiotic production in *Streptomyces coelicolor* by quantitative proteomics. *Mol. Microbiol.* **2012**, *86* (6), 1490–1507.
(90) Monk, J. M.; Lloyd, C. J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **2017**, *35* (10), 904–908.
(91) Kim, E. S.; Hong, H. J.; Choi, C. Y.; Cohen, S. N. Modulation of actinorhodin biosynthesis in *Streptomyces lividans* by glucose repression of afsR2 gene transcription. *J. Bacteriol.* **2001**, *183* (7), 2198–2203.
(92) Rueda, B.; Miguélez, E. M.; Hardisson, C.; Manzanal, M. B. Changes in glycogen and trehalose content of *Streptomyces brasiliensis* hyphae during growth in liquid cultures under sporulating and non-sporulating conditions. *FEMS Microbiol. Lett.* **2001**, *(1) Rueda,* (2), 181–185.
(93) Schneider, D.; Bruton, C. J.; Chater, K. F. Duplicated gene clusters suggest an interplay of glycogen and trehalose metabolism during sequential stages of aerial mycelium development in *Streptomyces coelicolor* A3(2). *Mol. Gen. Genet. MGG* **2000**, *263* (3), 543–553.
(94) Stelling, J.; Sauer, U.; Szallasi, Z.; Doyle, F. J.; Doyle, J. Robustness of Cellular Functions. *Cell* **2004**, *118* (6), 675–685.
(95) Alam, M. T.; Olin-Sandoval, V.; Stincone, A.; Keller, M. A.; Zelezniak, A.; Luisi, B. F.; Ralser, M. The self-inhibitory nature of metabolic networks and its alleviation through compartmentalization. *Nat. Commun.* **2017**, *8*, 16018.
(96) Nielsen, J.; Keasling, J. D. Engineering Cellular Metabolism. *Cell* **2016**, *164* (6), 1185–1197.
(97) Pandit, A. V.; Srinivasan, S.; Mahadevan, R. Redesigning metabolism based on orthogonality principles. *Nat. Commun.* **2017**, *8*, 15188.
(98) Hodgson, D. A. Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv. Microb. Physiol.* **2000**, *42*, 47–238.
(99) Nishiyama, T.; Hashimoto, Y.; Kusakabe, H.; Kumano, T.; Kobayashi, M. Natural low-molecular mass organic compounds with oxidase activity as organocatalysts. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (48), 17152–17157.
(100) Shin, J.-H.; Singh, A. K.; Cheon, D.-J.; Roe, J.-H. Activation of the SoxR Regulon in *Streptomyces coelicolor* by the Extracellular Form of the Pigmented Antibiotic Actinorhodin. *J. Bacteriol.* **2011**, *193* (1), 75 LP – 81.

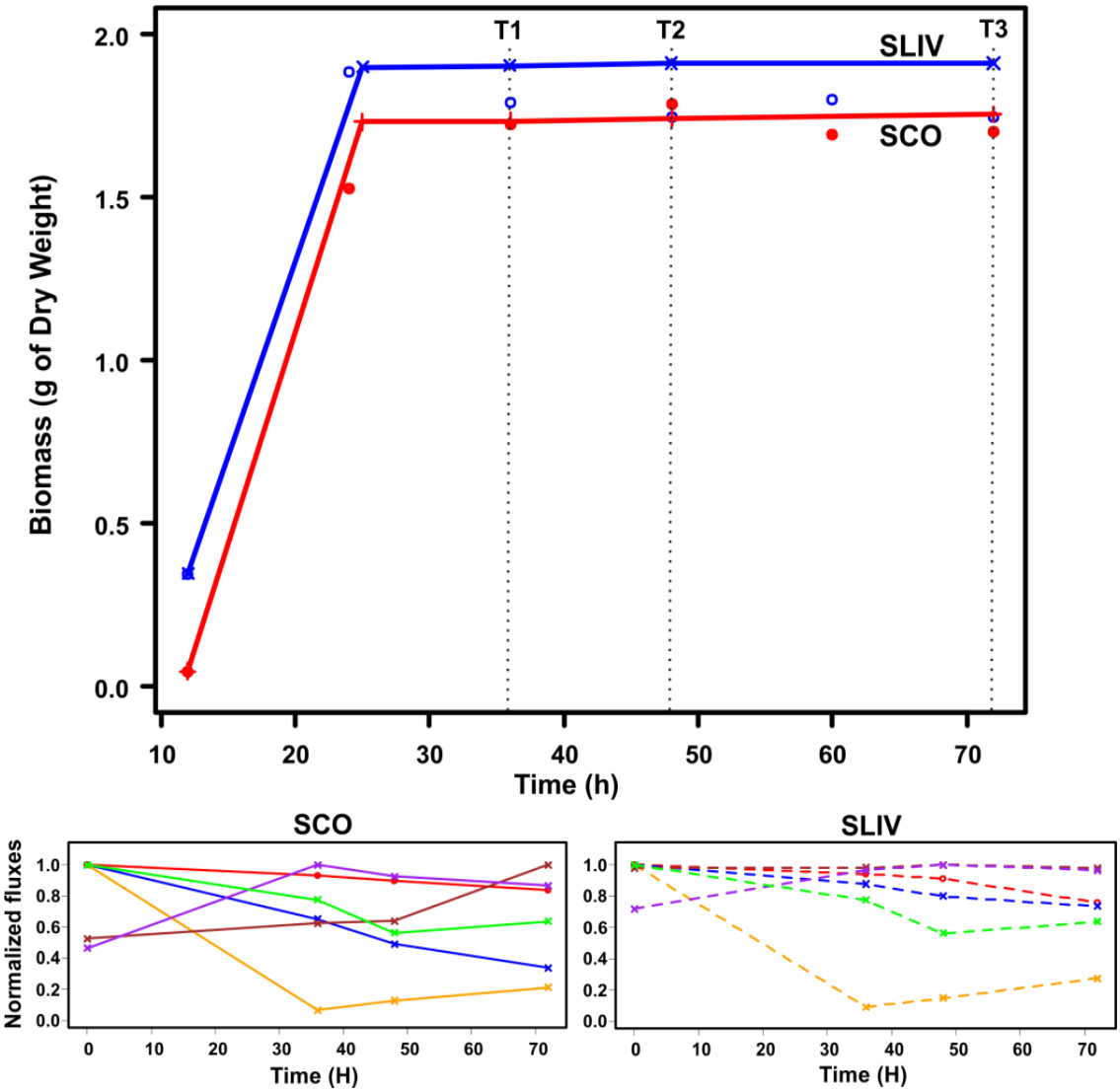
3.8 Supplementary Files

SCOIDs	Name	Feature
SCO0046	FMN-dependent NADH-azoreductase (EC 1.7.1.6)	Chromosome extremities
SCO0982	Isocitrate lyase (EC 4.1.3.1)	Same genomic island
SCO0983	Malate synthase (EC 2.3.3.9)	
SCO0985	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (EC 2.1.1.14)	
SCO4950	Respiratory nitrate reductase γ chain (EC 1.7.99.4)	Common protein complex
SCO4949	Respiratory nitrate reductase Δ chain (EC 1.7.99.4)	
SCO4948	Respiratory nitrate reductase β chain (EC 1.7.99.4)	
SCO4947	Respiratory nitrate reductase α chain (EC 1.7.99.4)	
SCO6834	Thioredoxin reductase (EC 1.8.1.9)	Largest genomic island
SCO6819	3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)	
SCO6818	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC 5.4.2.12)	
SCO6816	Phosphate ABC transporter, periplasmic phosphate-binding protein PstS (TC 3.A.1.7.1)	
SCO6815	Phosphate transport system permease protein PstA (TC 3.A.1.7.1)	
SCO6814	ABC transporter, ATP-binding protein	

Supplementary Table 3.1: Metabolic genes lost by *S. lividans* that are located in the lost genomic islands that are present in *S. coelicolor*

These genes were deleted from the *S. lividans* model and correspond to genes that are located in the lost genomic islands identified in *S. lividans* by previous comparative genomic studies (Jayapal et al., 2007). Multiple genes are involved in important primary and central metabolic functions.

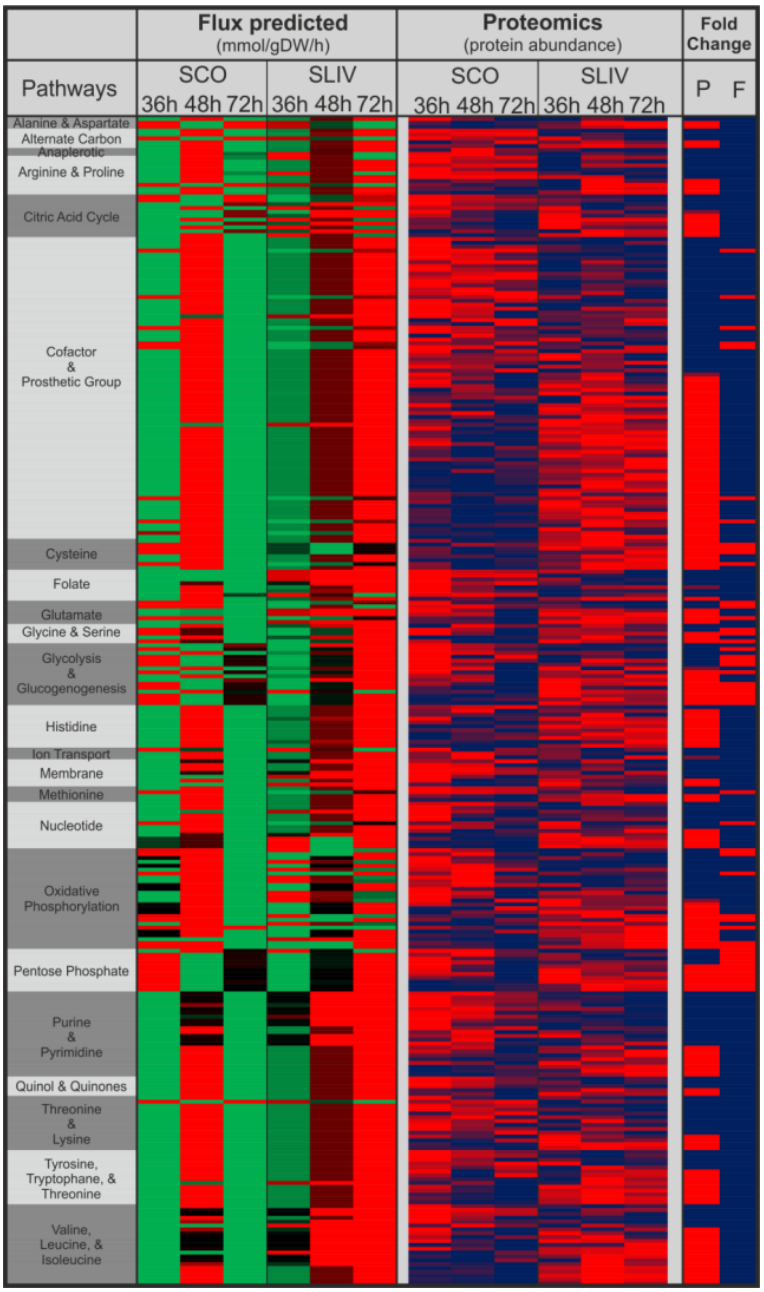
Interestingly, the phosphate transport operon (SCO6814-16) is duplicated in *S. coelicolor*. Also, there are 4 nitrate reductases operons in *S. coelicolor* while *S. lividans* only have 3 operons (SCO4947-50 is absent). Theses operons duplications in *S. coelicolor* suggests a higher oxidative metabolism in this strain compared to *S. lividans*, as *S. coelicolor* may consume more phosphate for oxidative phosphorylation and have a more active nitrate oxidative respiration.



Supplementary Figure 3.1: Predicted biomass validation and set of constraints used for the metabolic models of *S. coelicolor* and *S. lividans*

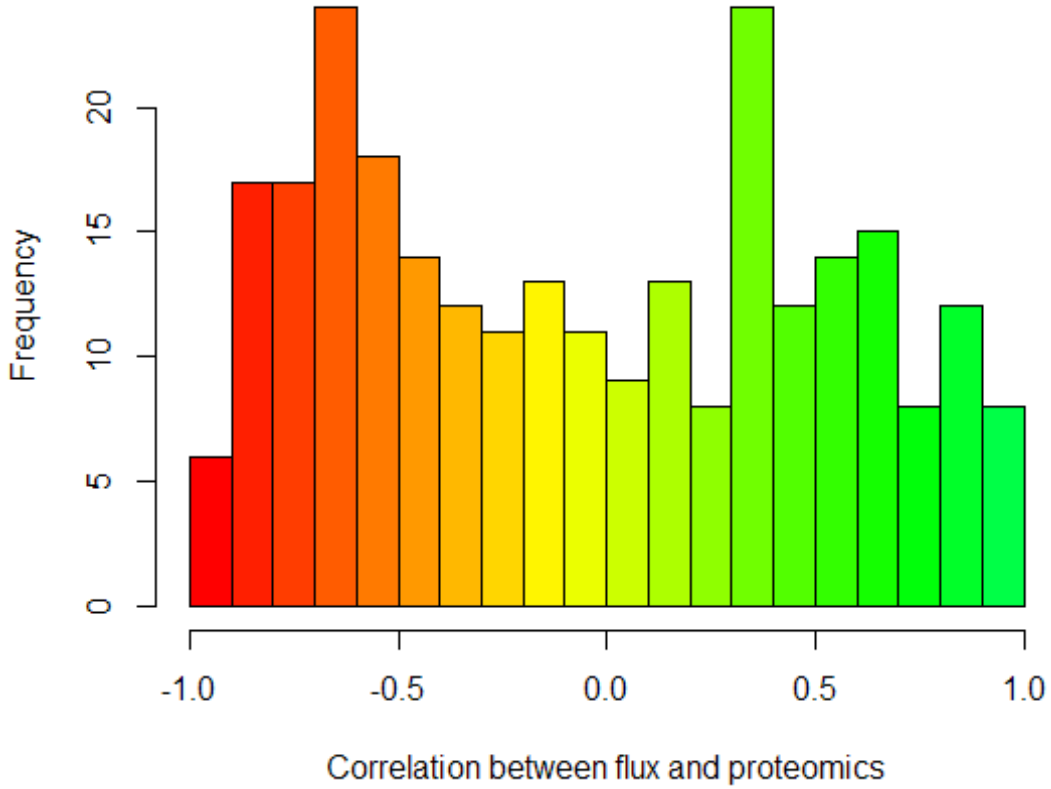
Top box: The model predictions with the iAA1259 model for *S. coelicolor* (red line) are in agreement with the observed biomass (red dots), however the last 4 points seems to be off which could be due to some metabolic shift happening in late stationary that is not predicted by the model. The model predictions with the *i*SLT1240 model for *S. lividans* (blue line) are in good agreement with the experimental data (blue dots). The times of proteomics sampling are indicated by dashed lines at 36, 48, and 72H.

Bottom left box: *S. lividans* uptakes and export fluxes used to constrain the *i*SLT1240 model
Bottom right box: *S. coelicolor* uptakes and export fluxes used to constrain the *i*AA1259 model
These constraints includes glucose (red), phosphate (orange), proline (blue), NO_2 (brown), NH_3 (purple), and NO_3 (green). The fluxes of consumption of glucose is higher in *S. lividans* (glycolytic metabolism), while the consumption of proline is higher in *S. coelicolor* (high amino-acid catabolism). Also it is interesting to note that the NO_3 fluxes are constant in *S. lividans* while they are lower in S. coelicolor and is exported in late stationary. Finally, the phophate depletion happens at the same moment in both strains.



Supplementary Figure 3.2: Heatmap of predicted fluxes compared to proteomics data

The fluxes for *S. coelicolor* and *S. lividans* were predicted over the three time points (36, 48, and 72h) and compared to the proteomics data corresponding to the enzyme catalysing the reaction. In the flux heatmap the green colour corresponds to relative high predicted flux and red correspond to relative low predicted flux between the two strains, black correspond to zero flux. In the proteomics heatmap red corresponds to the high protein concentration and dark blue corresponds to the low protein concentration over the two strains for one protein. The fold of changes of proteins and of fluxes between *S. coelicolor* and *S. lividans* were compared (high value in red and low value in dark blue). As it can be seen the fold of changes of predicted fluxes does not systematically match the fold of change of proteomics between the two strains. This suggests that the models are not able to accurately predict the metabolic difference between the two strains under these conditions. However, here the data available is in hypoxic conditions, at stationary phase, in a complex media, which is very complex to model due to the multiple regulations phenomenon occurring under these conditions as well as some uncertainty in the constraints.



Supplementary Figure 3.3: Pearson correlation between the predicted fluxes and the proteomics data

The Pearson correlation distribution of the predicted fluxes across time (36, 48, and 72h) to proteomics data shows that the overall correlation is poor. Here the mean correlation is r = -0.05. This confirms that the predictions under these conditions are not correlating with the experimental data, which suggests that the correlations are not sufficiently accurate. The colour code is a colour gradient from low correlations in red to high correlations in green.

3.8.1 Supplementary Document 1: Pyruvate formate lyase genomic analysis in *S. lividans*

There is some evidences that the Pyruvate Formate Lyase (and its activating enzyme) are present in *S. lividans* (1326 and TK24) but not in *S. coelicolor*, this is only based on a bioinformatics analysis.

Pyruvate Formate Lyase:

SSPG_07391 in *S. lividans* TK24 / SLI_0149 in *S. lividans* 1326

http://bacteria.ensembl.org/Streptomyces_lividans_1326/Gene/Summary?db=core;g=SLI_0149;r=chromosome:143590-144108;t=EOY44868

No hit in *S. coelicolor* genome with OrthoDB and Blast (using *E. coli* and *Mycobacterium tuberculosis* PFL amino-acid sequence). No enzymes were identified with high similarity to known PFL as in SLIV.

NCBI Blast Result:

BLASTP 2.8.0+ Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402. Reference for compositional score matrix adjustment: Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109. RID: UH921S9C015 Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects 170,547,179 sequences; 62,282,543,095 total letters Query= tr|A0A0E0U6Q8|A0A0E0U6Q8_ECOLX Pyruvate formate-lyase OS=Escherichia coli UMNK88 OX=696406 GN=UMNK88_4789 PE=4 SV=1 Length=765 **No significant similarity found.** Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects Posted date: Sep 21, 2018 10:53 PM Number of letters in database: 2,749,877 Number of sequences in database: 8,449 Lambda K H 0.320 0.136 0.395 Gapped Lambda K H 0.267 0.0410 0.140 Matrix: BLOSUM62 Gap Penalties: Existence: 11, Extension: 1 Number of Sequences: 8449 Number of Hits to DB: 11 Number of extensions: 0 Number of successful extensions: 0 Number of sequences better than 100: 0 Number of HSP's better than 100 without gapping: 0 Number of HSP's gapped: 0 Number of HSP's successfully gapped: 0 Length of query: 765 Length of database: 2749877 Length adjustment: 96 Effective length of query: 669 Effective length of database: 1938773 Effective search space: 1297039137 Effective search space used: 1297039137 T: 21 A: 40 X1: 16 (7.4 bits) X2: 38 (14.6 bits) X3: 64 (24.7 bits) S1: 37 (18.9 bits) S2: 50 (23.9 bits) ka-blk-alpha gapped: 1.9 ka-blk-alpha ungapped: 0.7916 ka-blk-alpha_v gapped: 42.6028 ka-blk-alpha_v ungapped: 4.96466 ka-blk-sigma gapped: 43.6362

Pyruvate Formate Lyase activating enzyme:

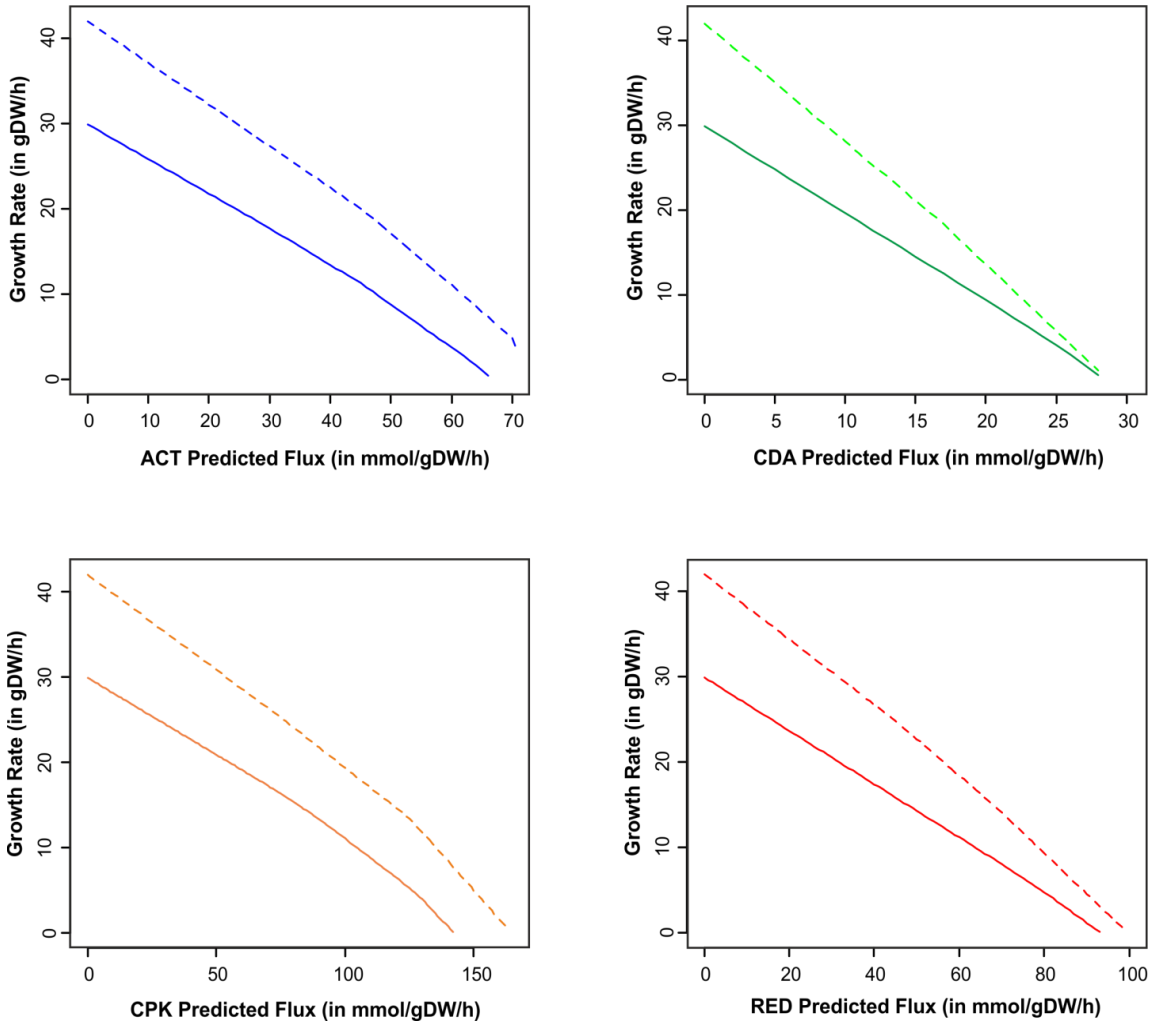
SSPG_07392 in *S. lividans* TK24 / SLI_0147 in *S. lividans* 1326

http://bacteria.ensembl.org/Streptomyces_lividans_1326/Gene/Summary?db=core;g=SLI_0147;r=chromosome:142832-143344;t=EOY44866

Once again no hit in *S. coelicolor* genome with OrthoDB and Blast (using *E. coli* and *Mycobacterium tuberculosis* PFL activating enzyme amino-acid sequence)

NCBI Blast Result:

BLASTP 2.8.0+ Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402. Reference for compositional score matrix adjustment: Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109. RID: UH8B1HNA015 Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects 170,547,179 sequences; 62,282,543,095 total letters Query= sp|P0A9N4|PFLA_ECOLI Pyruvate formate-lyase 1-activating enzyme OS=Escherichia coli (strain K12) OX=83333 GN=pflA PE=1 SV=2 Length=246 **No significant similarity found.** Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects Posted date: Sep 21, 2018 10:53 PM Number of letters in database: 2,749,877 Number of sequences in database: 8,449 Lambda K H 0.322 0.140 0.440 Gapped Lambda K H 0.267 0.0410 0.140 Matrix: BLOSUM62 Gap Penalties: Existence: 11, Extension: 1 Number of Sequences: 8449 Number of Hits to DB: 19 Number of extensions: 0 Number of successful extensions: 0 Number of sequences better than 100: 0 Number of HSP's better than 100 without gapping: 0 Number of HSP's gapped: 0 Number of HSP's successfully gapped: 0 Length of query: 246 Length of database: 2749877 Length adjustment: 86 Effective length of query: 160 Effective length of database: 2023263 Effective search space: 323722080 Effective search space used: 323722080 T: 21 A: 40 X1: 16 (7.4 bits) X2: 38 (14.6 bits) X3: 64 (24.7 bits) S1: 33 (17.3 bits) S2: 45 (21.9 bits) ka-blk-alpha gapped: 1.9 ka-blk-alpha ungapped: 0.7916 ka-blk-alpha_v gapped: 42.6028 ka-blk-alpha_v ungapped: 4.96466 ka-blk-sigma gapped: 43.6362



Supplementary Figure 3.4: Optimal antibiotics production versus optimal growth in *S. coelicolor* and in *S. lividans*

The optimal growth rates to antibiotics production were predicted with ACT (blue), CDA (green), CPK (orange), and RED (red). This was compared with *S. coelicolor* (SCO WT - solid lines) and *S. lividans* (SLIV WT - dashed lines) metabolic models. In every cases growth is in direct competition with the antibiotics production, when antibiotics production increases the optimal growth reduces.

However, it was odd to see that production was predicted as significantly higher in *S. lividans*, which has been further investigated as described in the main manuscript.

Rxn ID	Rxn Name	SCO WT	SCO + PFL	Difference
Objective	Actinorhodin Production:	66.7	71.4	4.70
o2(e)	Oxygen	-1000.00	-1000.00	0.00
glc-D(e)	D-Glucose	-1000.00	-537	-463.00
nh4(e)	Ammonium	-0.844	-0.844	0.00
pi(e)	Phosphate	-0.0747	-0.0747	0.00
so4(e)	Sulfate	-0.035	-0.035	0.00
k(e)	Potassium	-0.0186	-0.0186	0.00
fe2(e)	Ferrous iron	-0.00635	-0.0071	0.00
mg2(e)	Magnesium	-0.000825	-0.000825	0.00
fe3(e)	Ferric iron	-0.000743	0	0.00
ca2(e)	Calcium	-0.000495	-0.000495	0.00
cl(e)	Chloride	-0.000495	-0.000495	0.00
mobd(e)	Molybdene	-0.000364	-0.000364	0.00
cobalt2(e)	Cobalt	-0.000131	-0.000131	0.00
cu2(e)	Copper	-6.74E-05	-6.74E-05	0.00
mn2(e)	Manganese	-6.58E-05	-6.58E-05	0.00
zn2(e)	Zinc	-3.24E-05	-3.24E-05	0.00
ni2(e)	Nickel	-3.07E-05	-3.07E-05	0.00
dalua(c)	Dialuric acid	0.000129	0.000129	0.00
4crsol(c)	p-Cresol	0.000481	0.000481	0.00
amob(c)	S-Adenosyl-4-mt-2-oxobutanoate	0.000835	0.000835	0.00
meoh(e)	Methanol	0.000835	0.000835	0.00
5mtr(e)	5-Methylthio-D-ribose	0.00334	0.00334	0.00
5drib(c)	5'-Deoxyribose	0.00746	0.00746	0.00
gACT(e)	g-Actinorhodin	66.7	71.4	-4.70
glyc(e)	Glycerol	95.4	0.00288	95.40
for(e)	Formate	143	757	-614.00
idon-L(e)	Idonate	169	0	169.00
co2(e)	Carbone Dioxide	425	176	249.00
lac-D(e)	D-Lactate	664	0	664.00
h2o2(e)	Hydrogen Peroxide	733	678	55.00
h(e)	Hydrogen	1000.00	866	134.00
h2o(e)	Water	1000.00	1000.00	0.00

Rxn ID	Rxn Name	SLIV WT	SLIV - PFL	Difference
Objective	Actinorhodin Production:	71.4	66.5	4.90
o2(e)	Oxygen	-1000.00	-1000.00	0.00
glc-D(e)	D-Glucose	-549	-966	417.00
nh4(e)	Ammonium	-0.844	-0.844	0.00
pi(e)	Phosphate	-0.0746	-0.0746	0.00
so4(e)	Sulfate	-0.035	-0.035	0.00
k(e)	Potassium	-0.0186	0.0186	-0.04
fe2(e)	Ferrous iron	-0.0071	-0.00635	0.00
mg2(e)	Magnesium	-0.000825	-0.000825	0.00
ca2(e)	Calcium	-0.000495	-0.000495	0.00
cl(e)	Chloride	-0.000495	-0.000495	0.00
mobd(e)	Molybdene	-0.000364	-0.000364	0.00
cobalt2(e)	Cobalt	-0.000131	-0.000131	0.00
cu2(e)	Copper	-6.74E-05	-6.74E-05	0.00
mn2(e)	Manganese	-6.58E-05	-6.58E-05	0.00
zn2(e)	Zinc	-3.24E-05	-3.24E-05	0.00
ni2(e)	Nickel	-3.07E-05	-3.07E-05	0.00
fe3(e)	Ferric iron	0	-0.000743	0.00
glyc(e)	Glycerol	0	0.00175	0.00
lac-D(e)	D-Lactate	0	318	-318.00
glcn(e)	D-gluconate	0	396	-396.00
dalua(c)	Dialuric acid	0.000129	0.000129	0.00
4crsol(c)	p-Cresol	0.000481	0.000481	0.00
amob(c)	S-Adenosyl-4-mt-2-oxobutanoate	0.000835	0.000835	0.00
meoh(e)	Methanol	0.000835	0.000835	0.00
gcald(c)	Glycolaldehyde	0.00226	0.00226	0.00
5mtr(e)	5-Methylthio-D-ribose	0.00334	0.00334	0.00
5drib(c)	5'-Deoxyribose	0.00746	0.00746	0.00
gACT(e)	g-Actinorhodin	71.4	66.5	4.90
co2(e)	Carbone Dioxide	139	336	-197.00
h2o2(e)	Hydrogen Peroxide	641	734	-93.00
h(e)	Hydrogen	646	1000.00	-354.00
for(e)	Formate	867	0.00481	867.00
h2o(e)	Water	1000	1000	0.00

Supplementary Figure 3.5: Predicted exchange reactions in both models with optimized Actinorhodin production

Here the Actinorhodin production was optimized while growth rate was fixed at 0.1 gDW/h. The exchange reactions between the model and the media are represented here. The models for *S. coelicolor* and for *S. lividans* had respectively PFL added or deleted. This impacted the antibiotic production and the overall metabolism, which changed the metabolic exchanges. The exchange fluxes higher under WT conditions are in green while the ones lower are in red.

Rxn ID	Rxn Name	SCO WT	SCO - PFL	Difference
Objective	CDA Production:	28.4	30.5	2.1
glc-D(e)	D-Glucose	-1000.00	-1000.00	0.00
o2(e)	Oxygen	-1000.00	-1000.00	0.00
nh4(e)	Ammonium	-399	-427	28.00
pi(e)	Phosphate	-28.5	-30.5	2.00
so4(e)	Sulfate	-0.035	-0.035	0.00
k(e)	Potassium	-0.0186	-0.0186	0.00
fe2(e)	Ferrous iron	-0.00635	-0.00635	0.00
mg2(e)	Magnesium	-0.000825	-0.000825	0.00
fe3(e)	Ferric iron	-0.000743	-0.000743	0.00
ca2(e)	Calcium	-0.000495	-0.000495	0.00
cl(e)	Chloride	-0.000495	-0.000495	0.00
mobd(e)	Molybdene	-0.000364	-0.000364	0.00
cobalt2(e)	Cobalt	-0.000131	-0.000131	0.00
cu2(e)	Copper	-6.74E-05	-6.74E-05	0.00
mn2(e)	Manganese	-6.58E-05	-6.58E-05	0.00
zn2(e)	Zinc	-3.24E-05	-3.24E-05	0.00
ni2(e)	Nickel	-3.07E-05	-3.07E-05	0.00
idon-L(e)	Idonate	0	419	-419.00
dalua(c)	Dialuric acid	0.000129	0.000129	0.00
4crsol(c)	p-Cresol	0.000481	0.000481	0.00
amob(c)	S-Adenosyl-4-mt-2-oxobutanoate	0.000835	0.000835	0.00
meoh(e)	Methanol	0.000835	0.000835	0.00
5mtr(e)	5-Methylthio-D-ribose	0.00334	0.00334	0.00
for(e)	Formate	0.00456	492	-492.00
5drib(c)	5'-Deoxyribose	0.00746	0.00746	0.00
CDA(e)	Calcium-Dependant Antibiotic	28.4	30.5	-2.10
glyc(e)	Glycerol	103	110	-7.00
lac-D(e)	D-Lactate	328	141	187.00
glcn(e)	D-gluconate	399	0	399.00
co2(e)	Carbone Dioxide	405	200	205.00
h2o2(e)	Hydrogen Peroxide	697	586	111.00
h2o(e)	Water	1000.00	1000.00	0.00
h(e)	Hydrogen	1000.00	1000.00	0.00

Rxn ID	Rxn Name	SLIV WT	SLIV - PFL	Difference
Objective	CDA Production:	28.6	27.8	0.8
o2(e)	Oxygen	-1000.00	-1000.00	0.00
glc-D(e)	D-Glucose	-754	-1000.00	246.00
nh4(e)	Ammonium	-402	-390	-12.00
pi(e)	Phosphate	-28.7	-27.9	-0.80
so4(e)	Sulfate	-0.035	-0.035	0.00
k(e)	Potassium	-0.0186	-0.0186	0.00
fe2(e)	Ferric iron	-0.00635	-0.00635	0.00
mg2(e)	Magnesium	-0.000825	-0.000825	0.00
fe3(e)	Ferrous iron	-0.000743	-0.000743	0.00
ca2(e)	Calcium	-0.000495	-0.000495	0.00
cl(e)	Chloride	-0.000495	-0.000495	0.00
mobd(e)	Molybdene	-0.000364	-0.000364	0.00
cobalt2(e)	Cobalt	-0.000131	-0.000131	0.00
cu2(e)	Copper	-6.74E-05	-6.74E-05	0.00
mn2(e)	Manganese	-6.58E-05	-6.58E-05	0.00
zn2(e)	Zinc	-3.24E-05	-3.24E-05	0.00
ni2(e)	Nickel	-3.07E-05	-3.07E-05	0.00
h2o(e)	Water	0	205	-205.00
h(e)	Hydrogen	0.000129	0.000129	0.00
h2o2(e)	Hydrogen Peroxide	0.000481	0.000481	0.00
lac-D(e)	D-Lactate	0.000835	0.000835	0.00
co2(e)	Carbone Dioxide	0.000835	0.000835	0.00
for(e)	Formate	0.00175	0.00175	0.00
idon-L(e)	Idonate	0.00226	0.00226	0.00
CDA(e)	Calcium-Dependant Antibiotic	0.00334	0.00334	0.00
5drib(c)	5'-Deoxyribose	0.00746	0.00746	0.00
5mtr(e)	5-Methylthio-D-ribose	28.6	27.8	0.80
gcald(c)	Glycolaldehyde	177	410	-233.00
glyc(e)	Glycerol	285	0.00481	285.00
amob(c)	S-Adenosyl-4-mt-2-oxobutanoate	311	276	35.00
meoh(e)	Methanol	314	57.4	256.60
4crsol(c)	p-Cresol	716	640	76.00
dalua(c)	Dialuric acid	1000.00	1000.00	0.00
glcn(e)	D-gluconate	1000.00	1000.00	

Supplementary Figure 3.6: Predicted exchange reactions in both models with optimized Calcium-Dependent Antibiotic (CDA) production

Here the CDA production was optimized while growth rate was fixed at 0.1 gDW/h. The exchange reactions between the model and the media are represented here. The models for *S. coelicolor* and for *S. lividans* had respectively PFL added or deleted. This impacted the antibiotic production and the overall metabolism, which changed the metabolic exchanges. The exchange fluxes higher under WT conditions are in green while the ones lower are in red.

Rxn ID	Rxn Name	SCO WT	SCO - PFL	Difference
Objective	Coelimycin P1 Production:	142	164	22
glc-D(e)	D-Glucose	-1000.00	-1000.00	0.00
o2(e)	Oxygen	-1000.00	-1000.00	0.00
fe3(e)	Ferrous iron	-216	-0.313	-215.69
nh4(e)	Ammonium	-143	-165	22.00
pi(e)	Phosphate	-0.0747	-0.0747	0.00
so4(e)	Sulfate	-0.035	-0.035	0.00
k(e)	Potassium	-0.0186	-0.0186	0.00
mg2(e)	Magnesium	-0.000825	-0.000825	0.00
ca2(e)	Calcium	-0.000495	-0.000495	0.00
cl(e)	Chloride	-0.000495	-0.000495	0.00
mobd(e)	Molybdene	-0.000364	-0.000364	0.00
cobalt2(e)	Cobalt	-0.000131	-0.000131	0.00
cu2(e)	Copper	-6.74E-05	-6.74E-05	0.00
mn2(e)	Manganese	-6.58E-05	-6.58E-05	0.00
zn2(e)	Zinc	-3.24E-05	-3.24E-05	0.00
ni2(e)	Nickel	-3.07E-05	-3.07E-05	0.00
etoh(e)	Ethanol	0	137	-137.00
dalua(c)	Dialuric acid	0.000129	0.000129	0.00
4crsol(c)	p-Cresol	0.000481	0.000481	0.00
amob(c)	S-Adenosyl-4-mt-2-oxobutanoate	0.000835	0.000835	0.00
meoh(e)	Methanol	0.000835	0.000835	0.00
glyc(e)	Glycerol	0.00288	0.00288	0.00
5mtr(e)	5-Methylthio-D-ribose	0.00334	0.00334	0.00
5drib(c)	5'-Deoxyribose	0.00746	0.00746	0.00
cpkepox(e)	Coelimycin	142	164	-22.00
fe2(e)	Ferric iron	216	0.306	215.69
glcn(e)	D-gluconate	216	454	-238.00
co2(e)	Carbone Dioxide	259	-0.0995	259.10
for(e)	Formate	518	986	-468.00
h2o2(e)	Hydrogen Peroxide	716	671	45.00
lac-D(e)	D-Lactate	740	13.4	726.60
h(e)	Hydrogen	752	387	365.00
h2o(e)	Water	1000.00	1000.00	0.00

Rxn ID	Rxn Name	SLIV WT	SLIV - PFL	Difference
Objective	Coelimycin P1 Production:	164	142	22
glc-D(e)	D-Glucose	-1000.00	-1000.00	0.00
o2(e)	Oxygen	-1000.00	-1000.00	0.00
nh4(e)	Ammonium	-165	-143	-22.00
fe3(e)	Ferrous iron	-3.38	-218	214.62
co2(e)		-0.0874	259	-259.09
pi(e)	Phosphate	-0.0746	-0.0746	0.00
so4(e)	Sulfate	-0.035	-0.035	0.00
k(e)	Potassium	-0.0186	-0.0186	0.00
mg2(e)	Magnesium	-0.000825	-0.000825	0.00
ca2(e)	Calcium	-0.000495	-0.000495	0.00
cl(e)	Chloride	-0.000495	-0.000495	0.00
mobd(e)	Molybdene	-0.000364	-0.000364	0.00
cobalt2(e)	Cobalt	-0.000131	-0.000131	0.00
cu2(e)	Copper	-6.74E-05	-6.74E-05	0.00
mn2(e)	Manganese	-6.58E-05	-6.58E-05	0.00
zn2(e)	Zinc	-3.24E-05	-3.24E-05	0.00
ni2(e)	Nickel	-3.07E-05	-3.07E-05	0.00
dalua(c)	Dialuric acid	0.000129	0.000129	0.00
4crsol(c)	p-Cresol	0.000481	0.000481	0.00
amob(c)	S-Adenosyl-4-mt-2-oxobutanoate	0.000835	0.000835	0.00
meoh(e)	Methanol	0.000835	0.000835	0.00
glyc(e)	Glycerol	0.00175	0.00175	0.00
gcald(c)	Glycolaldehyde	0.00226	0.00226	0.00
5mtr(e)	5-Methylthio-D-ribose	0.00334	0.00334	0.00
5drib(c)	5'-Deoxyribose	0.00746	0.00746	0.00
fe2(e)	Ferric iron	3.37	218	-214.63
lac-D(e)	D-Lactate	12.8	740	-727.20
etoh(e)	Ethanol	137	0	137.00
cpkepox(e)	Coelimycin	164	142	22.00
h(e)	Hydrogen	388	753	-365.00
glcn(e)	D-gluconate	454	215	239.00
h2o2(e)	Hydrogen Peroxide	671	716	-45.00
for(e)	Formate	986	518	468.00
h2o(e)	Water	1000.00	1000.00	0.00

Supplementary Figure 3.7: Predicted exchange reactions in both models with optimized Coelimycin P1 (CPK) production

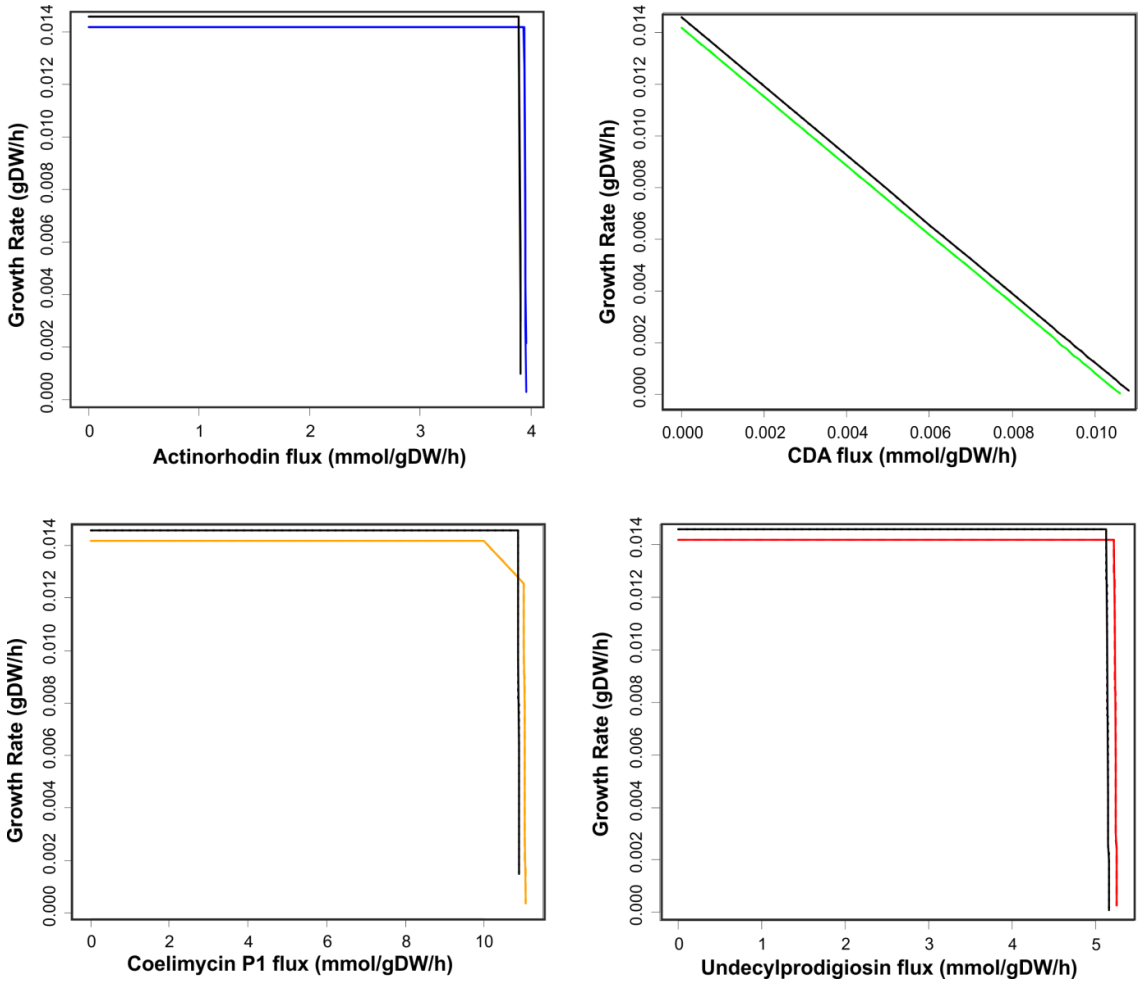
Here the CPK production was optimized while growth rate was fixed at 0.1 gDW/h. The exchange reactions between the model and the media are represented here. The models for *S. coelicolor* and for *S. lividans* had respectively PFL added or deleted. This impacted the antibiotic production and the overall metabolism, which changed the metabolic exchanges. The exchange fluxes higher under WT conditions are in green while the ones lower are in red.

Rxn ID	Rxn Name	SCO WT	SCO - PFL	Difference
Objective	Undecylprodigiosin Production:	93	101	7.30
glc-D(e)	D-Glucose	-1000.00	-1000.00	0.00
o2(e)	Oxygen	-1000.00	-1000.00	0.00
nh4(e)	Ammonium	-280	-305	25.00
pi(e)	Phosphate	-0.0747	-0.0747	0.00
so4(e)	Sulfate	-0.035	-0.035	0.00
k(e)	Potassium	-0.0186	-0.0186	0.00
fe2(e)	Ferrous iron	-0.00635	-0.0071	0.00
mg2(e)	Magnesium	-0.000825	-0.000825	0.00
fe3(e)	Ferric iron	-0.000743	0	0.00
ca2(e)	Calcium	-0.000495	-0.000495	0.00
cl(e)	Chloride	-0.000495	-0.000495	0.00
mobd(e)	Molybdene	-0.000364	-0.000364	0.00
cobalt2(e)	Cobalt	-0.000131	-0.000131	0.00
cu2(e)	Copper	-6.74E-05	-6.74E-05	0.00
mn2(e)	Manganese	-6.58E-05	-6.58E-05	0.00
zn2(e)	Zinc	-3.24E-05	-3.24E-05	0.00
ni2(e)	Nickel	-3.07E-05	-3.07E-05	0.00
idon-L(e)	Idonate	0	303	-303.00
dalua(c)	Dialuric acid	0.000129	0.000129	0.00
4crsol(c)	p-Cresol	0.000481	0.000481	0.00
amob(c)	S-Adenosyl-4-methylthio-2-oxobutanoate	0.000835	0.000835	0.00
meoh(e)	Methanol	0.000835	0.000835	0.00
5mtr(e)	5-Methylthio-D-ribose	0.00334	0.00334	0.00
for(e)	Formate	0.00456	797	-797.00
5drib(c)	5'-Deoxyribose	0.00746	0.00746	0.00
lac-D(e)	D-Lactate	60.9	0	60.90
glyc(e)	Glycerol	63.6	59.8	3.80
RED(c)	Undecylprodigiosin	93	101	-8.00
glcn(e)	D-gluconate	398	0	398.00
co2(e)	Carbone Dioxide	909	673	236.00
h2o2(e)	Hydrogen Peroxide	963	830	133.00
h2o(e)	Water	1000.00	1000.00	0.00
h(e)	Hydrogen	1000.00	1000.00	0.00

Rxn ID	Rxn Name	SLIV WT	SLIV - PFL	Difference
Objective	Undecylprodigiosin Production:	99.4	92.1	7.30
glc-D(e)	D-Glucose	-1000.00	-852	-148.00
o2(e)	Oxygen	-1000.00	-1000.00	0.00
nh4(e)	Ammonium	-299	-277	-22.00
pi(e)	Phosphate	-0.0746	-0.0746	0.00
so4(e)	Sulfate	-0.035	-0.035	0.00
k(e)	Potassium	-0.0186	-0.0186	0.00
mg2(e)	Magnesium	-0.000825	-0.000825	0.00
fe3(e)	Ferric iron	-0.000743	-0.000743	0.00
ca2(e)	Calcium	-0.000495	-0.000495	0.00
cl(e)	Chloride	-0.000495	-0.000495	0.00
mobd(e)	Molybdene	-0.000364	-0.000364	0.00
cobalt2(e)	Cobalt	-0.000131	-0.000131	0.00
cu2(e)	Copper	-6.74E-05	-6.74E-05	0.00
mn2(e)	Manganese	-6.58E-05	-6.58E-05	0.00
zn2(e)	Zinc	-3.24E-05	-3.24E-05	0.00
ni2(e)	Nickel	-3.07E-05	-3.07E-05	0.00
dalua(c)	Dialuric acid	0.000129	0.000129	0.00
4crsol(c)	p-Cresol	0.000481	0.000481	0.00
amob(c)	S-Adenosyl-4-methylthio-2-oxobutanoate	0.000835	0.000835	0.00
meoh(e)	Methanol	0.000835	0.000835	0.00
glyc(e)	Glycerol	0.00175	0.00175	0.00
gcald(c)	Glycolaldehyde	0.00226	0.00226	0.00
5mtr(e)	5-Methylthio-D-ribose	0.00334	0.00334	0.00
fe2(e)	Ferrous iron	0.00635	-0.00635	0.01
5drib(c)	5'-Deoxyribose	0.00746	0.00746	0.00
RED(c)	Undecylprodigiosin	99.4	92.1	7.30
glcn(e)	D-gluconate	361	319	42.00
co2(e)	Carbone Dioxide	645	893	-248.00
for(e)	Formate	701	0.00481	701.00
h2o2(e)	Hydrogen Peroxide	841	1.00E+03	-159.00
h2o(e)	Water	1000.00	1000.00	0.00
h(e)	Hydrogen	1000.00	1000.00	0.00

Supplementary Figure 3.8: Predicted exchange reactions in both models with optimized Undecylprodigiosin (RED) production

Here the RED production was optimized while growth rate was fixed at 0.1 gDW/h. The exchange reactions between the model and the media are represented here. The models for *S. coelicolor* and for *S. lividans* had respectively PFL added or deleted. This impacted the antibiotic production and the overall metabolism, which changed the metabolic exchanges. The exchange fluxes higher under WT conditions are in green while the ones lower are in red.



Supplementary Figure 3.9: Optimal antibiotics production versus optimal growth in *S. coelicolor* and in *S. lividans*, with or without pyruvate formate lyase in a complex media

The optimal growth rates to antibiotics production were predicted with ACT (blue), CDA (green), CPK (orange), and RED (red) for *S. coelicolor* and black for *S. lividans*. This was compared with *S. coelicolor* (SCO WT - solid lines), S. coelicolor + PFL (SCO + PFL - dotdash lines), *S. lividans* (SLIV WT - dashed lines), and *S.lividans* - PFL (SLIV - PFL - dotted lines) metabolic models. The models were constrained for a growth in in Pi-limited R2YE media using experimental uptakes (Esnault et al., 2017). Here, the presence of PFL does not seem to significantly impact neither antibiotic production nor growth (slightly with ACT). Only CDA shows a direct competition between growth and CDA production, with very similar values for both. For ACT, RED, and CPK the biomass does not seem to be in competition with production until a "breaking" point where it dramatically decreases. This shows the optimum point under these conditions where the strains can sustain optimal growth and antibiotic production (for ACT, RED, & CPK). This graph is the full version of the one with a truncated x-axis to focus on the differences (*Figure 3.11*).

3.9 Acknowledgments

We would like to thanks Aaron Millan-Oropeza and Marie-Joelle Virolle for sending us the raw quantitative data for the growth and nutrients concentrations of *S. coelicolor* and *S. lividans* published in J. Prot. Res. paper by Millan-Oropeza et al.

Chapter IV

4. Comparative analysis of the predicted metabolic capabilities of biotechnologically relevant Actinobacteria

Adam Amara, Eriko Takano, and Rainer Breitling*

Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, Faculty of Science and Engineering, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom.

***Corresponding author**: Rainer Breitling, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom, Tel.: +44161 306 5117, E-mail: rainer.breitling@manchester.ac.uk

4.1 Preface

The work carried out here was an *in silico* study, so all the experimental data used and discussed were from published or publicly available data. This is a manuscript in preparation for submission.

4.2 Abstract

Actinobacteria are found in many different ecological niches from soil to marine ecosystems. This makes Actinobacteria a very diverse phylum phylogenetically and metabolically, with many species of high biotechnological or clinical importance (e.g., *Streptomyces*, *Corynebacterium*, or *Mycobacterium*). A major resource of these organisms is their diverse secondary metabolism; particularly in the genus *Streptomyces* where most of the antibiotics on market were discovered. To learn more about the metabolism of Actinobacteria and *Streptomyces* species, a comparative metabolic model reconstruction pipeline was built to reconstruct metabolic models of

these microbes. The metabolic models of more than 50 different Actinobacteria of biotechnological interest, with a particular emphasis on *Streptomyces* species, were used to compare their predicted metabolic capabilities. The metabolic distances between these strains were analysed and compared to their phylogenetic distances. The quality of the metabolic models was estimated using key quality control criteria. Finally, the metabolic diversity of all strains was studied in terms of their predicted ability to grow in different media conditions, and differences in their core and accessory metabolism were identified.

4.3 Introduction

Actinobacteria is a phylum of Gram-positive bacteria; these microbes are found in different ecological niches, including soil-dwelling bacteria(e.g., *Streptomyces*, *Salinispora* species), marine bacteria (e.g., *Micromonospora, Micrococcus* species), and the microbiota of a diverse range of animals (e.g., *Corynebacterium*, *Mycobacterium*, *Nocardia*, *Frankia* species) [1]. These organisms are important in biotechnological, industrial, and clinical contexts [2–4]. For example, the *Streptomyces* species are responsible for the discovery of 2/3 of the antimicrobials on the market [5,6], the *Corynebacterium* species play a major role in the industrial production of amino-acids and enzymes [3,7], and *Mycobacterium* species are responsible for major human diseases such as tuberculosis [8]. The adaptation to different ecological niches makes the Actinobacteria a very diverse phylum phylogenetically [9], genetically [10,11], physiologically [1,12], metabolically [13] and morphologically [1]. Their adaptations to different environments lead to a diverse metabolism with the ability to grow using many different nutrients (e.g., polysaccharides or nucleosides) and produce diverse

secondary metabolites (e.g., polyketide antibiotics or non-ribosomal peptide siderophores). This makes the Actinobacteria an attractive phylum for the discovery and production of secondary metabolites with industrial interest. In this phylum, a large number of secondary metabolite biosynthetic gene clusters remain undiscovered [14]. The production of secondary metabolites is directly linked to primary metabolism; primary metabolism influences regulation of secondary metabolism, and provides the necessary precursors. Primary metabolism has been studied and engineered to increase or trigger secondary metabolites production in Actinobacteria [15–17]. Many of the Actinobacteria are used as chassis strains for production of valuable metabolites; such as *Streptomyces coelicolor* for antibiotics production [18], or *Corynebacterium glutamicum* for amino acid and enzyme production [3]. The host metabolism can become a limiting factor in target compound production, so choosing a chassis strain with a compatible metabolism for the target compound is critical. This approach buildup on previous studies that used metabolic modelling to help identify potential Actinomycetes chassis for secondary metabolites production [13,19]. A metabolic comparison between the Actinobacteria strains used for bioproduction could help in the decision process to select the most suitable chassis strains for particular categories of compounds.

However, there is a lack of experimental comparative studies of the global metabolism of these strains or of less studied parts of metabolism; for example, purine and pyrimidine catabolism, or alternate carbon source metabolism. In this context, the application of computational modelling methods to the Actinobacteria metabolism opens opportunities to study their global metabolic capacities and identify important metabolic differences between strains or identify potential targets to increase

secondary metabolites production. One of the main computational methods used by researchers to study the metabolism of microorganisms is constraint-based genome-scale metabolic modelling (GSMM), which enables to quickly reconstruct organism-specific metabolic networks from their genome sequence [20]. Many Actinobacteria of biotechnological interest have their full genome sequenced, so many researchers reconstructed multiple high-quality metabolic models to study the metabolism or identify gene targets to increase metabolic production. The models reconstructed includes models for *Streptomyces coelicolor* [21–24], *Streptomyces hygroscopicus* [25,26], *Streptomyces clavuligerus* [27,28], *Streptomyces ambofaciens* [29], *Streptomyces leeuwenhoeki* [30], *Mycobacterium tuberculosis* [31–33], and *Corynebacterium glutamicum* [34,35]. Many of these metabolic models present major problems of standardisation (e.g., non-standard models annotations, or absence of databases annotation, models not shared in a standard format) to allow systematic comparison. Furthermore, for many strains of biotechnological interest we do not have a metabolic model; for example, *Streptomyces albus* is widely used by the research community as a host for heterologous secondary metabolites production, but no metabolic model is available. So, to study the biotechnologically interesting Actinobacteria metabolism, it is necessary to reconstruct more metabolic models. However, building high-quality metabolic models requires manual curation, improvements and validation based on experimental data, so reconstructing a large number of strains for comparative studies is more challenging. Many automated model reconstruction tools have been published [36–40], but these often generate low quality models, and manual steps are still necessary to identify key biochemical processes not represented in databases. Comparative model reconstruction has become an alternative to generate better quality models

from high-quality metabolic models. This has proved useful to identify metabolic differences of interest between organisms [41,42]. However, these tools are not always simple to use, as they require installing many dependencies that can be obsolete if the tool was published a few years ago.

Here, a comparative metabolic model reconstruction pipeline was built to reconstruct metabolic models of Actinobacteria on the basis of the *S. coelicolor* metabolic model *i*AA1259 [21] previously reconstructed and validated (*Chapter I*). The advantage of this pipeline is that it is entirely based on COBRApy [43] and basic maintained Python and *R* dependencies, making it easy and fast to use. Metabolic models were reconstructed and compared for more than 50 different Actinobacteria with a biotechnological or research interest, to help identify major metabolic differences between these strains. The metabolic distance between the strains was analysed, and the quality of the reconstructed models was verified. The core and accessory metabolism of the Actinobacteria were determined and studied, as well as their active core metabolism. Finally, the metabolic differences between the strains were further investigated by predicting and comparing their growth ability in different media conditions.

4.4 Methods & Material

4.4.1 Reconstruction of the phylogenetic tree

The phylogenetic tree was reconstructed using the whole proteome of the strains in the "Phylogenetic Tree" tool in Pathosystems Resource Integration Center (PATRIC) [44] based on BLAST and FastTree [45,46]. The genomes of *Escherichia coli* DH5α, *Pseudomonas fluorescens* SRM1, *Micromonospora echinospora* ATCC15837,

Mycobacterium tuberculosis, and *Corynebacterium glutamicum* R were used to root the tree. The tree visualisation and branch distribution were modified in iTOL [47].

4.4.2 *Streptomyces* strains metabolic model reconstruction

The strains were reconstructed from the reference metabolic model *i*AA1259 [21] using a tailored pipeline for comparative metabolic model reconstruction in *R* and Python. In this pipeline (*Figure 4.1*), the first step is the protein comparison of the reference strain *S. coelicolor* to the predicted proteins from the genome of the strain to reconstruct. The protein comparison is carried out in the PATRIC tool using the "Proteome Comparison" tool, running bidirectional BLASTP to compare genomes based on the protein sequence [44]. The output file in .csv is then processed in an *R* script to filter the genes matching. The genes kept are encoding proteins with >40% protein sequence identity between the two strains, and if the sequence identity is >25% and <40% but it has a sequence coverage >95% the sequence is kept [48,49]. The genes filtered are matched to the reference model (StrepDB IDs for *S. coelicolor*), which is then used to match the reactions from the template model *i*AA1259 in *R*. The genes missing are used to delete the reactions respecting the gene-product-relationship. When a gene missing was essential for a reaction, the reaction was deleted from the template model, if the gene is not essential only the gene was deleted from the model. The modifications on the model were done using the COBRAPy package in Python [43]. This constitutes a draft v0, the genes and reactions deleted were investigated to identify the genes and reactions that are essential for the model to predict growth. If an essential reaction was deleted it was reintroduced in the model v1 with a gene annotation "ESSENTIAL_GENE". For the strain in reconstruction,

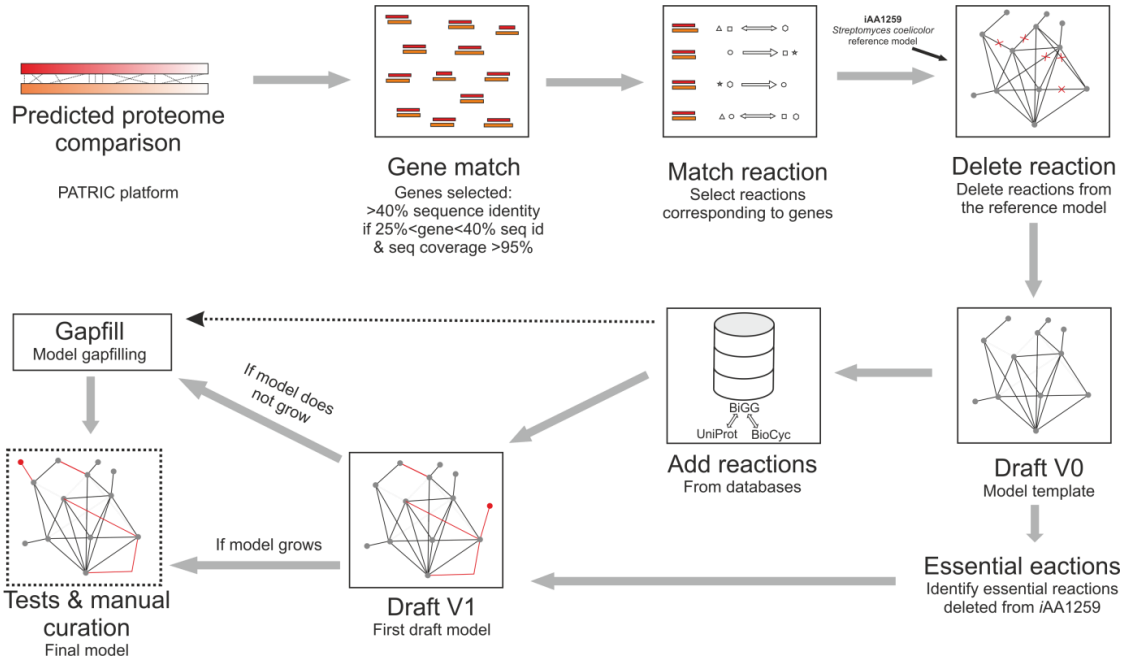


Figure 4.1: Comparative reconstruction pipeline developed for the Actinobacteria genome-scale metabolic models

The reconstruction pipeline starts with a file from the PATRIC platform for the comparison of the predicted proteome of the target organism with the reference model (i.e. *i*AA1259). This file is used to do a "gene match" generating a genes list where the homologous protein sequences corresponding to a given gene are selected based on a threshold of proteins sequence identity and coverage. The target organism genes list is used to query the reactions from the reference model corresponding to these genes. The reactions with no genes match (respecting the gene-reaction-product relationship) are deleted from the reference model, hence generating a draft model V0. This model V0 is used to add new reactions matching enzymatic genes identified in the target organism from UniProt and BioCyc database. At the same time the model is tested to verify if reactions essential for the model were deleted. If this model V1 grows it can go on further testing or manual curation. If the model does not grow, it is gapfilled by reintroducing any essential reaction delete, which was detected in the previous step.

the genes identified in Uniprot as encoding potential metabolic proteins were used to identify metabolic reactions to add to the model v1 [50]. The reactions were queried from the BiGG database based on the Uniprot and BioCyc IDs associated with the proteins [50–52]. The reactions queried from the universal BiGG model [51] were added in the draft model using COBRAPy [43]. If the model does not grow, it requires gapfilling [53]. If the model grows, then it goes on to further testing and manual curation. The model

capability to grow is tested in a minimal medium using FBA. The whole pipeline built in *R* and Python is available in the Jupyter notebook (*Electronic Supplementary 4.1*).

4.4.3 Distance and comparison methods for the metabolic models

The Euler diagram was built using the "VennDiagram" library in *R* [54]. A three-set Euler diagram was preferred to a Venn diagram, as the *S. coelicolor* area has no discrete elements it is totally included in the Streptomycetaceae area. To avoid overestimating the core-metabolism size, the reactions data used did not include the exchanges, biomass, and ATPM reactions; as these are functional reactions for the model and kept automatically.

The matrix of metabolic model reaction comparisons was built by creating a list of all the reaction IDs (pan-metabolism) and matching the model reactions to the list. The strain-matched lists of reactions were ranked from left column (with the *i*AA1259 model) to right column based on the sum of reactions added to and deleted from the reference model. The metabolic models comparison using the bubble plots was performed using the "Plotly" library in *R* [55]. The metabolic models size (total reactions number) is represented by the bubble size, the colours of the bubbles correspond to the ratio of reactions number change (reference reactions number minus reactions deleted plus reactions added) to the model size (total reactions in the model). The pairwise distances were calculated by subtraction of the number of reactions different (between two strains) to the number of reactions in common divided by the total number of reactions in the strains. The distance tree was built using the "dendextend" library in *R* [56]. First, the matrix of pairwise reactions matching between the models is converted to 1 when a reaction is present and 0 when absent. Then, from this, a

dissimilarity matrix was built with the Euclidian distance between the models. The models are then clustered using agglomerative hierarchical clustering (Ward's method), which is plotted as a dendrogram.

4.4.4 Metabolic models quality control

The published metabolic models for *S. hygroscopicus* [26], *S. clavuligerus* [27], *C. glutamicum* [35], and *M. tuberculosis* [31] were compared to the models reconstructed here for the corresponding strains and the *i*AA1259 model. The *C. glutamicum*, *M. tuberculosis*, and *S. hygroscopicus* metabolic models reactions were compared based on their IDs compliant with the BiGG standards [51]. The *S. clavuligerus* model was built not using BiGG IDs but many reactions were associated to one or multiple Enzyme Commission numbers (EC numbers), so the EC numbers were used to query the BiGG IDs from the BiGG database. However, this method also means that reactions without EC numbers associated were not taken into account. In the comparison, exchange reactions (starting by EX_ or DM_) were ignored, as well as the reactions for biomass, ATP maintenance, and the reactions for ACT, RED, CDA, and CPK (only produced by *S. coelicolor*), to reduce the matrix size and avoid overestimations/underestimations of similarities between these strains.

The number of blocked reactions were identified using the COBRApy [43] function "find_blocked_reactions" using flux variability analysis to identify reactions unable to carry flux. The exchange reactions were unconstrained to allow fluxes to in and avoid these reactions becoming bottlenecks. The essential reactions gapfilled come from the outputs of the reconstruction pipeline, the number of essential reactions are then calculated for each strain.

4.4.5 Comparative analysis of the metabolic models

The metabolic models pathways categories are based on the subsystem annotations of the individual reactions from the *i*AA1259 metabolic model [21] and the BiGG database metabolic models [51].

The growth capability of the different strains was tested by switching-off all the carbon sources import from the media, then switching on the exchange reactions one by one and optimising the growth with FBA in COBRApy [43]. The growth outputs values of the different metabolic models using different carbon sources were clustered using hierarchical clustering in *R*, then plotted as a heatmap. The cases where no model grew were ignored from the analysis.

The metabolic models core active fluxes were predicted by constraining all the models with the universal media that contain all the essential metabolites for the strains chosen. This is a similar universal media to the one used by Alam et al, to reconstruct and compare Actinomycetes [13]. The universal media composition and corresponding exchange reactions are available in the *Supplementary Table 4.1*. The metabolic fluxes are then predicted for all the metabolic models using the universal minimal media constraints with parsimonious FBA (pFBA) in COBRApy [43]. A pan-metabolic network was built by adding all the reactions present in all the models to the *S. coelicolor* network. These metabolic fluxes are compared across all the models to identify the fluxes universally active or only active in one group or another. The reactions involved in the core active metabolism and the active metabolism in 45 to 50 strains were mapped on the pan-metabolic network using Cytoscape [57].

4.5 Results & Discussion

Multiple Actinobacteria strains of biotechnological importance were selected to reconstruct their genome-scale metabolic model using the reconstruction pipeline created for this study (described in *Methods*). A total of 49 Actinobacteria metabolic models were reconstructed (*Table 4.1*); with a majority coming from the family Streptomycetaceae (31 strains, including 29 *Streptomyces* strains), and 18 other Actinobacteria strains. In the model comparisons the *S. coelicolor* (from *Chapter I*) and the *Streptomyces lividans* (from *Chapter II*) metabolic models were included. All of the Actinobacteria metabolic models described here were reconstructed on the basis of the *S. coelicolor* metabolic model *i*AA1259 [21] (as described in *Methods*).

The phylogenetic tree of the different strains (inferred on the basis of the genome sequences) has two major strains clusters, the Streptomycetaceae and the other Actinobacteria (*Figure 4.2*). As expected the strains closest to *S. coelicolor* are *Streptomyces* species, with *Kitasatospora griseola* (also called *Streptomyces griseolosporeus*) between the *Streptomyces* group and the other Actinobacteria. The other Actinobacteria are very diverse, so these are not all phylogenetically close to each other. However, there are two clear clusters associated with the *Mycobacterium* and *Corynebacterium* strains. This phylogenetic tree helps to visualise the genetic distance between the strains and will be useful to compare with the metabolic distance based on the metabolic models.

Streptomycetaceae		Other Actinobacteria
Kitasatospora griseola (*Streptomyces griseolosporeus*)	***Streptomyces coelicolor* (reference)**	*Corynebacterium casei*
Streptomyces alboflavus	*Streptomyces formicae*	*Corynebacterium glutamicum*
Streptomyces bingchenggensis	*Streptomyces fradiae*	*Corynebacterium phocae*
Streptomyces griseoflavus	*Streptomyces griseoruber*	*Frankia alni*
Streptomyces hygroscopicus	*Streptomyces griseus*	*Micromonospora echinaurantiaca*
Streptomyces ipomoeae	*Streptomyces lavendulae*	*Micromonospora rhizosphaerae*
Streptomyces lydicus	*Streptomyces leeuwenhoekii*	*Mycobacterium agri*
Streptomyces peucetius	*Streptomyces lividans*	*Mycobacterium celatum*
Streptomyces scabies	*Streptomyces rapamycinicus*	*Mycobacterium fragae*
Streptomyces afghaniensis	*Streptomyces rimosus*	*Mycobacterium smegmatis*
Streptomyces albidoflavus	*Streptomyces rimosus ATCC 10970*	*Mycobacterium tuberculosis*
Streptomyces albus	*Streptomyces roseus*	*Nocardia brasiliensis*
Streptomyces ambofaciens	*Streptomyces venezuelae*	*Planobispora rosea*
Streptomyces antibioticus	*Streptomyces violaceoruber*	*Pseudonocardia autotrophica*
Streptomyces avermitilis	*Streptomyces xiamenensis*	*Rhodococcus sp.*
Streptomyces clavuligerus	Streptomycetaceae bacterium MP113-05	*Saccharopolyspora antimicrobica*
		Saccharopolyspora flava
		Thermobispora bispora

Table 4.1: Reconstructed Actinobacteria strains

4.5.1 Metabolic and phylogenetic distances

As the GSMMs are reconstructed from a reference model (i.e., *i*AA1259) the strains with the most deleted and added reactions should be the most distant strains phylogenetically from *S. coelicolor* (i.e., *Corynebacterium*). Here, the 5 strains with the most reactions deleted and added are the three *Corynebacterium* strains and two of the *Mycobacterium* strains (*Figure 4.3*), which are part of the most phylogenetically distant strains from *S. coelicolor* (*Figure 4.2*). The opposite is also true; the closest strains to *S. coelicolor* are all Streptomycetaceae strains (*Figure 4.3*). There is an overall trend where the other Actinobacteria have more reactions deleted than the Streptomycetaceae family (*Figure 4.3*). The strain model with the least modification compared to the reference model *i*AA1259 is the *Streptomyces lividans* model; this strain is genetically very close to *S. coelicolor* and was manually curated and reconstructed on the basis of the *i*AA1259 model (see *Chapter II*). The overall trend shows that Streptomycetaceae models necessitate less modification of the reference model than other Actinobacteria (*Figure 4.3*). An interactive version of *Figure 4.3* is available in *Electronic Supplementary File 4.2 - Interactive Plot 1* to explore this plot in detail. Otherwise, there are some interesting individual cases, the three *Corynebacterium* strains are the smallest models with the most reactions deleted (*Figure 4.3*), while the *Mycobacterium smegmatis* and the *Mycobacterium tuberculosis* strains have the models with the most reactions added (*Figure 4.3*). The *Saccharopolyspora* strains are metabolically very close to the Streptomycetaceae group, despite being distant phylogenetically, while the *S. clavuligerus* strain is very close to the other Actinobacteria (*Electronic Supplementary File 4.2 - Plot 2*).

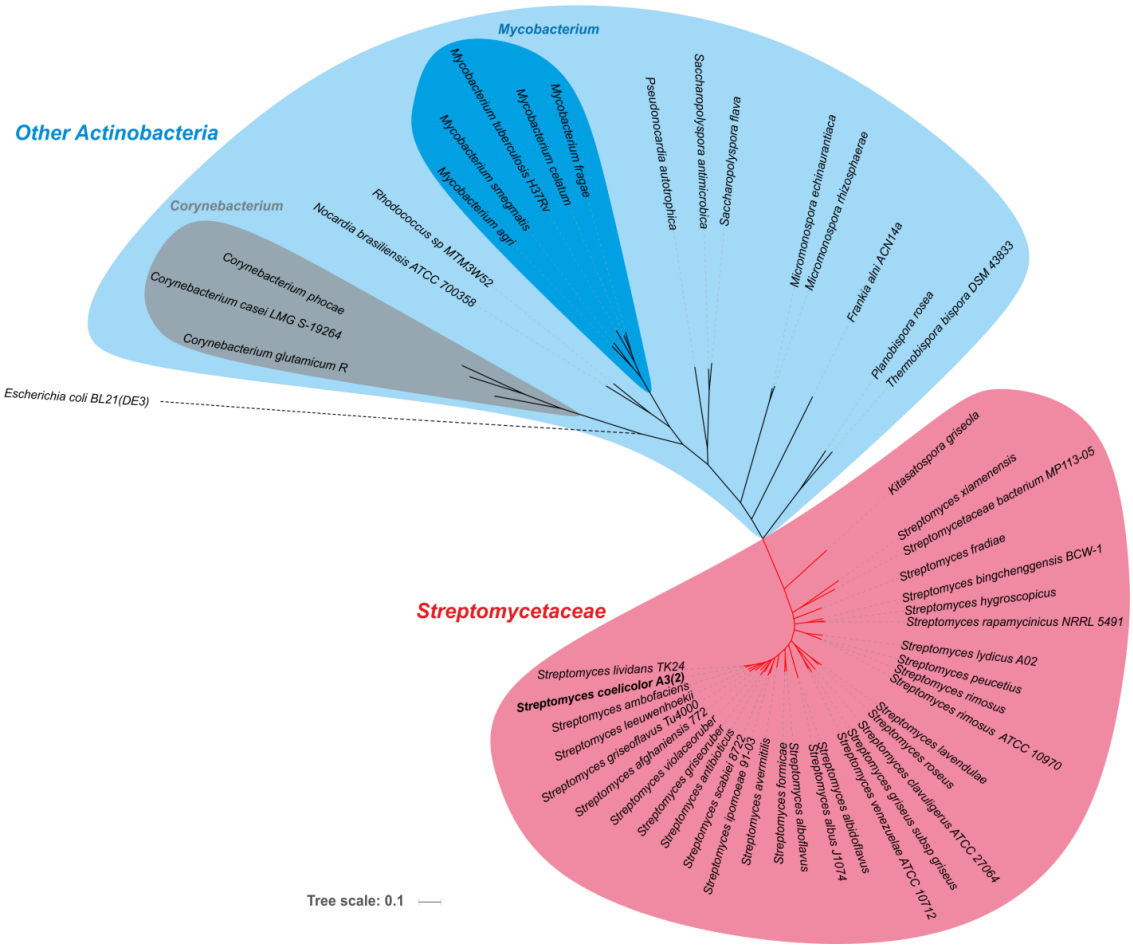


Figure 4.2: Phylogenetic Tree of the Actinobacteria strains studied here

The phylogenetic tree of the 51 Actinobacteria selected for a genome-scale metabolic model reconstruction was based on the full-genome sequence of the strains. Two main clusters are formed one for the *Streptomycetaceae* (in red) and one for the other Actinobacteria (light blue). In the other Actinobacteria there are two main clusters one of *Corynebacterium* strains (in grey) and one of *Mycobacterium* (in dark blue). The *Escherichia coli* BL21 strain was used to anchor the tree.

To further investigate the metabolic distances between the strains, the pairwise distance between all the metabolic models was calculated (*Figure 4.4*). It confirmed that the most distant strains are the *M. smegmatis*, *M. tuberculosis*, and the three *Corynebacterium* strains (*Figure 4.4*). Groups of similar strains seemed to form clusters of similar models, such as the *Corynebacterium* group. This was further investigated by generating a distance tree (*Figure 4.5*) using clustering algorithms based on the

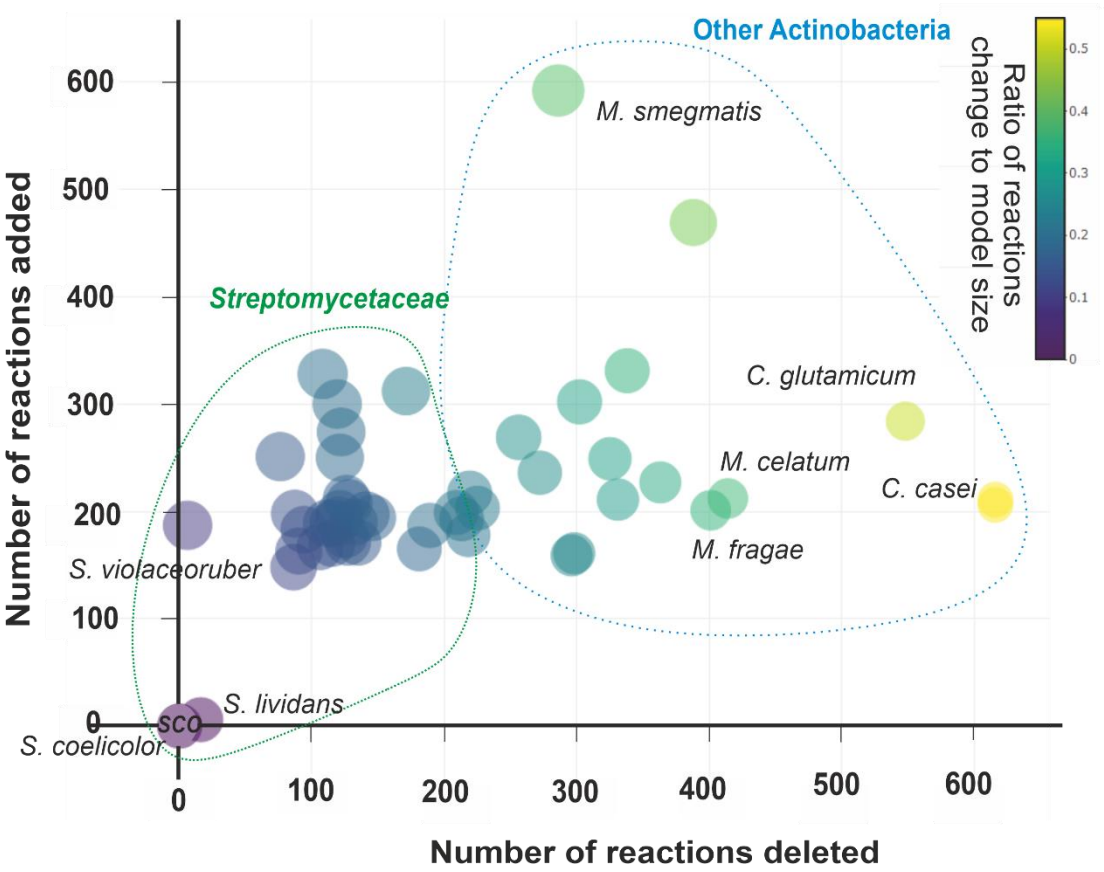


Figure 4.3: Plot of reactions added versus reactions deleted in the individual metabolic models

The metabolic models with the most reactions added and deleted are the metabolic models the furthest from the metabolic model of reference (SCO) at the origin of the coordinate plane. The *Streptomycetacea* group of strain is the closest group to *S. coelicolor* model (green dashed circle), while the furthest strains are other Actinobacteria (blue dashed circle). The circle size represents the total number of reactions in the model, and the colour correspond to the ratio of reactions change (reference reactions number minus reactions deleted plus reactions added) to the model size (total reactions in the model).

reactions in the models (see *Methods*). Some unexpected clustering happened, such as the Streptomycetaceae *bacterium* MP113-05 that clusters with other Actinobacteria like *Micromonospora* strains instead of other Streptomycetaceae strains (*Figure 4.5*). Also, all the *Corynebacterium* strains clustered in the same branches, but not the *Mycobacterium* strains that are divided into two clusters (*Figure 4.5*). This suggests that *M. smegmatis* and *M. tuberculosis* are metabolically more different from the *Mycobacterium agri*, *Mycobacterium celatum*, and *Mycobacterium fragae* than from

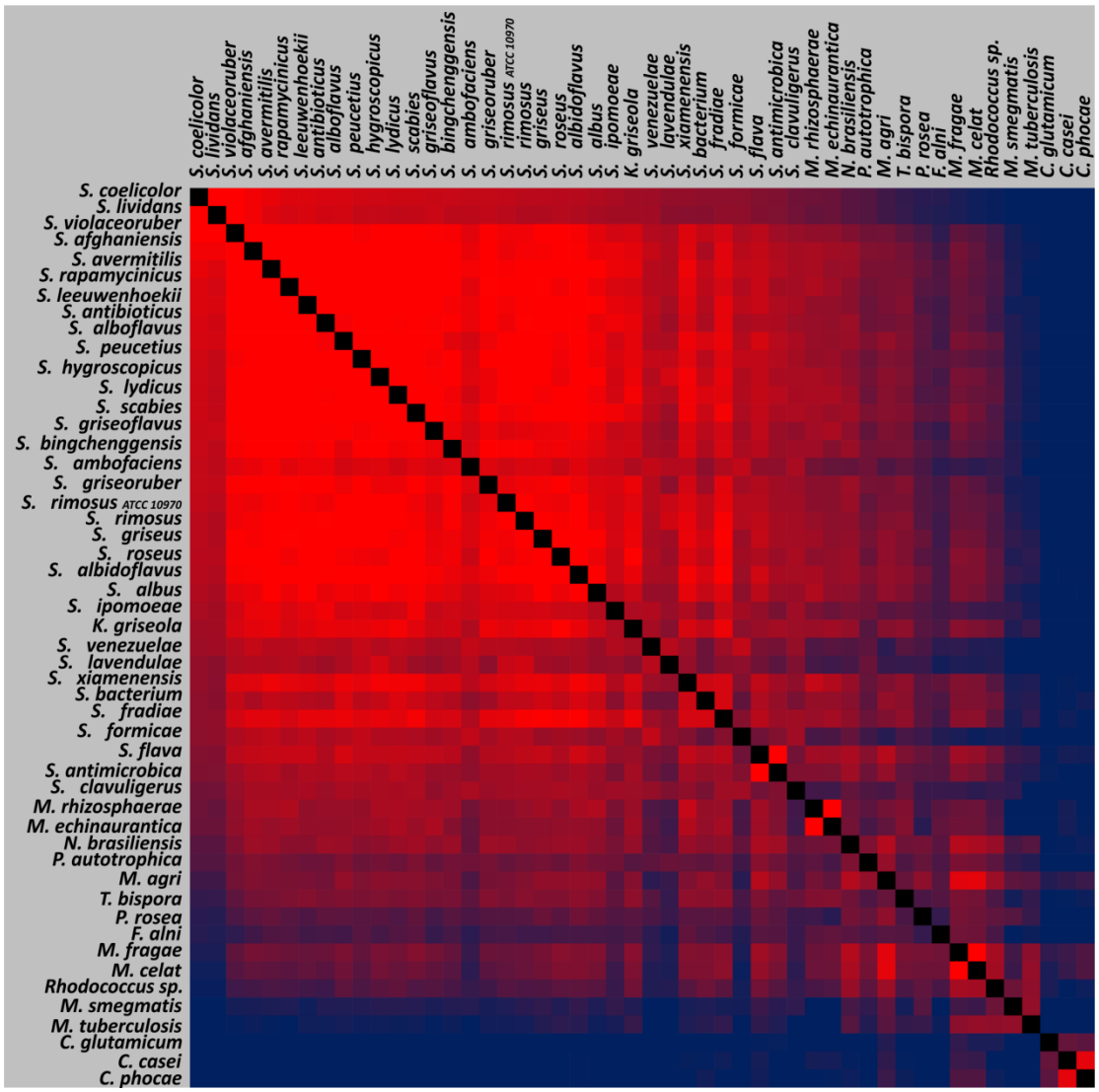


Figure 4.4: Heatmap of pairwise distance matrix of the metabolic models reconstructed based on the metabolic reactions

The distance between each pair of the strains specific metabolic models is represented by a heatmap, with red representing the smallest distance values and blue the highest distance. The pairwise distance is calculated by subtraction of the number of reactions different (between two strains) from the number of reactions in common divided by the total number of reactions in the strains. The strains are ranked from the closest strain to *S. coelicolor* to the furthest. The group of *Streptomycetacea* is the group with the smallest distance between the models (red block).

the *Corynebacterium* strains (*Figure 4.5*). This is in opposition to the phylogenetic distance analysis, where the *Mycobacterium* strains are all clustered together (*Figure 4.2*). However, these strains are the ones that had the most reactions added and/or deleted compared to the reference strain, so if these strains are too distant

phylogenetically from the *S. coelicolor* model, the comparative reconstruction pipeline might not allow to build good quality models.

4.5.2 Quality control of the metabolic models

The tree based on the metabolic models has – with a few exceptions (*S. bacterium* MP113-05, *M. smegmatis*, and *M. tuberculosis*) – an overall high similarity with the phylogenetic tree (*Figure 4.2* and *Figure 4.5*). There are two major clusters, one for *Streptomyces* strains and one for the other Actinobacteria, with the strains most distant from *S. coelicolor* being *Corynebacterium* and *Mycobacterium*. However, in the details, there are some significant differences that might help to identify the limits of the reconstruction pipeline (*Mycobacterium* separated clusters) and identify some interesting metabolic differences (*S. clavuligerus*, or *S. bacterium* MP113-05). To define the limits of the pipeline capabilities and estimate the quality of the metabolic models reconstructed, the metabolic models were compared to published manually curated metabolic models. Many models were reconstructed for Actinobacteria strains, such as *S. hygroscopicus* [25,26], *S. clavuligerus* [27,28], *S. ambofaciens* [29], *S. leeuwenhoeki* [30], *M. tuberculosis* [31–33], or *C. glutamicum* [34,35]. However, most of these models did not comply with BiGG [51] and/or MIRIAM standards [58], or the models were not shared by the authors, making them impossible or too complex to compare with the models reconstructed here. For example, many models had reactions and metabolites IDs based on in-house annotations which meant it would require the creation of a conversion table for each of the published models (this was only done for the *S. clavuligerus* model). Therefore, the comparison was mostly limited to models compliant with the BiGG standards, this included metabolic models for

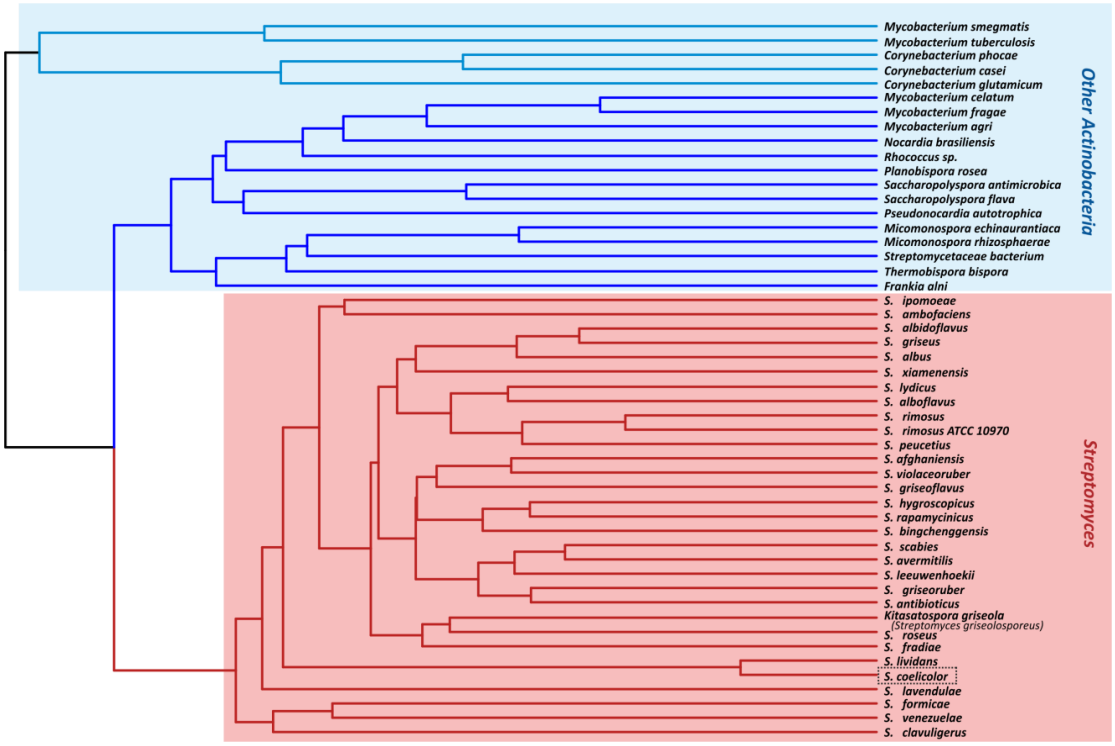


Figure 4.5: Tree based on the distance between the metabolic models

This tree represents the metabolic distance between the Actinobacteria strains studied here, which is comparable to the full-genome based phylogenetic tree (*Figure 4.1*). There are two major clusters, one corresponding to *Streptomyces* (red part) and one corresponding to the other Actinobacteria (blue part of the tree). However, the other Actinobacteria part of the tree shows two separate clusters one containing two of the *Mycobacterium* strains and all the *Corynebacterium* strains (light blue branches) and the rest of the strains (dark blue branches). The *S. coelicolor* strain metabolic model used as a reference is in the dashed box. The tree is build based on the pairwise distance between the different metabolic models.

S. hygroscopicus [26], *C. glutamicum* [35], and *M. tuberculosis* [31], as well as a converted model for *S. clavuligerus* [27] (*see Supplementary Data 4.8.1*). The comparison showed that the automatically reconstructed model for *S. hygroscopicus* had an overall 83% similarity with the manually curated and published model. The overall similarity was of 24% between the reconstructed and the published *C. glutamicum* models, 22% for *M. tuberculosis*, and 17% (45% adjusted similarity) for the *S. clavuligerus* models. The comparison of metabolic models from different sources present major challenges because of the ambiguities in models annotations [58], and in the databases used such as the Enzyme Commission (EC) numbers that are also ambiguous [59]. The details of the

analyses to reduce these ambiguities in the comparison are discussed in detail in the *Supplementary Data 4.8.1*. In summary, after filtering of ambiguities the reconstructed model for *S. hygroscopicus* model covered 94% of the published metabolic model, 74% for *S. clavuligerus*, 75% for *C. glutamicum*, and 47% for *M. tuberculosis*. However, the reconstructed model for *C. glutamicum* was almost two times bigger than the published models, because the numbers of reactions were overestimated in this organism due to the reintroduction of essential reactions described in the section below. Otherwise, the models for *M. tuberculosis* and *C. glutamicum* correspond to phylogenetically more distant strains from the reference strain, so the pipeline may not reconstruct with sufficient precision these models. This indicates that the quality is lower for more distant models such as *M. tuberculosis* and *C. glutamicum*. While for closer strains like *S. hygroscopicus* seems to reach a high level of similarity with high-quality manually curated models [26].

A good indicator of quality for genome-scale metabolic models is a low number of blocked reactions [60]. If metabolic reactions cannot carry any flux, this means the reactions are associated with dead-end metabolites, creating an unbalanced pathway due to a blocked compound [61]. This is important to take into account during simulations as it can create false-negative for gene deletion or media testing, as the metabolic pathway cannot carry flux. This is corrected in manual curation steps by filling these reactions, if there are supporting information or if these are essential for growth [60]. So, the number of blocked reactions is a relevant quality control for automated metabolic models reconstruction. Here, the blocked reactions in the metabolic models were identified using flux variable analysis [60–62]. The numbers of reactions blocked in the metabolic models were overall higher in the most distant

strains from *S. coelicolor* (*Figure 4.6a*). The strains with the lowest ratio of blocked reactions are the manually curated metabolic model of *S. lividans*. While the two models with the highest number of blocked reactions are *M. smegmatis* and *M. tuberculosis*, which are the two models that have an unexpected clustering (*Figure 4.5*). Also, the Streptomycetaceae model that seems to have the lowest quality is the *S. clavuligerus* model (*Figure 4.6a* and *Electronic Supplementary 4.2- Interactive Plot 3*). This is overall in accordance with the expectations that the more distant the strain is the lower the quality of the model will be, when using the automated comparative reconstruction pipeline.

Another useful indicator of quality in this reconstruction pipeline is the number of essential reactions "gapfilled", so reactions reintroduced in the model to ensure the model predicts growth. As in the reconstruction process reactions essential for growth that were deleted from the reference model are re-introduced in the model to ensure the model predicts growth. Lower quality models are more likely to have more essential reactions "gapfilled" (i.e., reactions reintroduced in the model). In the reconstruction pipeline, some reactions can be reintroduced in the reconstructed metabolic model if these are essential to predict growth. However, the organism might have an alternative pathway leading to the same essential biomass precursor, or this precursor might not be essential. Hence, biomass composition is critical to the metabolic model accuracy [20,60], and the pipeline does not modify the biomass as it requires experimental data to define it. As a consequence, the lower the quality of the model is, the more essential reactions are gapfilled. As previously observed, the most distant strains are the ones with the lowest quality and the most essential reactions gapfilled (*Figure 4.6b*). The strains with almost no essential reaction gapfilled are the

two closest strains phylogenetically to *S. coelicolor*; the manually curated model *S. lividans* (0 gapfilled) and the automatically reconstructed model for *S. violaceoruber* (1 gapfilled). In contrast, the strains with the most essential reactions gapfilled were the three most distant strains phylogenetically from *S. coelicolor*, the 3 *Corynebacterium* strains (*Figure 4.6b* and *Electronic Supplementary 4.2- Interactive Plot 4*). An important criterion to consider here is also the size of the genome that is also influencing the model size (*Supplementary Figure 4.1*). The trend shows that the smaller the genome, the smaller the model is, then the more likely essential reactions are deleted and gapfilled. This was previously observed with another automated reconstruction for Actinomycetes models [13]. *Corynebacterium* strains have the smallest genomes with the smallest models. However, the *M. smegmatis* and *M. tuberculosis* also have smaller genomes but bigger models than expected. These models seem to be the ones of lowest quality due to their distance to the reference metabolic model. This is likely due to the biomass reactions of the models that are based on *S. coelicolor* biomass, representing biomass accurately in metabolic models is critical for prediction accuracy and is known to be challenging [63]. The more distant models are very likely to represent a different physiology (e.g., different membrane lipids content) to the reference strain, so these are more likely to have a different biomass reaction. Hence, to reconstruct models for strains more distant phylogenetically from the reference model it would require a more thorough readjustment of the biomass.

In the application and down-stream comparative analysis, it is important to take into account that the most distant models from the reference model (e.g., *Mycobacterium* and *Corynebacterium*) are probably not an accurate representation of these organisms' metabolism.

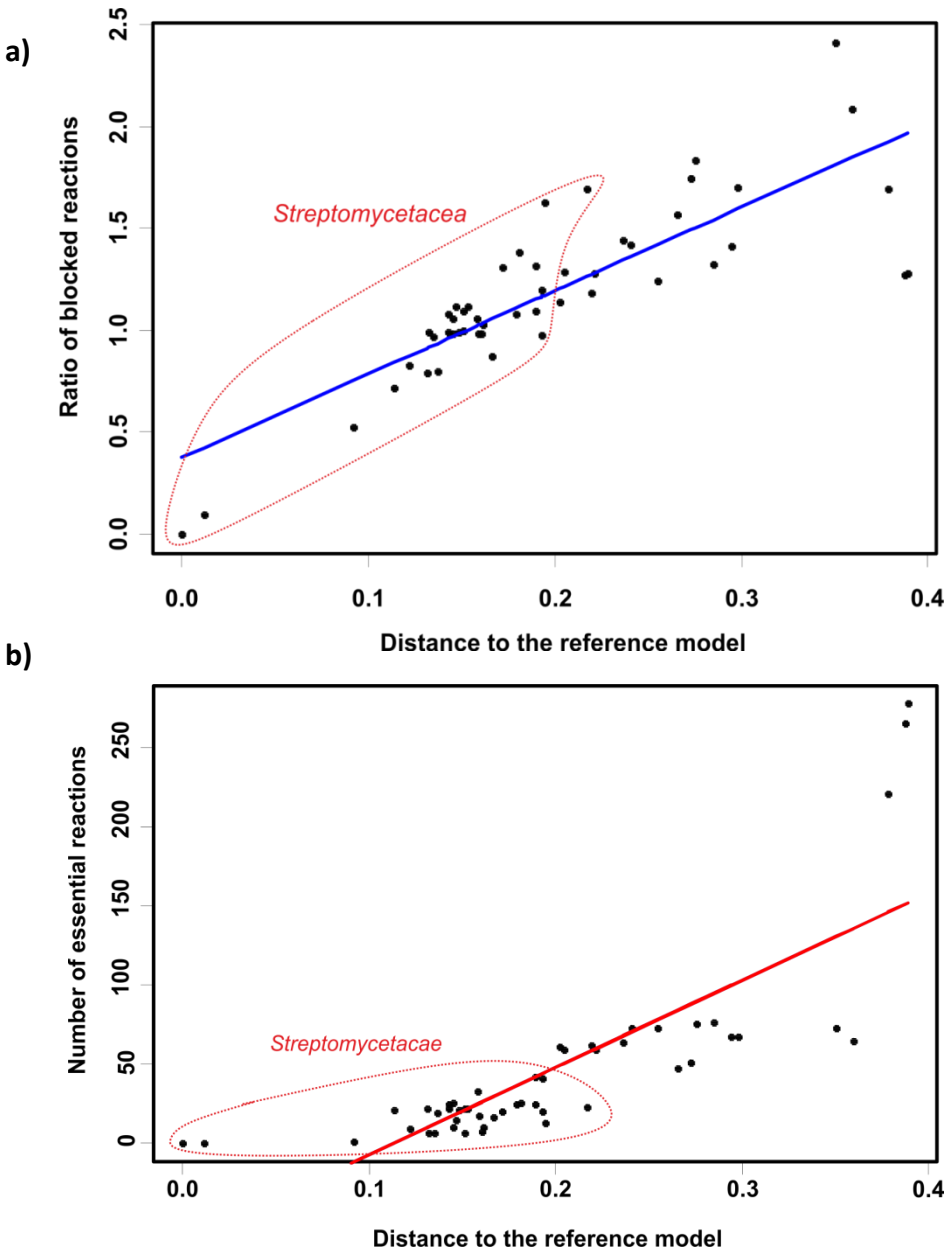


Figure 4.6: Quality checks on the metabolic models

a) Ratio of blocked reactions to the distance to the reference model. The blue line is the fitted linear equation to the data, it represents the general trend of the dataset. The *Streptomycetacea* are the closest strains (red dashed circled area) to the reference organism and the strains with the least number of blocked reactions. The outlier in the *Streptomycetacea* is the *S. clavuligerus* model.

b) Number of essential reactions gapfilled to the distance to the reference model. The red line is the fitted linear equation to the data, it represents the general trend of the dataset. The *Streptomycetacea* are the closest strains (red dashed circled area) to the reference organism and the strains with the least number of essential reactions gapfilled. The 3 Corynebacterium strains model are the ones with the most gapfilled essential reactions.

To enable identification of the individual models, these plots are available as interactive plots in *Electronic Supplementary 2- Interactive Plot 3 & Interactive Plot 4*.

4.5.3 Metabolic similarities and differences between the Actinobacteria strains

The metabolic networks of the Actinobacteria strains selected have some reactions that are universally conserved between all strains, forming the core metabolism, while the reactions only present in some strains constitute the accessory metabolism. The combination of the core and accessory metabolism creates the entire set of metabolic reactions of all the metabolic networks called the pan-metabolism. The core metabolism of these Actinobacteria represents about 63% of the pan-metabolism with 1870 reactions conserved against 2968 pan-metabolism reactions (*Figure 4.7*). The *S. coelicolor* strain does not possess any unique reactions compared to other Streptomycetaceae strains, but the Streptomycetaceae strains have 136 unique reactions compared to other Actinobacteria. The Actinobacteria have 245 unique reactions (*Figure 4.7*). There are 31 Streptomycetaceae and 18 other Actinobacteria; however, the number of unique reactions is about 45% higher in other Actinobacteria. This suggests that there might be a higher diversity of reactions across the other Actinobacteria than in the Streptomycetaceae group. This is probably due to the more homogenous group of Streptomycetaceae (with 29 *Streptomyces* spp.) whereas the other Actinobacteria include a mix of different genera such as *Mycobacteria*, *Corynebacterium*, and *Micromonospora*.

The core and accessory metabolism are visualised in a matrix representing the reactions added and deleted from the reference metabolic model (*i*AA1259) (*Figure 4.8*). There are some reactions conserved in many strains but not all, with 422 reactions (14% of the pan-metabolism) conserved in more than 45 strains, with many mostly absent from *Corynebacterium* strains (*Figure 4.8*).

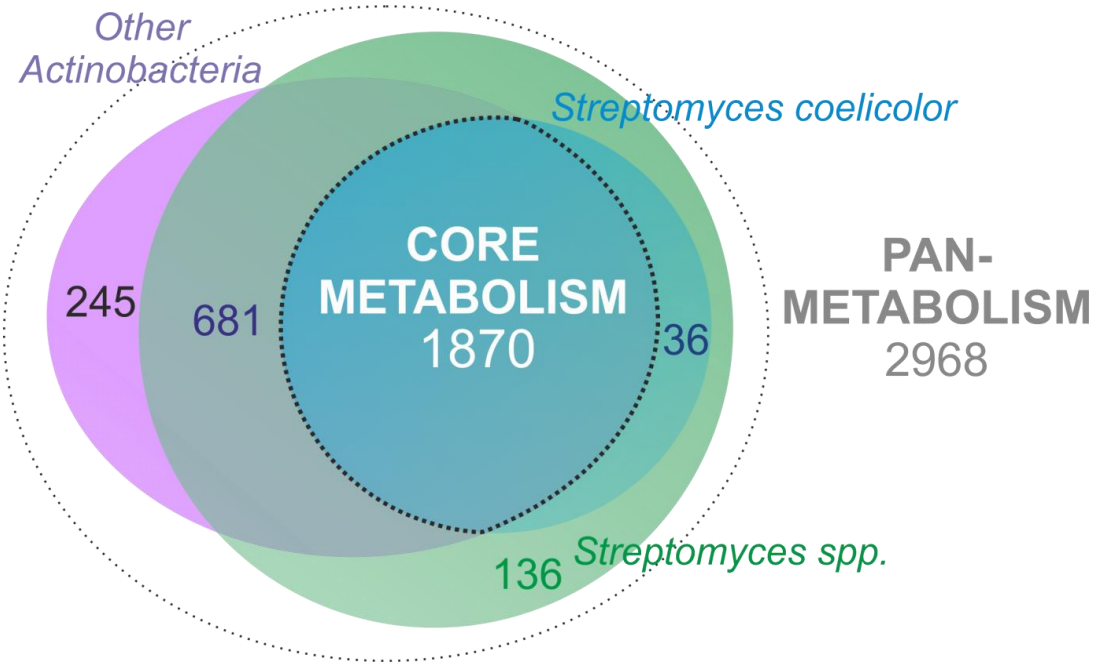


Figure 4.7: Euler diagram of the metabolic reactions overlapping *S. coelicolor*, *Streptomyces* species, and the other *Actinobacteria*.

The Euler diagram delimitates the number of reactions (1870) shared between all the strains (in blue inside the dashed black line), this represents the core metabolism of the Actinobacteria reconstructed. *Streptomyces coelicolor* does not present any unique reaction, but the *Streptomyces* species have 136 unique reactions, and the other Actinobacteria have 245 unique reactions. That forms a pan-metabolism of 2968 reaction for the 51 strains compared.

The comparison of the metabolic models aims to identify major metabolic differences between the Actinobacteria strains studied. The main metabolic processes found in the metabolic models core and accessory metabolism are the cell envelope and fatty-acids biosynthesis, and the amino-acid metabolism (*Figure 4.9*). Despite a high conservation of these pathways as many are universally conserved, there is still a high diversity as new reactions are acquired by different strains.

This is not the case for highly conserved metabolic pathways but with a lower presence in accessory metabolisms, such as the cofactor and prosthetic group biosynthesis, central metabolism, and tRNA charging (*Figure 4.9*). Cofactor metabolism and tRNA

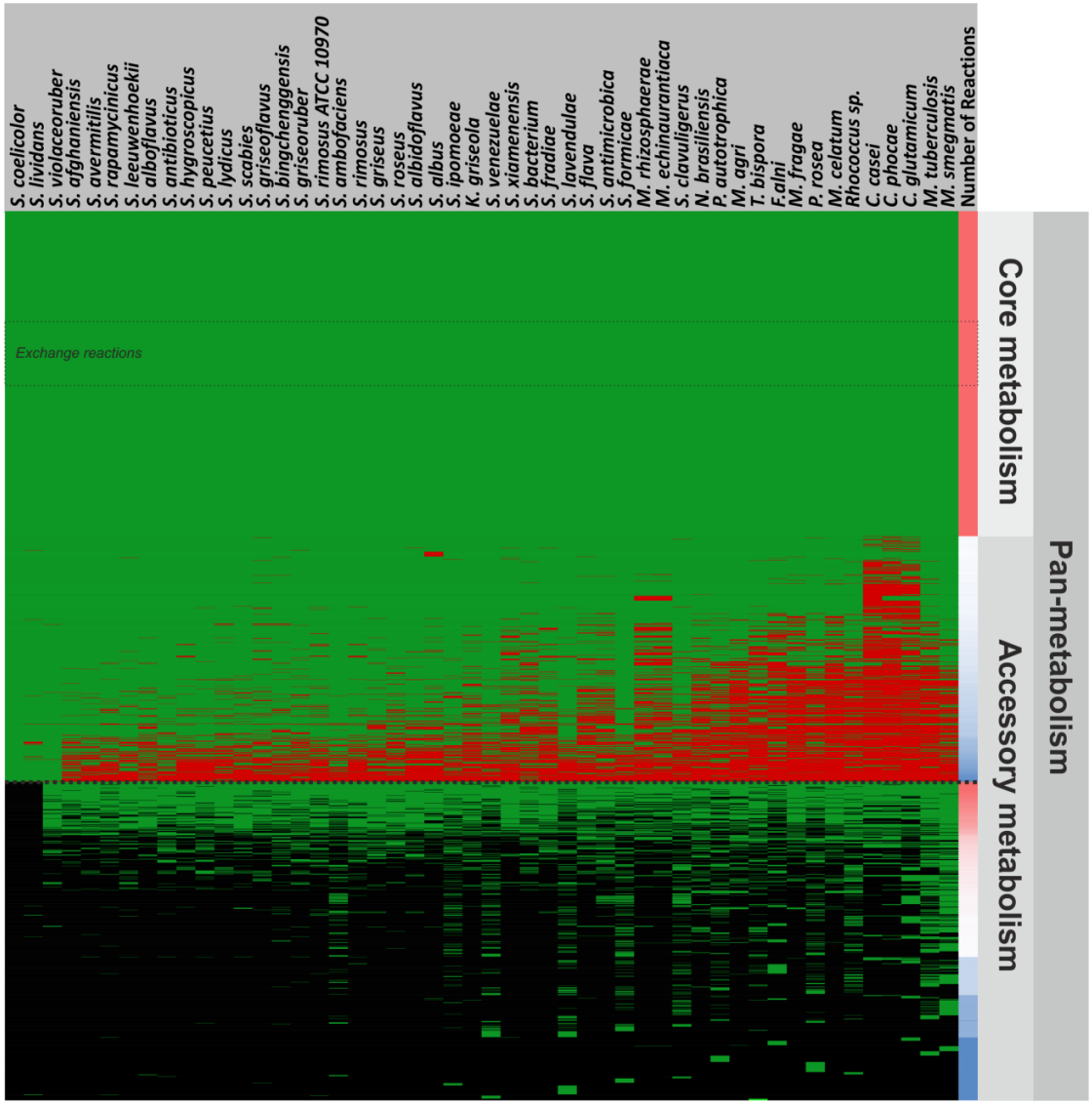


Figure 4.8: Comparison of the metabolic models based on the reactions conserved, added, or deleted from the reference model

The metabolic models' reactions were aligned with the models ordered from left closest to the reference strain *S. coelicolor* model to the most distant models on the right based on the number of reactions added + reactions deleted; the first column correspond to the *S. coelicolor* model. The green bands correspond to the reaction present in the model, the green bands below the black dashed line are reactions added in the model compared to the reference model. The black bands correspond to reactions absent from the model but present in other models, while the red bands correspond to reactions deleted from the reference model. The last heatmap column corresponds to the frequency at which a reaction is present in the models, with red for the highly frequent reactions and blue for the less frequent reactions. The reactions present in all the Actinobacteria models constitute the pan-metabolism for these strains, while the reactions not present in all the models constitute the accessory metabolism. The reactions conserved across all the strains forms the core metabolism of the Actinobacteria strains studied here. However, some of the reactions included in this visualisation and conserved are exchange reactions (including biomass and ATP maintenance) that are functional reactions for the model (to constrain the models) but do not represent metabolic functions (within dashed box).

charging were shown to be highly conserved across prokaryote species and to be essential for growth [64]. Otherwise, the alternate carbon metabolisms and secondary metabolisms are the metabolic pathways with the highest diversity as these are mostly present in accessory metabolism (*Figure 4.9*). It was shown that secondary metabolism in Actinobacteria is highly diverse [14], but here the secondary metabolism size seems quite small. This is due to the metabolic databases used for the automated metabolic modelling reconstruction that do not contain many secondary metabolic pathways so these cannot be introduced automatically [65]. The increase in alternate carbon metabolism in the accessory metabolism is likely due to the different adaptation of the strain to different nutrients (e.g., degradation of polysaccharides), as Actinobacteria have diverse ecological niches [15,66].

The metabolic adaptations of Actinobacteria to different ecological niches from soil to human microbiota lead to the ability to growth on different substrates. The metabolic models are able to grow on different sole carbon sources depending on the strains (*Figure 4.10*). There are two major groups once again corresponding to Streptomycetaceae and other Actinobacteria. The Streptomycetaceae have a clear capacity to grow with more diverse carbon-sources, from amino-acids to polysaccharides. Whereas other Actinobacteria are able to use fewer carbon sources (*Figure 4.10*). For example, the *C. casei* metabolic model can grow on only 5 substrates whereas *S. coelicolor* can grow on 134 reactions. This shows a major metabolic and phenotypic difference between *Streptomyces* species and other Actinobacteria with a capacity to grow with diverse carbon-sources. The importance of carbon sources has been repeatedly shown in the regulation of secondary metabolism in *Streptomyces* [16,67–69]. This capacity to grow on the different substrate is likely to originate from the

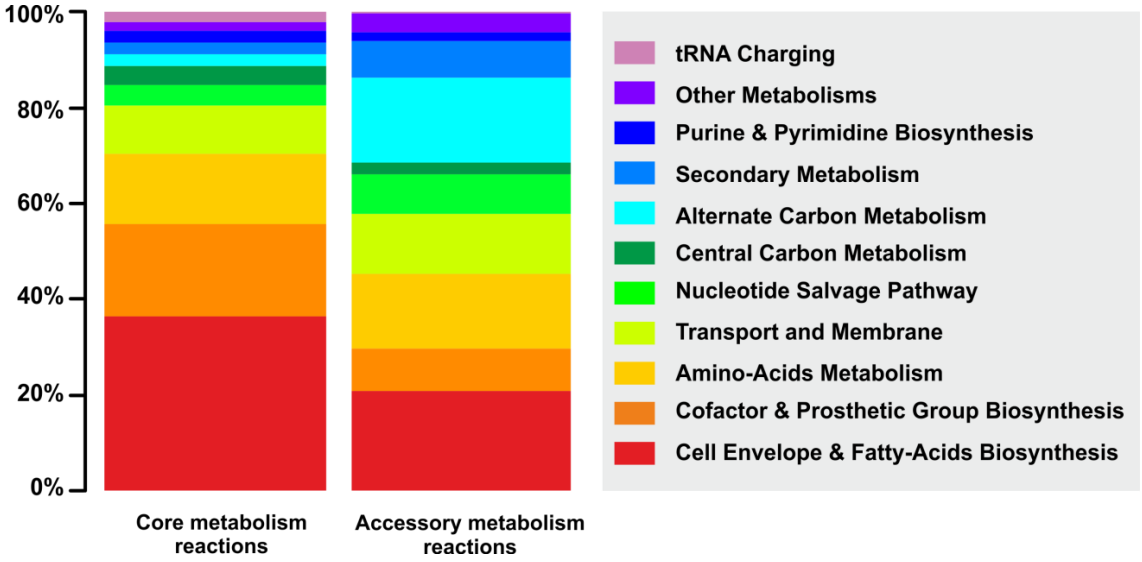


Figure 4.9: Metabolic pathways distribution for the reactions involved in the core metabolism and in the accessory metabolism

The share of reactions (in percent) per subsystems categories (i.e., metabolic pathways) are represented by different colours. The metabolic pathways come from the reference model (*i*AA1259) and the BiGG database models associated with these reactions. The shares of reactions are determined for the universally conserved reactions metabolic models (i.e., core metabolism) and for the non-core reactions (i.e., accessory metabolism).
This figure ignores the reactions with unassigned pathways; the figure with the unassigned pathways version is available in *Supplementary Figure 2.*

Streptomyces complex ecological niches (e.g., soil) with the need to compete for nutrients [15]. However, there are some differences within the Streptomycetaceae group, with two groups, one with *Streptomyces* strains able to grow using more purines and pyrimidine compounds and another with strains unable to do so (*Figure 4.10*). Most of the Streptomycetaceae can grow using amino-acids, monosaccharides, and polysaccharides. At the same time, the predictions on media growth might give an insight into the quality of the models. The models unable to predict growth on more than 5 carbon sources are the models of the strains most distant phylogenetically from *S. coelicolor*. A possible cause is absent or incomplete alternative metabolic pathways in the models necessary to degrade carbon sources. This is complicated to test as most of the Actinobacteria do not have systematic

growth data on unique carbon sources. However, in the case of the more studied strains, *C. glutamicum* [35] and *M. tuberculosis* [32], these are frequently tested with multiple carbon sources and were shown to need multiple carbon sources. For example, *M. tuberculosis* requires glucose or glycerol supplementation to degrade any other carbon sources [32]. If other Actinobacteria require multiple carbon sources to grow, this could also explain the inability to predict growth on unique carbon sources for other Actinobacteria models compared to *Streptomyces* (*Figure 4.10*) – in this case, the observed differences would not indicate a lack of model quality, but a real physiological difference.

Between the two subgroups of Streptomycetaceae, one is more capable to grow on purines and pyrimidines (*Figure 4.10*). The metabolic models of these two groups have only 1 enzyme absent from all the strains unable to grow on these substrates; this is a phosphopentosemutase (PPM) involved in the purine and pyrimidine salvage pathway. The PPM is a phosphotransferase converting D-ribose 1-phosphate to D-ribose 5-phosphate, and 2-deoxy-D-ribose 1-phosphate to 2-deoxy-D-ribose 5-phosphate [70]. The PPM connects the nucleotide salvage pathway to the central carbon metabolism independently from the pentose phosphate pathway [71]. This enables the use of the ribose moiety in the nucleoside into the central metabolism. Mutant strains of *Escherichia coli* were not able to grow on purine and pyrimidines as substrates when the PPM enzymes were knocked out [72,73]. The gene associated with the PPM enzymes in *S. coelicolor* (SCO7443) is the same gene encoding for the phosphoglucomutase. A study deleting the SCO7443 gene showed that it resulted in a decrease of actinorhodin production while an increase was expected as it was supposed to cut fluxes toward glycogen [17]. The study did not take into account the

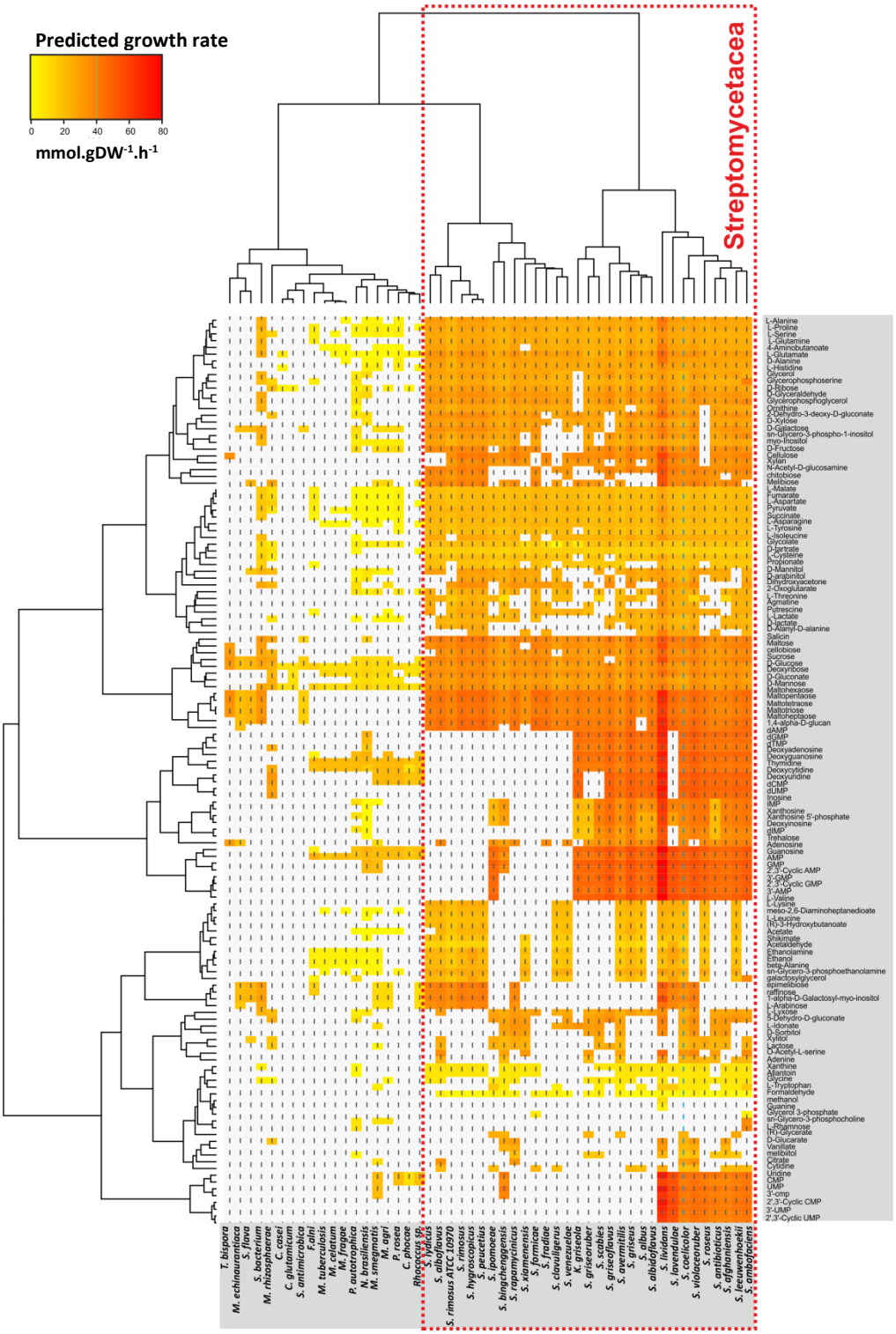


Figure 4.10: Growth capabilities of the different strains' metabolic models

The metabolic models were optimised using as sole carbon source the compounds listed on the right side of the heatmap, if a model grow it has a coloured cell, if it does not grow it is a white cell. The rows correspond to the metabolites used as a carbon source, while the columns represent the strains ability to grow on this source. If the strain is not predicted as growing the colour is white, then the colour goes from lowest growth rate in yellow to highest growth rate in red. The strains and metabolites are clustered by rows and columns using hierarchical clustering. The *Streptomycetacea* group clustered in two sub-groups highlighted with a red dashed box.

impact PPM activity of this enzyme that would impact the nucleosides catabolism. Hence, this enzyme is an interesting target to study in the *Streptomyces* species. The dynamics in the metabolism are not represented in the comparison of the metabolic models based on reactions, so the metabolic models were used to compare predicted metabolic fluxes (*Figure 4.11*). The models were optimised on a universal medium (see *Methods*). Based on the metabolic fluxes the strains clustered once again in Streptomycetaceae and other Actinobacteria (*Supplementary Figure 4.3*). There is also a clear core minimum active metabolism, where reactions carry fluxes in all the models under these minimal conditions. This core active metabolism was visualised within the pan-metabolic network of Actinobacteria under minimal conditions (*Figure 4.11*). This core minimum metabolism represents 46% of all the active reactions and 14% of the pan-metabolism. The main metabolic pathways represented in the core active metabolic reactions are cell envelope and lipid biosynthesis, cofactor and prosthetic group biosynthesis, and amino-acids metabolism (*Supplementary Table 4.2*).

The core active metabolism is a representation of the minimal metabolism of the Actinobacteria, containing the key enzymes for metabolic processes necessary for growth under minimal conditions. The core active metabolism represents only a small portion of the pan-metabolism, the rest of the metabolism probably has an importance in the adaptation to other environmental conditions. This is confirmed in the strains capability to grow on different substrates (*Figure 4.10*) and the diversity of metabolic pathways found in the accessory metabolism (*Figure 4.9* and *Supplementary Table 4.2*). The metabolic pathways mostly utilised in Streptomycetaceae but not in the other Actinobacteria are mostly involved in transport, and catabolism of amino-acid and nucleotides, as well as oxidative metabolism (*Supplementary Figure 4.4*). The

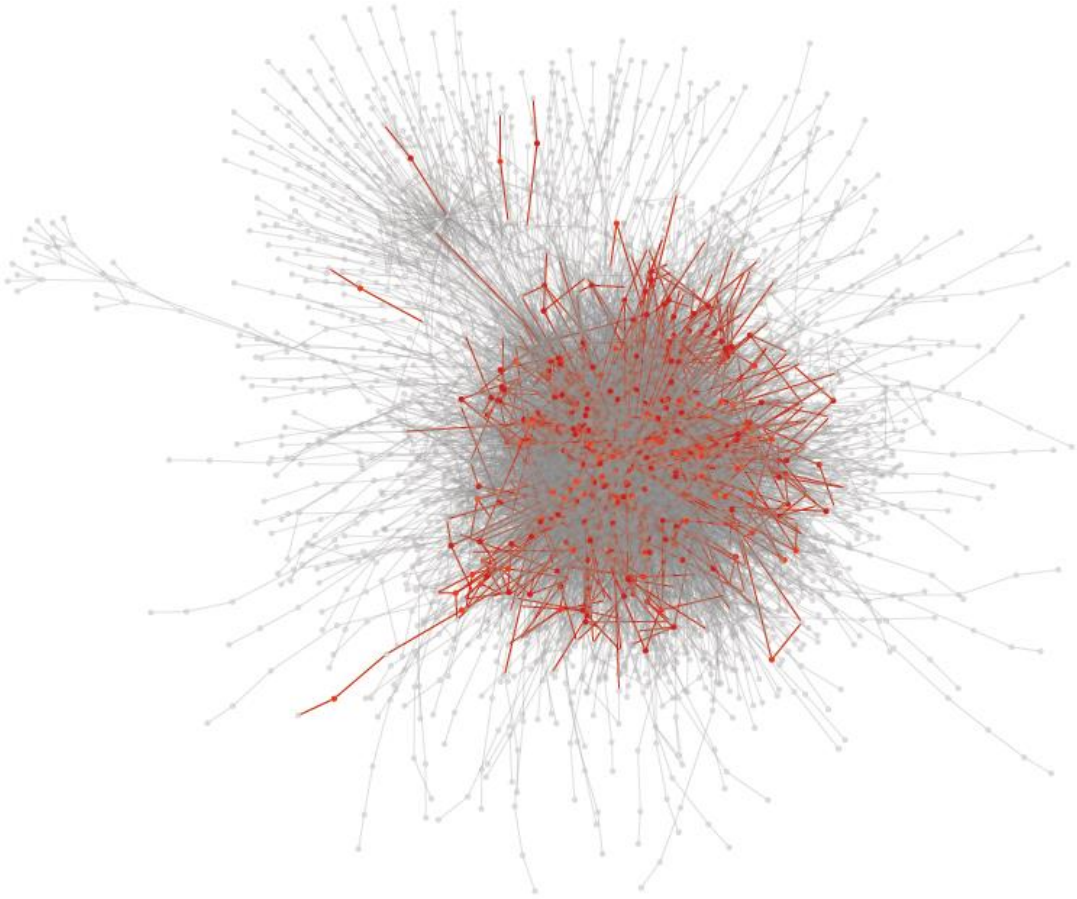


Figure 4.11: Core metabolic model fluxes in the pan-metabolic network

The pan-metabolic network is represented here, with the reactions as the edges and metabolites as nodes. The core metabolic fluxes (reactions carrying a flux in all the metabolic models in red here). The reactions carrying fluxes in at least 45 models (85% of the strains) are in orange. The reactions and metabolites involved in the accessory metabolism under these minimal conditions are in grey in the network. A representation of this network using an edge-weighted layout is available in the *Supplementary Figure 6*.

other Actinobacteria are more actively using some of the cofactor and prosthetic group pathways involved with purines and pyrimidines metabolism, glycolysis and trehalose metabolism, as well as other amino-acids pathways (*Supplementary Figure 4.5*). This is linked to the differences in nutrients usage capacities of the Streptomycetaceae compared to the other Actinobacteria (*Figure 4.10*). Interestingly, the Streptomycetaceae are using the nucleotide salvage pathway passing by the phosphoglucomutase to go into glycolysis, while the other Actinobacteria used glucose directly.

4.6 Conclusion

The development of an easy to use comparative metabolic model reconstruction pipeline helped to compare the metabolism of more than 50 Actinobacteria strains used in industrial and academic research but that were not previously studied using metabolic modelling. The quality of these models was checked using key quality controls for metabolic models (e.g., blocked reactions, or “gapfilled” reactions). The metabolic distance between the strains models were analysed in details, showing an overall similar metabolism between *Streptomyces* strains (and more widely Streptomycetaceae), but not with other Actinobacteria. These observations were reinforced by comparing their growth capabilities that showed the capability of Streptomycetaceae to grow in theory on more diverse nutrients (i.e., polysaccharides) than other Actinobacteria. Also, this highlighted two different groups in the Streptomycetaceae with only one of the two groups able to grow on purines and pyrimidine. Finally, the core and accessory metabolism of the Actinobacteria were determined, with a large core metabolic network (46% of the pan-metabolism) and a small active core metabolic network (14% of the pan-metabolism).

These automatically reconstructed metabolic models can now be used as a base to easily reconstruct high-quality metabolic model; by manually curating the strain-specific modifications (e.g. introduce secondary metabolites pathways or modify biomass) and validating them with available data. These models can be used in the future to compare the production capabilities of different secondary metabolites in different Actinobacteria, to identify the best chassis strain for a given compound.

4.7 References

(1) Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; van Wezel, G. P. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **2016**, *80* (1), 1 LP – 43.

(2) Manivasagan, P.; Venkatesan, J.; Sivakumar, K.; Kim, S.-K. Pharmaceutically Active Secondary Metabolites of Marine Actinobacteria. *Microbiol. Res.* **2014**, *169* (4), 262–278.

(3) Tatsumi, N.; Inui, M. Corynebacterium Glutamicum*: Biology and Biotechnology*; Springer Science & Business Media, 2012; Vol. 23.

(4) Watve, M.; Tickoo, R.; Jog, M.; Bhole, B. How Many Antibiotics Are Produced by the Genus Streptomyces ? *Arch. Microbiol.* **2001**, *176* (5), 386–390.

(5) Hopwood, D. A. Streptomyces *in Nature and Medicine : The Antibiotic Makers*; Oxford University Press, 2007.

(6) Watve, M. G.; Tickoo, R.; Jog, M. M.; Bhole, B. D. How Many Antibiotics Are Produced by the Genus *Streptomyces*? *Arch. Microbiol.* **2001**, *176* (5), 386–390.

(7) Ikeda, M.; Nakagawa, S. The Corynebacterium Glutamicum Genome: Features and Impacts on Biotechnological Processes. *Appl. Microbiol. Biotechnol.* **2003**, *62* (2–3), 99–109.

(8) Gagneux, S.; Small, P. M. Global Phylogeography of Mycobacterium Tuberculosis and Implications for Tuberculosis Product Development. *Lancet Infect. Dis.* **2007**, *7* (5), 328–337.

(9) Verma, M.; Lal, D.; Kaur, J.; Saxena, A.; Kaur, J.; Anand, S.; Lal, R. Phylogenetic Analyses of Phylum Actinobacteria Based on Whole Genome Sequences. *Res. Microbiol.* **2013**, *164* (7), 718–728.

(10) Doroghazi, J. R.; Metcalf, W. W. Comparative Genomics of Actinomycetes with a Focus on Natural Product Biosynthetic Genes. *BMC Genomics* **2013**, *14* (1), 611.

(11) Kim, J.-N.; Kim, Y.; Jeong, Y.; Roe, J.-H.; Kim, B.-G.; Cho, B.-K. Comparative Genomics Reveals the Core and Accessory Genomes of <i>Streptomyces<\i> Species. *J. Microbiol. Biotechnol.* **2015**, *25* (10), 1599–1605.

(12) Genilloud, O. Physiology of Actinobacteria. In *Biology and Biotechnology of Actinobacteria*; Springer International Publishing: Cham, 2017; pp 151–180.

(13) Alam, M. T.; Medema, M. H.; Takano, E.; Breitling, R. Comparative Genome-scale Metabolic Modeling of Actinomycetes: The Topology of Essential Core Metabolism. *FEBS Lett.* **2011**, *585* (14), 2389–2394.

(14) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Brown, L. C. W.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **2014**, *158* (2), 412–421.

(15) Hodgson, D. A. Primary Metabolism and Its Control in Streptomycetes: A Most Unusual Group of Bacteria. *Adv. Microb. Physiol.* **2000**, *42*, 47–238.

(16) Butler, M. J.; Bruheim, P.; Jovetic, S.; Marinelli, F.; Postma, P. W.; Bibb, M. J. Engineering of Primary Carbon Metabolism for Improved Antibiotic Production in *Streptomyces Lividans*. *Appl. Environ. Microbiol.* **2002**, *68* (10), 4731 LP – 4739.

(17) Ryu, Y.-G.; Butler, M. J.; Chater, K. F.; Lee, K. J. Engineering of Primary Carbohydrate Metabolism for Increased Production of Actinorhodin in *Streptomyces Coelicolor*. *Appl. Environ. Microbiol.* **2006**, *72* (11), 7132–7139.

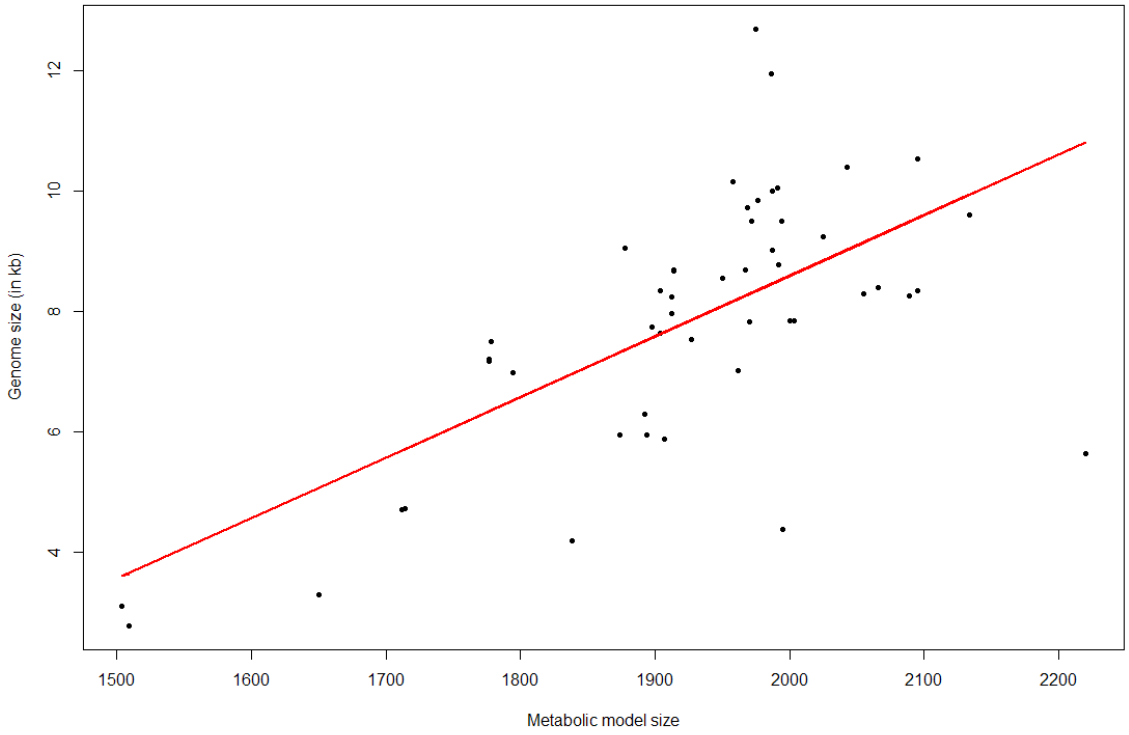
(18) Gomez-Escribano, J. P.; Bibb, M. J. Heterologous Expression of Natural Product Biosynthetic Gene Clusters in Streptomyces Coelicolor: From Genome Mining to Manipulation of Biosynthetic Pathways. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 425–431.
(19) Zakrzewski, P.; Medema, M. H.; Gevorgyan, A.; Kierzek, A. M.; Breitling, R.; Takano, E. MultiMetEval: Comparative and Multi-Objective Analysis of Genome-Scale Metabolic Models. *PLoS One* **2012**, *7* (12), e51511.
(20) Schellenberger, J.; Que, R.; Fleming, R. M. T.; Thiele, I.; Orth, J. D.; Feist, A. M.; Zielinski, D. C.; Bordbar, A.; Lewis, N. E.; Rahmanian, S.; et al. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6* (9), 1290–1307.
(21) Amara, A.; Takano, E.; Breitling, R. Development and Validation of an Updated Computational Model of *Streptomyces Coelicolor* Primary and Secondary Metabolism. *BMC Genomics* **2018**, *19* (1), 519.
(22) Borodina, I.; Krabben, P.; Nielsen, J. Genome-Scale Analysis of Streptomyces Coelicolor A3(2) Metabolism. *Genome Res.* **2005**, *15* (6), 820–829.
(23) Alam, M. T.; Merlo, M. E.; Hodgson, D. A.; Wellington, E. M. H.; Takano, E.; Breitling, R. Metabolic Modeling and Analysis of the Metabolic Switch in Streptomyces Coelicolor. *BMC Genomics* **2010**, *11* (1), 1.
(24) Kim, M.; Yi, J. S.; Kim, J.; Kim, J. N.; Kim, M. W.; Kim, B. G. Reconstruction of a High-Quality Metabolic Model Enables the Identification of Gene Overexpression Targets for Enhanced Antibiotic Production in Streptomyces Coelicolor A3(2). *Biotechnol. J.* **2014**, *9* (9), 1185–1194.
(25) Dang, L.; Liu, J.; Wang, C.; Liu, H.; Wen, J. Enhancement of Rapamycin Production by Metabolic Engineering in Streptomyces Hygroscopicus Based on Genome-Scale Metabolic Model. *J. Ind. Microbiol. Biotechnol.* **2017**, *44* (2), 259–270.
(26) Wang, J.; Wang, C.; Song, K.; Wen, J. Metabolic Network Model Guided Engineering Ethylmalonyl-CoA Pathway to Improve Ascomycin Production in Streptomyces Hygroscopicus Var. Ascomyceticus. *Microb. Cell Fact.* **2017**, *16* (1), 169.
(27) Toro, L.; Pinilla, L.; Avignone‐rossa, C.; Ríos‐estepa, R.; Avignone-Rossa, C.; Ríos-Estepa, R. An Enhanced Genome-Scale Metabolic Reconstruction of Streptomyces Clavuligerus Identifies Novel Strain Improvement Strategies. *Bioprocess Biosyst. Eng.* No. 5.
(28) Medema, M. H.; Alam, M. T.; Heijne, W. H. M.; van den Berg, M. A.; Müller, U.; Trefzer, A.; Bovenberg, R. A. L.; Breitling, R.; Takano, E. Genome-Wide Gene Expression Changes in an Industrial Clavulanic Acid Overproduction Strain of Streptomyces Clavuligerus. *Microb. Biotechnol.* **2011**, *4* (2), 300–305.
(29) Fondi, M.; Pinatel, E.; Talà, A.; Damiano, F.; Consolandi, C.; Mattorre, B.; Fico, D.; Testini, M.; De Benedetto, G. E.; Siculella, L.; et al. Time-Resolved Transcriptomics and Constraint-Based Modeling Identify System-Level Metabolic Features and Overexpression Targets to Increase Spiramycin Production in Streptomyces Ambofaciens. *Front. Microbiol.* **2017**, *8*, 835.
(30) Razmilic, V.; Castro, J. F.; Andrews, B.; Asenjo, J. A. Analysis of Metabolic Networks of *Streptomyces Leeuwenhoekii* C34 by Means of a Genome Scale Model: Prediction of Modifications That Enhance the Production of Specialized

Metabolites. *Biotechnol. Bioeng.* **2018**, *115* (7), 1815–1828.
(31) Kavvas, E. S.; Seif, Y.; Yurkovich, J. T.; Norsigian, C.; Poudel, S.; Greenwald, W. W.; Ghatak, S.; Palsson, B. O.; Monk, J. M. Updated and Standardized Genome-Scale Reconstruction of Mycobacterium Tuberculosis H37Rv, IEK1011, Simulates Flux States Indicative of Physiological Conditions. *BMC Syst. Biol.* **2018**, *12* (1), 25.
(32) Jamshidi, N.; Palsson, B. Ø. Investigating the Metabolic Capabilities of Mycobacterium Tuberculosis H37Rv Using the in Silico Strain INJ 661 and Proposing Alternative Drug Targets. *BMC Syst. Biol.* **2007**, *1* (1), 26.
(33) Garay, C. D.; Dreyfuss, J. M.; Galagan, J. E. Metabolic Modeling Predicts Metabolite Changes in Mycobacterium Tuberculosis. *BMC Syst. Biol.* **2015**, *9* (1), 57.
(34) Kjeldsen, K. R.; Nielsen, J. In Silico Genome-Scale Reconstruction and Validation of the *Corynebacterium Glutamicum* Metabolic Network. *Biotechnol. Bioeng.* **2009**, *102* (2), 583–597.
(35) Zhang, Y.; Cai, J.; Shang, X.; Wang, B.; Liu, S.; Chai, X.; Tan, T.; Zhang, Y.; Wen, T. A New Genome-Scale Metabolic Model of Corynebacterium Glutamicum and Its Application. *Biotechnol. Biofuels* **2017**, *10* (1), 169.
(36) Henry, C. S.; DeJongh, M.; Best, A. A.; Frybarger, P. M.; Linsay, B.; Stevens, R. L. High-Throughput Generation, Optimization and Analysis of Genome-Scale Metabolic Models. *Nat. Biotechnol.* **2010**, *28* (9), 977–982.
(37) Devoid, S.; Overbeek, R.; DeJongh, M.; Vonstein, V.; Best, A. A.; Henry, C. Automated Genome Annotation and Metabolic Model Reconstruction in the SEED and Model SEED; Humana Press, Totowa, NJ, 2013; pp 17–45.
(38) Agren, R.; Liu, L.; Shoaie, S.; Vongsangnak, W.; Nookaew, I.; Nielsen, J. The RAVEN Toolbox and Its Use for Generating a Genome-Scale Metabolic Model for Penicillium Chrysogenum. *PLoS Comput. Biol.* **2013**, *9* (3), e1002980.
(39) Faria, J. P.; Rocha, M.; Rocha, I.; Henry, C. S. Methods for Automated Genome-Scale Metabolic Model Reconstruction. *Biochem. Soc. Trans.* **2018**, *46* (4), 931–936.
(40) Karlsen, E.; Schulz, C.; Almaas, E. Automated Generation of Genome-Scale Metabolic Draft Reconstructions Based on KEGG. *BMC Bioinformatics* **2018**, *19* (1), 467.
(41) Pitkänen, E.; Jouhten, P.; Hou, J.; Syed, M. F.; Blomberg, P.; Kludas, J.; Oja, M.; Holm, L.; Penttilä, M.; Rousu, J.; et al. Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. *PLoS Comput. Biol.* **2014**, *10* (2), e1003465.
(42) Lee, D.-S.; Burd, H.; Liu, J.; Almaas, E.; Wiest, O.; Barabási, A.-L.; Oltvai, Z. N.; Kapatral, V. Comparative Genome-Scale Metabolic Reconstruction and Flux Balance Analysis of Multiple Staphylococcus Aureus Genomes Identify Novel Antimicrobial Drug Targets. *J. Bacteriol.* **2009**, *191* (12), 4015–4024.
(43) Ebrahim, A.; Lerman, J. A.; Palsson, B. Ø.; Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7* (1), 74.
(44) Wattam, A. R.; Abraham, D.; Dalay, O.; Disz, T. L.; Driscoll, T.; Gabbard, J. L.; Gillespie, J. J.; Gough, R.; Hix, D.; Kenyon, R.; et al. PATRIC, the Bacterial Bioinformatics Database and Analysis Resource. *Nucleic Acids Res.* **2014**, *42* (D1), D581–D591.

(45) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.
(46) Price, M. N.; Dehal, P. S.; Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **2010**, *5* (3), e9490.
(47) Letunic, I.; Bork, P. Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees. *Nucleic Acids Res.* **2016**, *44* (W1), W242–W245.
(48) Rost, B. Twilight Zone of Protein Sequence Alignments. *Protein Eng.* **1999**, *12* (2), 85–94.
(49) Pearson, W. R. An Introduction to Sequence Similarity ("Homology") Searching. *Curr. Protoc. Bioinforma.* **2013**, *Chapter 3*, Unit3.1.
(50) The UniProt Consortium. UniProt: A Hub for Protein Information. *Nucleic Acids Res.* **2015**, *43* (D1), D204–D212.
(51) King, Z. A.; Lu, J.; Drager, A.; Miller, P.; Federowicz, S.; Lerman, J. A.; Ebrahim, A.; Palsson, B. Ø.; Lewis, N. E.; J., H. BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models. *Nucleic Acids Res.* **2016**, *44* (D1), D515–D522.
(52) Caspi, R.; Billington, R.; Ferrer, L.; Foerster, H.; Fulcher, C. A.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L. A.; et al. The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2016**, *44* (D1), D471–D480.
(53) Aurich, M. K.; Fleming, R. M. T.; Thiele, I. MetaboTools: A Comprehensive Toolbox for Analysis of Genome-Scale Metabolic Models. *Front. Physiol.* **2016**, *7*, 327.
(54) Chen, H.; Boutros, P. C. VennDiagram: A Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R. *BMC Bioinformatics* **2011**, *12* (1), 35.
(55) Sievert, C.; Parmer, C.; Hocking, T.; Chamberlain, S.; Ram, K. Plotly: Create Interactive Web Graphics via'plotly. Js'. 2016. R Package Version 3.6. 0. **2018**.
(56) Galili, T. Dendextend: An R Package for Visualizing, Adjusting and Comparing Trees of Hierarchical Clustering. *Bioinformatics* **2015**, *31* (22), 3718–3720.
(57) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504.
(58) Le Novere, N.; Finney, A.; Hucka, M.; Bhalla, U. S.; Campagne, F.; Collado-Vides, J.; Crampin, E. J.; Halstead, M.; Klipp, E.; Mendes, P.; et al. Minimum Information Requested in the Annotation of Biochemical Models (MIRIAM). *Nat. Biotechnol.* **2005**, *23* (12), 1509–1515.
(59) Green, M. L.; Karp, P. D. Genome Annotation Errors in Pathway Databases Due to Semantic Ambiguity in Partial EC Numbers. *Nucleic Acids Res.* **2005**, *33* (13), 4035–4039.
(60) Thiele, I.; Palsson, B. Ø. A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction. *Nat. Protoc.* **2010**, *5* (1), 93–121.
(61) Satish Kumar, V.; Dasika, M. S.; Maranas, C. D. Optimization Based Automated Curation of Metabolic Reconstructions. *BMC Bioinformatics* **2007**, *8* (1), 212.

(62) Latendresse, M. Efficiently Gap-Filling Reaction Networks. *BMC Bioinformatics* **2014**, *15* (1), 225.
(63) Dikicioglu, D.; Kırdar, B.; Oliver, S. G. Biomass Composition: The "Elephant in the Room" of Metabolic Modelling. *Metabolomics* **2015**, *11* (6), 1690–1701.
(64) Xavier, J. C.; Patil, K. R.; Rocha, I. Metabolic Models and Gene Essentiality Data Reveal Essential and Conserved Metabolism in Prokaryotes. *PLOS Comput. Biol.* **2018**, *14* (11), e1006556.
(65) Mohite, O. S.; Weber, T.; Kim, H. U.; Lee, S. Y. Genome-Scale Metabolic Reconstruction of Actinomycetes for Antibiotics Production. *Biotechnol. J.* **2019**, *14* (1), 1800377.
(66) Zhu, H.; Sandiford, S. K.; van Wezel, G. P. Triggers and Cues That Activate Antibiotic Production by Actinomycetes. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 371–386.
(67) Nazari, B.; Kobayashi, M.; Saito, A.; Hassaninasab, A.; Miyashita, K.; Fujii, T. Chitin-Induced Gene Expression in Secondary Metabolic Pathways of Streptomyces Coelicolor A3(2) Grown in Soil. *Appl. Environ. Microbiol.* **2013**, *79* (2), 707–713.
(68) Kim, E. S.; Hong, H. J.; Choi, C. Y.; Cohen, S. N. Modulation of Actinorhodin Biosynthesis in *Streptomyces Lividans* by Glucose Repression of AfsR2 Gene Transcription. *J. Bacteriol.* **2001**, *183* (7), 2198–2203.
(69) Gao, C.; Hindra; Mulder, D.; Yin, C.; Elliot, M. A. Crp Is a Global Regulator of Antibiotic Production in *Streptomyces*. *MBio* **2012**, *3* (6), e00407-12.
(70) Barsky, D. L.; Hoffee, P. A. Purification and Characterization of Phosphopentomutase from Rat Liver. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1983**, *743* (1), 162–171.
(71) Sgarrella, F.; Poddie, F. P. A.; Meloni, M. A.; Sciola, L.; Pippia, P.; Tozzi, M. G. Channelling of Deoxyribose Moiety of Exogenous DNA into Carbohydrate Metabolism: Role of Deoxyriboaldolase. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* **1997**, *117* (2), 253–257.
(72) Karlström, O. Mutants of Escherichia Coli Defective in Ribonucleoside and Deoxyribonucleoside Catabolism. *J. Bacteriol.* **1968**, *95* (3), 1069–1077.
(73) Munch-Petersen, A. Deoxyribonucleoside Catabolism and Thymine Incorporation in Mutants of Escherichia Coli Lacking Deoxyriboaldolase. *Eur. J. Biochem.* **1970**, *15* (1), 191–202.

4.8 Supplementary Data



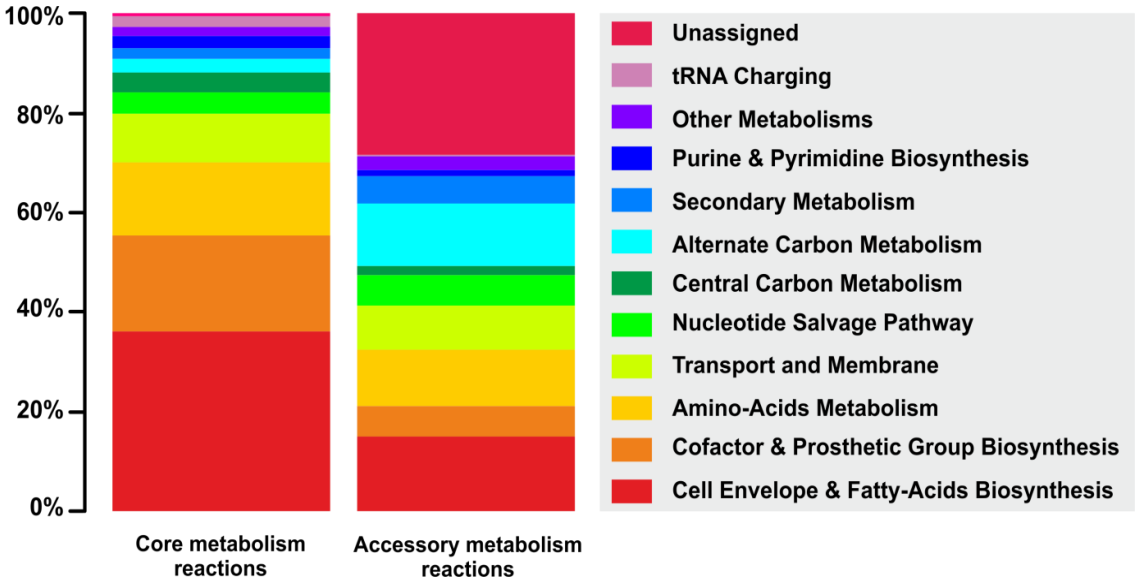
Supplementary Figure 4.1: Metabolic model size to the genome size

The metabolic model size is compared to the genome size of the strains. With an overall trend showing that the model size increases with the genome size. The trend (red line) is based on a linear regression fitted to the data.

Metabolites	Reaction_IDs
Asparagine	EX_asn_DASH_L_LPAREN_e_RPAREN_
Arginine	EX_arg_DASH_L_LPAREN_e_RPAREN_
Biotin	EX_btn_LPAREN_e_RPAREN_
Calcium	EX_ca2_LPAREN_e_RPAREN_
Cellobiose	EX_celb_LPAREN_e_RPAREN_
Chloride	EX_cl_LPAREN_e_RPAREN_
Citrate	EX_cit_LPAREN_e_RPAREN_
CO2	EX_co2_LPAREN_e_RPAREN_
Cobalt	EX_cobalt2_LPAREN_e_RPAREN_
Copper2	EX_cu2_LPAREN_e_RPAREN_
D-Fructose	EX_fru_LPAREN_e_RPAREN_
D-Galactose	EX_gal_LPAREN_e_RPAREN_
D-Glucose	EX_glu_DASH_L_LPAREN_e_RPAREN_
D-Mannose	EX_man_LPAREN_e_RPAREN_
D-Ribose	EX_rib_DASH_D_LPAREN_e_RPAREN_
Fe(III)dicitrate	EX_fe3_LPAREN_e_RPAREN_
Fe2+	EX_fe2_LPAREN_e_RPAREN_
Glycerol	EX_glyc_LPAREN_e_RPAREN_
H+	EX_h_LPAREN_e_RPAREN_
H2O	EX_h2o_LPAREN_e_RPAREN_
Lactose	EX_lac_DASH_D_LPAREN_e_RPAREN_
L-Cysteine	EX_cys_DASH_L_LPAREN_e_RPAREN_
L-Glutamate	EX_glu_DASH_L_LPAREN_e_RPAREN_
L-Glutamine	EX_gln_DASH_L_LPAREN_e_RPAREN_
L-Methionine	EX_met_DASH_L_LPAREN_e_RPAREN_
Magnesium	EX_mg2_LPAREN_e_RPAREN_
Maltose	EX_malt_LPAREN_e_RPAREN_
Manganese	EX_mn2_LPAREN_e_RPAREN_
Molybdate	EX_mobd_LPAREN_e_RPAREN_
NH4	EX_nh4_LPAREN_e_RPAREN_
Orthophosphate	EX_pi_LPAREN_e_RPAREN_
Oxygen	EX_o2_LPAREN_e_RPAREN_
Pantothenate	EX_pnto_DASH_R_LPAREN_e_RPAREN_
Propionic acid	EX_ppa_LPAREN_e_RPAREN_
Potassium	EX_k_LPAREN_e_RPAREN_
Sulfate	EX_so4_LPAREN_e_RPAREN_
Thiamin	EX_thm_LPAREN_e_RPAREN_
Urea	EX_urea_LPAREN_e_RPAREN_
Zinc	EX_zn2_LPAREN_e_RPAREN_

Supplementary Table 4.1: Metabolites constituting the universal minimal media

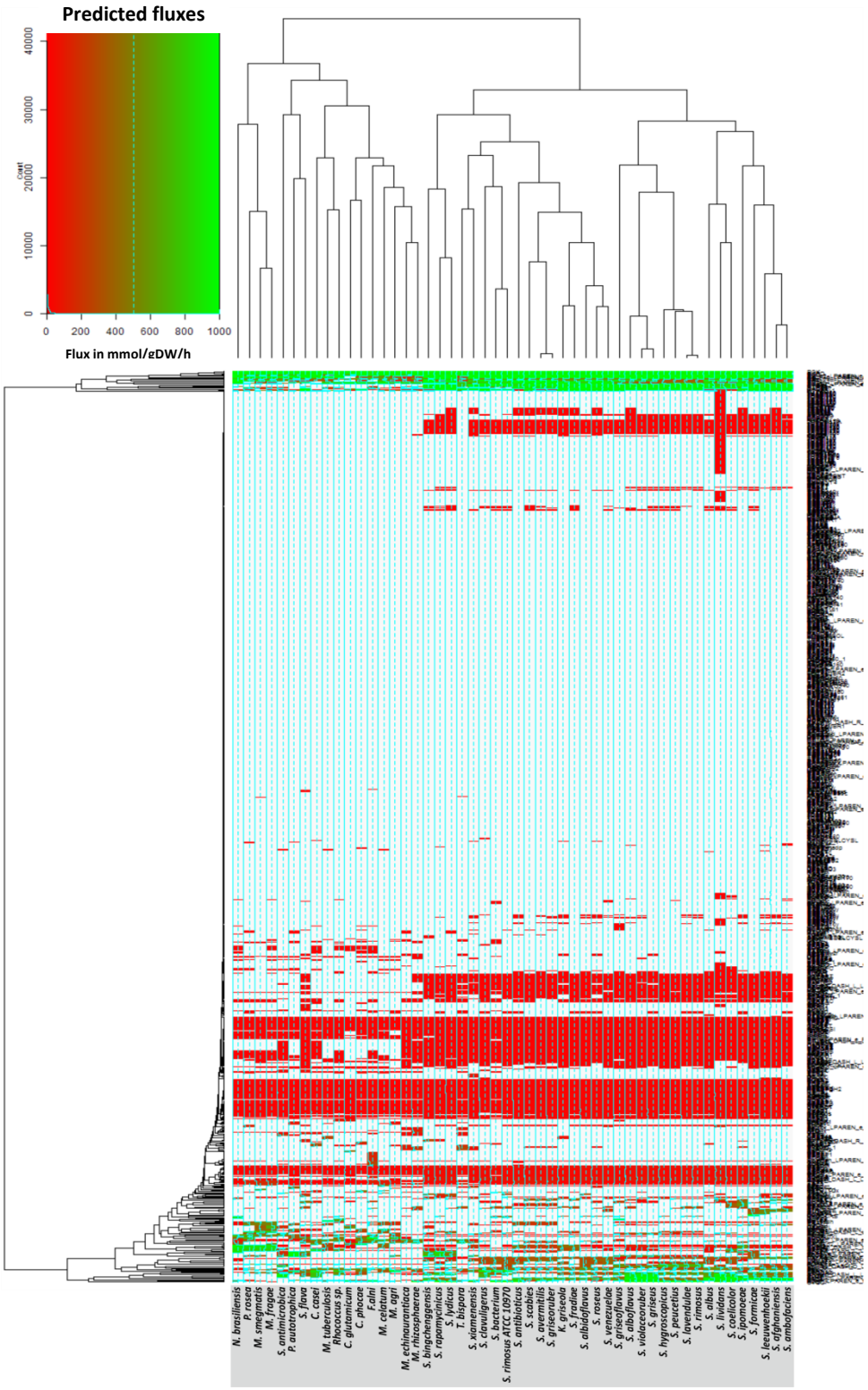
This is the composition of the universal minimal media used to test the strains. The nutrients in this list are either essential for all the strains growth or essential for at least one of the strains. These are based on experimental data available in literature



Supplementary Figure 4.2: Metabolic pathways associated to core and accessory metabolism

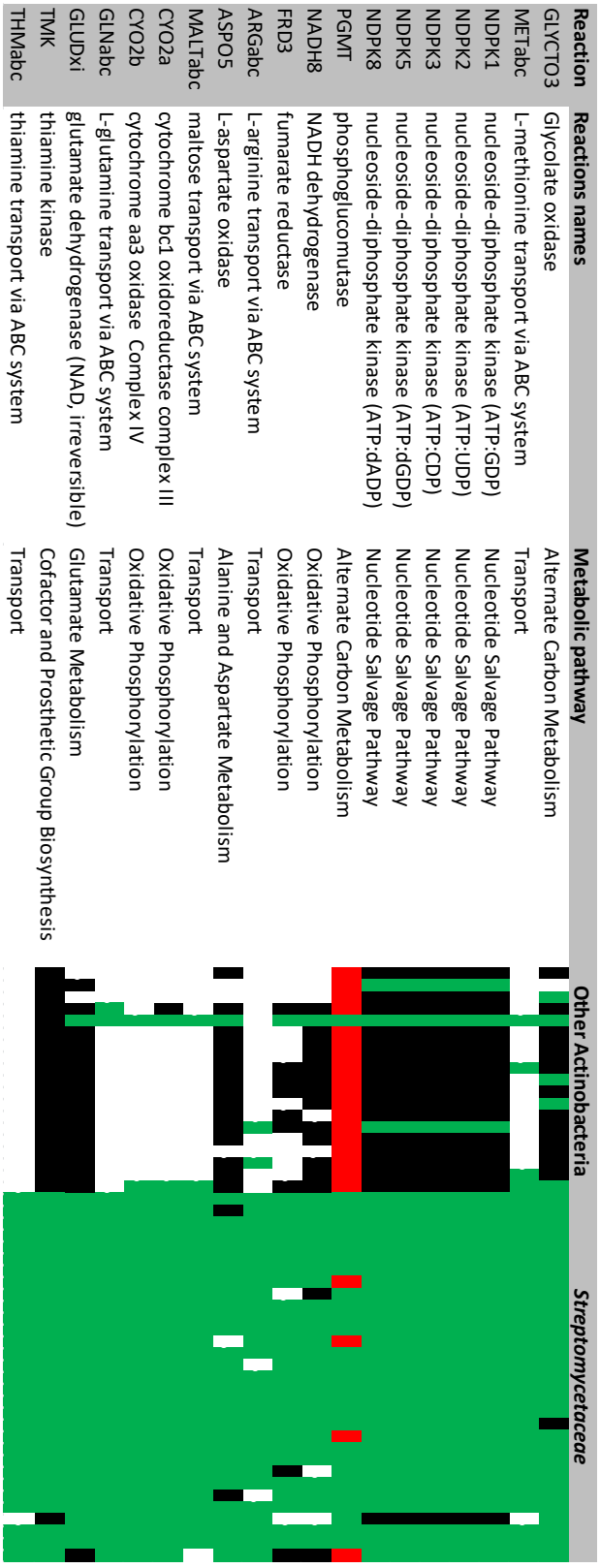
The share of reactions (in percent) per subsystems categories (i.e., metabolic pathways) are represented by different colours. The metabolic pathways come from the reference model (*i*AA1259) and the BiGG database models associated with these reactions. The shares of reactions are determined for the universally conserved reactions metabolic models (i.e., core metabolism) and for the non-core reactions (i.e., accessory metabolism).

This figure includes unassigned pathways, which are reactions added from the databases but that do not have any subsystem annotations in the databases.



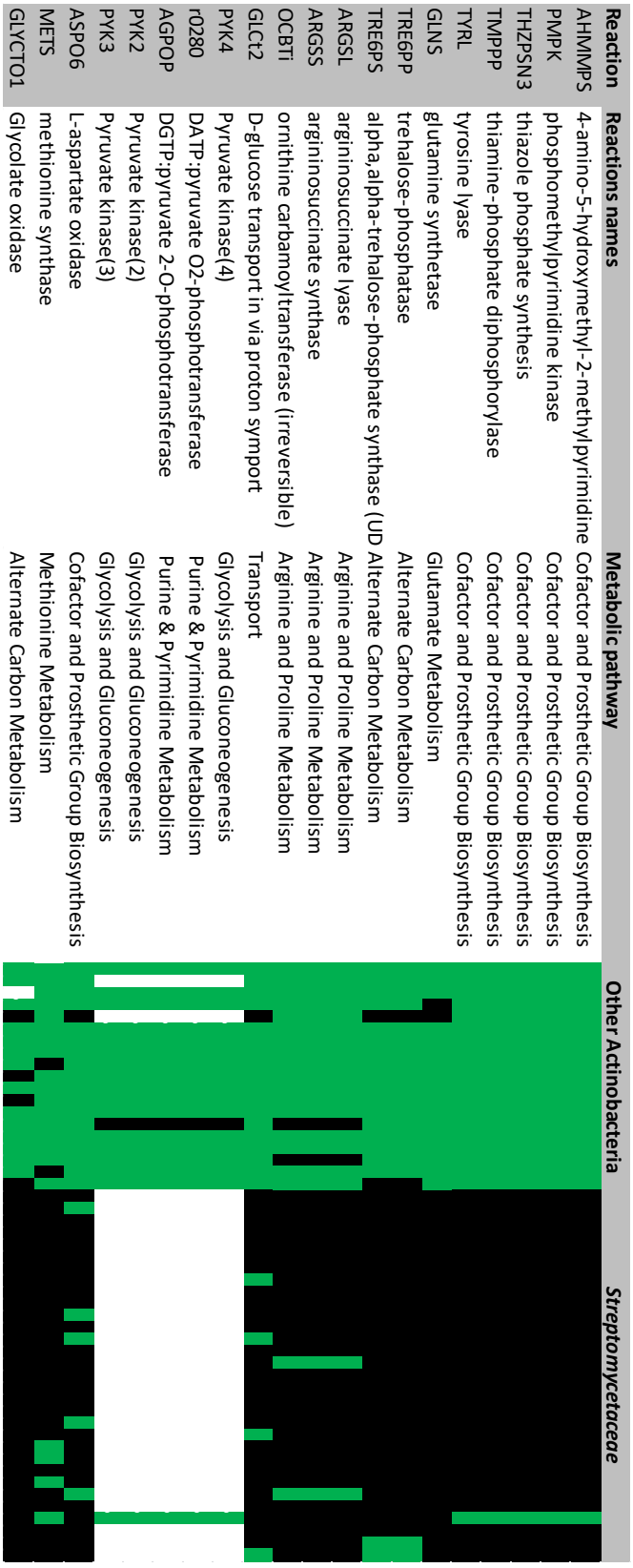
Supplementary Figure 4.3: Heatmap of the metabolic fluxes predicted in the different strains metabolic models

The metabolic fluxes predicted in by pFBA in the different metabolic models using the universal minimal media are visualised as a heatmap. The strains are in columns and the reactions are in rows, with the colours gradient from red to green corresponding to the predicted fluxes from low to high flux. The group of strains and reactions were clustered using hierarchical clustering, with the *Streptomycetaceae* clustering together and other Actinobacteria together (unless *Thermobispora bispora*).



Supplementary Figure 4.4: Reactions mostly active in *Streptomycetaceae* and not in other Actinobacteria

In the heatmap white cells are absent reactions, black are null fluxes reactions, in green are reactions carrying fluxes, red is opposite flux direction (negative values). Here only the reactions mostly used (>70%) by *Streptomycetaceae* group compared to other Actinobacteria are showed here.



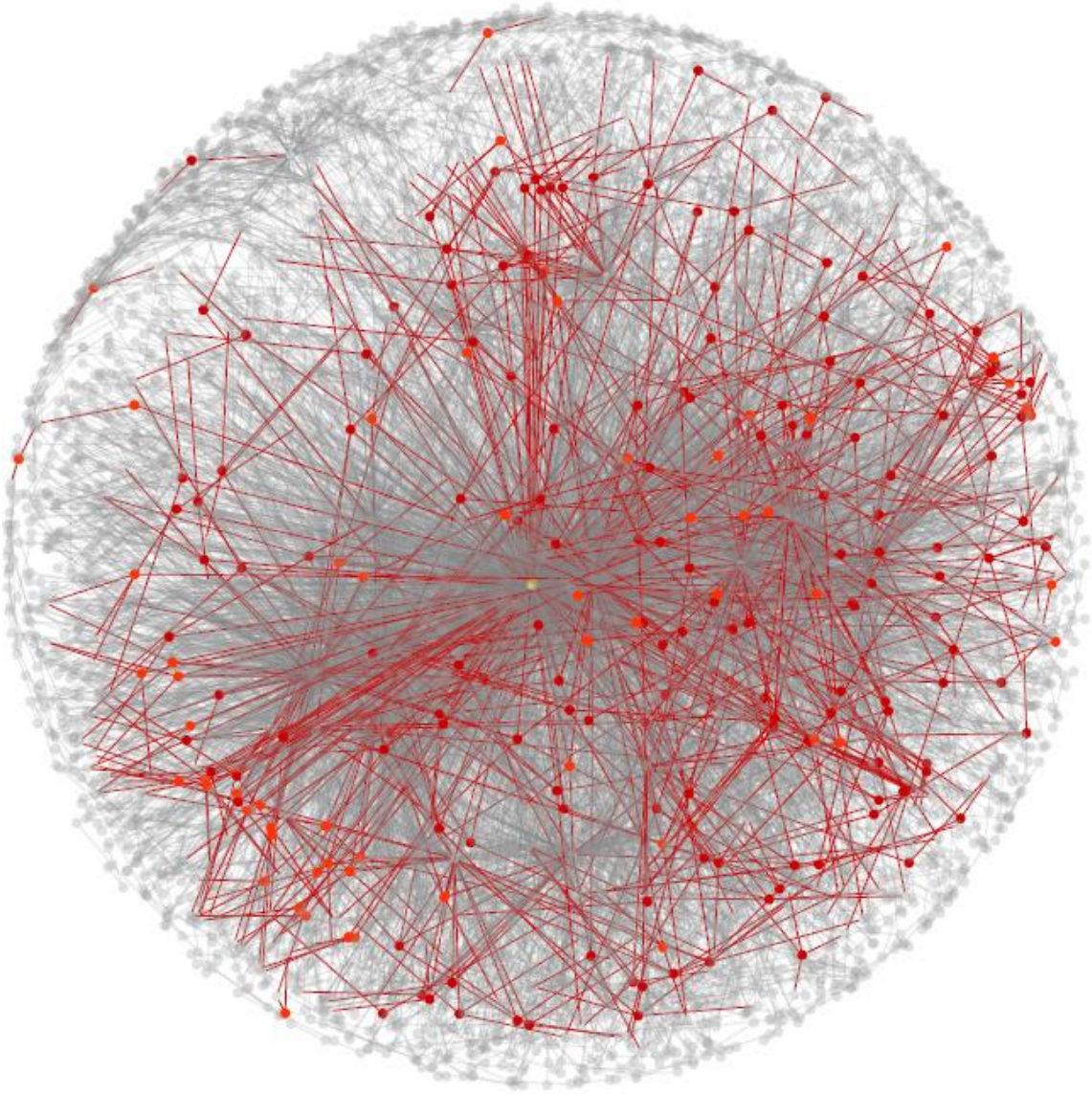
Supplementary Figure 4.5: Reactions mostly active in other Actinobacteria and not in *Streptomycetaceae*

In the heatmap white cells are absent reactions, black are null fluxes reactions, in green are the reactions carrying fluxes. Here only the reactions mostly used (>70%) by other Actinobacteria compared to the *Streptomycetaceae* group are showed here.

Subsystem	Number of reactions
Alternate Carbon Metabolism	1
Citric Acid Cycle	1
Oxidative Phosphorylation	1
Inositol Phosphate Metabolism	2
Unassigned	2
Pentose Phosphate Pathway	3
Transport and Membrane	5
Nucleotide Salvage Pathway	7
Quinol and Quinone Biosynthesis	7
Glycolysis and Gluconeogenesis	8
Inorganic Ion Transport and Metabolism	8
Purine and Pyrimidine Biosynthesis	17
Amino-Acids	70
Cofactor and Prosthetic Group Biosynthesis	87
Cell Envelope and Lipids Biosynthesis	236

Supplementary Table 4.2: Reactions active in all the Actinobacteria strains

The pathways with the number of reactions universally active in the Actinobacteria strains are shown in this table. The different metabolic pathways are based on the BiGG model database subsystems and the iAA1259 subsystems. The main pathways universally used are involved in producing the precursors for the biomass, such as lipids, cofactors, and amino-acids.



Supplementary Figure 4.6: Core metabolic model fluxes in the pan-metabolic network using edge-weighted layout

The pan-metabolic network is represented here, with the reactions as the edges and metabolites as nodes. The core metabolic fluxes (reactions with carrying a flux in all the metabolic models n red here). The reactions carrying fluxes in at least 45 models (85% of the strains) are in orange. The reactions and metabolites involved in the accessory metabolism under these minimal conditions are in grey in the network.

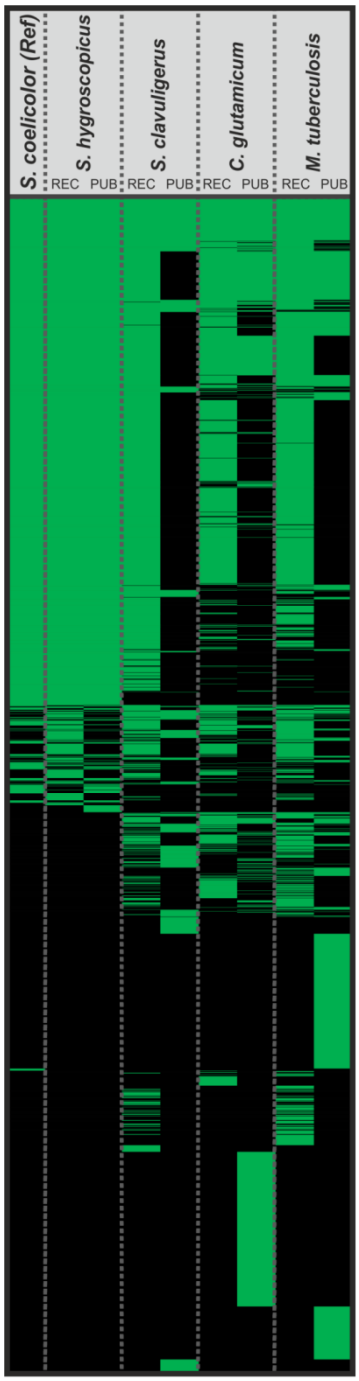
4.8.1 Comparison of reconstructed metabolic models with published metabolic models

Many models had reactions and metabolites IDs based on in-house annotations which meant it would require the creation of a conversion table for each of the published models (this was only done for the *S. clavuligerus* model). Therefore, the comparison was mostly limited to models compliant with the BiGG standards, this included metabolic models for *S. hygroscopicus* [26], *C. glutamicum* [35], and *M. tuberculosis* [31], as well as a converted model for *S. clavuligerus* [27] (*Supplementary Figure 4.7*).

The comparison showed that the automatically reconstructed model for *S. hygroscopicus* had an 83% similarity with the manually curated and published model. The similarity was of 24% between the reconstructed and the published *C. glutamicum* models, 22% for *M. tuberculosis*, and 17% (45% adjusted similarity) for the *S. clavuligerus* models. As the comparison method is based on the BiGG IDs of the reactions, the comparison was also done ignoring all transport reactions, secondary metabolite pathways, and spontaneous reactions (with no gene associated), as these are either ambiguous, annotated differently, or absent depending on the organism model. The results were 94% similarity between the manually curated and published model of *S. hygroscopicus*, 37% for *C. glutamicum* models, 39% for *M. tuberculosis*, and 21% (57% adjusted similarity) for the *S. clavuligerus* models. The significant increase in the similarity between the published and reconstructed models shows the impact of ambiguous annotations in the models comparison. A major problem in the comparison is the ambiguity in the reaction and metabolite annotations used in the different GSMMs and in the databases; which is a well-known issue in the community [59]. This is one of the reasons why the Actinobacteria models in this study were all reconstructed using the same standards rather than re-using published metabolic models; this reconstruction ensured that the models were comparable. The published *S. clavuligerus* model is the most different from the automatically reconstructed model, which would be surprising regarding the closeness of this strain to the reference *S. coelicolor* strain. But, this is likely due to the method used that compares the models based on BiGG standards the reactions IDs (see *Methods*); and this model did not use BiGG IDs, and Enzyme Commission (EC) numbers had to be used instead,

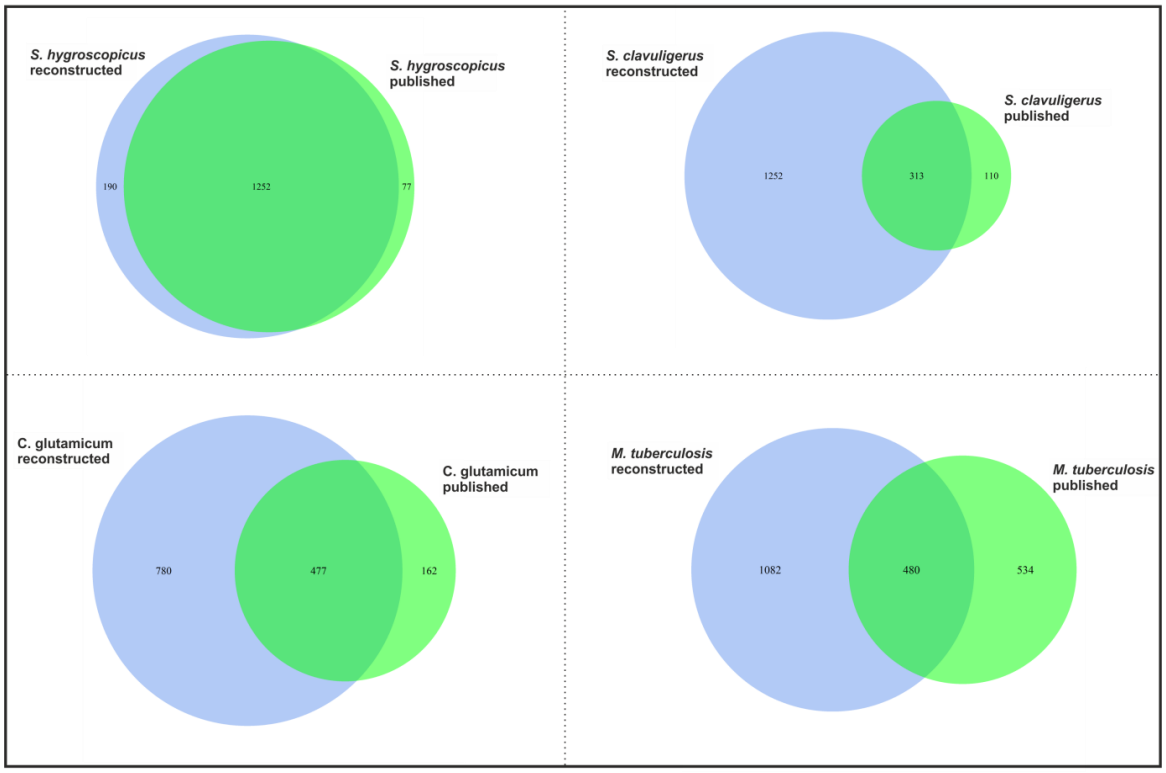
but EC numbers are known for their ambiguity [60]. Only 37% of the reactions present in the published model were converted in BiGG IDs, so the level of similarity were adjusted to the number of reactions actually extracted, making *S. clavuligerus* the second closest strain to the *S. coelicolor* model.

The models were further compared by identifying the overlapping of the metabolic models, the *S. hygroscopicus* model covered 94% of the published metabolic model; this reach 74% for *S. clavuligerus*, 75% for *C. glutamicum*, and 47% for *M. tuberculosis* (*Supplementary Figure 4.8*). However, the *C. glutamicum* reconstructed model is almost two times bigger than the published model. The pipeline overestimates the number of reactions present in the organism, because many reactions are transferred from the reference model because these are essential; this is further discussed in the Chapter. Another problem in the comparison is the ambiguity in annotations; some BiGG IDs are different but describe the same reaction. For example, in *M. tuberculosis*, when 10 reactions not matching in both models are selected, 6 reactions were found with the exact same reaction in the published model (*Supplementary Figure 4.9*). Otherwise, the models for *M. tuberculosis* and *C. glutamicum* correspond to phylogenetically more distant strains from the reference strain, so the pipeline may not reconstruct with sufficient precision these models. This indicates that the quality is lower for more distant models such as *M. tuberculosis* and *C. glutamicum*. While for closer strains like *S. hygroscopicus* seems to reach a high level of similarity with high-quality manually curated models [26].



Supplementary Figure 4.7: Overall comparison of published metabolic models with the reconstructed metabolic models.

The published metabolic models (PUB) were compared to the reconstructed metabolic models (REC) for *Streptomyces hygroscopicus*, *Streptomyces clavuliugerus*, *Corynebacterium glutamicum*, and *Mycobacterium tuberculosis*. Reactions present in the different metabolic models including the reference model of *Streptomyces coelicolor* are represented in green, while the absence of reactions is represented in black. The closest model to the reference strain is the *S. coelicolor*, while *S. clavuligerus* is more distant due to the comparison method based on EC numbers (see *Methods*). The most distant strains *C. glutamicum* and M. *tuberculosis* are the most different from the automatically reconstructed models.



Supplementary Figure 4.8: Venn diagram of the overlapping reactions in the reconstructed and published models of *S. hygroscopicus*, *S. clavuligerus*, *C. glutamicum*, and *M. tuberculosis*

Coverage of the published models reactions set (in green) by the reconstructed models reaction set (in blue).

The *S. hygroscopicus* reconstruction covers 94% of the published model.

The *S. clavuligerus* reconstruction covers 74% of the published model; but it has many more reaction because the published model reactions were queried using EC numbers, which only covered 37% of the total reactions of the published metabolic model.

The *C. glutamicum* reconstructed model covers 75% of the published model, however the reconstruction have many more reactions than the published metabolic model. This is likely due to two main factors, the number of essential reactions reintroduced in the model to ensure growth (as described in the Chapter), and due to the ambiguity in reactions BiGG IDs that have different names for the same reactions (*Table 4.1*).

The *M. tuberculosis* reconstructed model covers 47% of the published model. The same factors described above for *C. glutamicum* seems responsible for the low overlap between these two models.

M. tuberculosis reconstruction	*M. tuberculosis* published	Name in reconstruction	Name in published model	Notes
FACOAL120	FAS120	Fatty-acid--CoA ligase	Fatty acid synthase n C120	Different annotation of the same metabolic function
MPTS	MOAE1	Molybdopterin synthase	MNXR85919	Different annotation of the same metabolic function
AHSERL2r	ACHMSSH	O-acetylhomoserine (thiol)-lyase	ACHMSSH	Different annotation of the same metabolic function
MNLD	X	Mannitol 2-dehydrogenase	X	Absent
PSD150	X	Phosphatidylserine decarboxylase (n-C15:0)	X	Absent
ACONTa/ACONTb	ACONT	aconitase (half-reaction A, Citrate hydro-lyase)/ aconitase (half-reaction B, Isocitrate hydro-lyase)	Aconitase	Reaction more detailed in reconstruction
ARGD	ARGDr	Arginine Deiminase	arginine deiminase	Different annotation of the same metabolic function
X	MPML	X	MPML	Absent
GLYCK2	GLYCK	Glycerate Kinase (2pg product)	glycerate kinase (3pg product)	Close reaction exists but slightly different product
NTD11e	NTD11	5'-nucleotidase (IMP)	5'-nucleotidase (IMP)	Different annotation of the same metabolic function

Supplementary Table 4.3: Individual analysis of 10 random reactions only found in one of the reconstructed or published models of *M. tuberculosis*

The 10 reactions picked randomly had 5 reactions that were only present in the model reconstruction (5 upper part yellow reactions) and 5 reactions only present in the published model (5 lower part yellow reactions).

6/10 of the reactions were found to have the same reactions (exact same metabolites and stoichiometry). 3 reactions were completely absent, and 1 reaction had one product different 2phosphoglycerate (2pg) instead of 3 phosphoglycerate (3pg).

This shows that it is very likely that many reactions considered as different in the comparison are actually the exact same reaction but with a different ID.

Chapter V

5. An automated *R* tool to integrate exometabolome fluxes in constraints-based metabolic modelling

5.1 Preface

The work carried here was an *in silico* study, so most of the experimental data used and discussed were from published or publicly available data, apart from the first dataset I acquired for *Streptomyces coelicolor*. This is a manuscript in preparation for submission.

5.2 Abstract

The microbial production of industrially and medically relevant compounds has expanded during the last decade due to the rapid development of synthetic biology and metabolic engineering. A major challenge when engineering a production strain is to rapidly test and debug the strain. For this purpose, many researchers use constraint-based metabolic modelling to analyse the metabolic impact on the engineered strains and identify new targets to further increase production. This modelling method relies on the introduction of flux constraints, particularly exchange fluxes with the media; consequently, the estimation of these fluxes is crucial to generate accurate predictions. In order to automate this step and directly input test data results into the model construction, an *R* tool was developed to calculate the exchange rates with their confidence interval based on metabolite concentration and biomass measurements in

the culture (chemostat or flask). The tool was tested with multiple experimental data to calculate the exchange rates and use these to constrain a genome-scale metabolic model of the antibiotic producer model strain *Streptomyces coelicolor* A3(2). The tool's outputs were also used to apply an ensemble modelling approach, in which the model was constrained using exchange fluxes sampled within the confidence interval and hundreds of metabolic states were predicted using the resulting ensemble of metabolic models. This automated tool can now be used to bridge the test and design phase for strain engineering using synthetic biology.

5.3 Introduction

Bioproduction became a key manufacturing technology in the last few decades, from the production of biopharmaceuticals to the sustainable production of commodity chemicals and biofuels [1,2]. This was accelerated by synthetic biology and metabolic engineering that enabled fast engineering and optimisation of biological systems [3,4]. The application of engineering principles in synthetic biology created opportunities to produce more complex and diverse compounds in microorganism [5]. One main engineering principle applied in synthetic biology is the design, build, test, and learn (DBTL) cycle. During the last decade, this cycle, especially in the build, the pace has substantially accelerated, as a result of the decreasing cost of gene synthesis and sequencing, as well as easier and lower-cost technologies for gene-editing (e.g., CRISPR, TALENs) and genetic construction (e.g., Gibson assembly). However, a bottleneck in the test-to-design transition through the learn phase is stalling this cycle: the large amount of data generated are complex to interpret and integrate into the design phase [6,7]. So, when debugging an engineered strain, a major issue encountered

is to rapidly understand the effects of the metabolic modifications on the strains based on the test data (e.g., metabolomics) and feed it back in the design phase (e.g., metabolic modelling). In the design phase, the application of constraint-based metabolic modelling has opened opportunities to integrate omics data from engineered strains and analyse the effect of compound production on the biological system [8]. Constraint-based metabolic modelling with a genome-scale metabolic model (GSMM) requires condition-specific constraints, such as uptake of nutrients and export of metabolites in the culture media [9]. Integration of exometabolomic flux data is a powerful method in constraint-based metabolic modelling to predict intracellular metabolic fluxes [10,11]. These constraints are crucial to accurately predict the strain metabolic state. However, the introduction of these constraints requires calculating and inputting the exchange rates at multiple time points to simulate the metabolism across a growth curve. Furthermore, constraint-based methods like flux-balance analysis (FBA) predict one optimum state based on a unique set of constraints, so this does not take in to account the inherent uncertainty around the possible exchange rates and possible noise in the measurements.

To automate this step and take into account uncertainty in the process, an automated tool was developed to use quantitative exometabolomic data as input (*Figure 5.1*); the tool estimates the exchange fluxes with a 95% confidence interval. The tool generates an output with the exchange rates, which are then used to constrain the GSMM and predict the intracellular metabolic fluxes with methods like FBA [10–12]. The tool output is usable to create an ensemble of simulations within the 95% confidence of exchange flux constraints used. Hence applying a "respectful modelling" approach [13], rather than limiting the analysis to the single FBA optimum,

this method takes into account the uncertainty and shows the possible states of the metabolic systems. This tool is easily used with time series data from a chemostat or a replicate flasks experiment.

The first advantage of this tool is the ability to estimate fluxes at any time point on a growth curve. This is useful for datasets with a small number of time points or datasets with gaps (e.g. no metabolite or growth measurement at a given time). The tool also eases the calculation of uptake/secretion rates, these are quickly calculated based on the input. Furthermore, this also ensures reproducibility in exchange rate calculations, avoiding errors to be introduced at the calculation step. The main advantage of this tool is the analysis of uncertainty in the flux estimation by giving an output with the upper, lower, and median uptake/secretion rates within the 95% confidence interval (CI). However, the tool still presents some limitations; the input data is limited to a certain type of input well adapted to time series studies. It is also necessary to inform the tool prior to the analysis of the type of uptake or secretion trend (linear or non-linear). Finally, the tool would not work for uptakes and secretions without simple trends (e.g., increase then decrease).

As a first step, the tool structure and functions were described. The tool was then tested with experimental growth and antibiotics data I had collected, then the tool was tested with a published chemostat dataset [14]. The tool output of this chemostat dataset was used to constrain the metabolic model *i*AA1259 of *Streptomyces coelicolor* A3(2) [15]. The predictions were validated using the chemostat experimental growth data. Finally, the tool was used to apply an ensemble modelling approach by sampling exchange fluxes within the confidence interval estimated, and these exchange fluxes were constrained in the metabolic model. This enabled to

predict an ensemble of hundreds of plausible metabolic states, which were compared and validated by the experimental growth data.

5.4 Methods and Materials

5.4.1 Tool development and dependencies

The tool was developed using the open-source *R* language in the format of an *R* Markdown notebook. It is built in a modular way to ease adaptation of the scripts to the data used. Experimental data is used as an input table and automatically analysed by the tool generating outputs at different steps and a final script output for COBRA in Python and MATLAB. Two different versions of the tool are available, one for exported metabolites and one for metabolites that are taken up. The *R* notebooks cover each step described in *Figure 5.1*, the conversion of data, the interpolation of biomass, uncertainty estimations, calculation of the predicted concentrations, and the output. The *R* notebooks for the different tests and experiments described above are available in the *Electronic Supplementary 5.1*.

Input and conversions: The data input is in a .csv format containing the biomass quantity and the metabolites' concentrations across time. The tool will convert these depending on the user input type indicated in the first notebook *R* chunk, indicating if the metabolite concentration is in OD, mg/L, or mmol/L, and if the biomass input is in OD (which necessitates further input of a conversion curve in .csv) or in gram of dry weight (gDW). The conversion curve used was provided by Prof. Eriko Takano [16]. The user must indicate if the data is in technical, or biological replicates, or just averages. As well as input the conversion factors based on the Beer–Lambert law if the metabolite input is in OD, indicate the culture volume, the time points to analyse, the

molecular mass, the compound name, the paths for the input/output folders, and the reaction ID in the metabolic model (e.g., BiGG ID). Then, the tool will convert biomass in gDW, and metabolites concentrations in mmol/L.

Data interpolations and fittings: For the growth interpolation, the data is fitted with a logistic equation of population growth using the *R* package "grofit" [17]. Then, for the equation fitting of the metabolite concentration data, including parameter estimation and confidence intervals of the parameter estimations the packages "minpack.lm" and "nlstool" [18] are used. The equation parameters are estimated using linear and non-linear least square regression from the core *R* package [18]. Then, artificial data with noise was generated within a 0.1 standard deviation; this is used to re-estimate the parameters with a noisy dataset using the linear and non-linear least square regression [18] and compare the fitting of the equation with the experimental data and with the noisy artificial data. This is a verification step to make sure the "nls" package is working with this dataset, as this package cannot fit artificial data without noise, so if a dataset is small (e.g. 3 or 4 points) the algorithm could not fit the data [18]

Uncertainty estimations: The confidence interval around the parameters estimation is calculated. The 95% CI calculated for the parameter estimation is calculated using the "confint" function from the "nlstool" package [18], the minimum and maximum values of the parameters are used to plot the fittings with the upper and lower parameters equation curves (this is only used to visualise the difference). Then, the residuals from the equation fitting are analysed. The residuals and standardised residuals are plotted against the fitted values to identify any pattern in the data and see if a linear or non-linear equation is more adapted to the dataset. The autocorrelation is analysed by plotting residuals at a time (i) against residuals at the

next time point (i+1). The normal Q-Q plot of standardised residuals is plotted to verify if some data points are poorly fitted (potential outliers). The confidence interval of the equation fit to the data is estimated using an inverse estimation approach, this computes the maximum likelihood within the 95% confidence interval using the "investr" package [19]. The maximum and minimum values of the 95% CI interval are used as a data input to fit a lower and upper equation. This leads to a max and min equation representing the upper and lower predicted 95% CI for the metabolite concentration.

For antibiotics export data, further analysis is done, because the equation used is a modified version of the logistic growth equation. First, the estimation of the parameters couples' confidence is analysed using Beale's 95% unlinearized confidence region with the "nlsConfRegions" function; followed with an analysis of the RSS contour using "nlsContourRSS" from the "nlstool" package [18]. In addition, a 500 bootstrap sampling using "nlsBoot" [18] was performed to further analyse the possible couples of parameters. The bootstrap parameter couples were inputted in the equation to plot the 500 curves showing the confidence interval on the concentration prediction. Another analysis included a Jackknife resampling to further estimate the two parameters' distribution showing the most probable couples.

Output: The tool gives multiple outputs to ensure the user can verify if each step worked properly. At the different steps of analysis, the tool generates plots visualised in the *R* notebook and as images in the output folder. The tool also outputs a file with the time points queried, the average metabolite concentrations (in mmol/L), and the upper and lower concentration of the 95% CI (based on the max and min equation of the inverse estimation method). This file also includes the biomass (in

gDW), the specific growth rate in (h^{-1}), the absolute flux, and the uptake/secretion rates. Finally, the tool also generates a script to constrain the exchange fluxes at the queried time points in a MATLAB script for COBRA and a Python script for COBRApy.

5.4.2 Experimental data for antibiotics and growth curve dataset

I acquired the experimental dataset used to test the ACT and RED data, which were performed using the WT plasmid-free ($SCP1^-$ $SCP2^-$) strain M145 of *Streptomyces coelicolor* A3(2) M145 [20]. All the media were based on ion-free water, and all chemicals used were of analytical grade. The growth medium used for the growth curves and for the antibiotics analysis was a minimal defined media developed for 'omics analysis of *S. coelicolor* [14,21], phosphate-limited and with glucose and glutamate as a carbon source and glutamate as a sole nitrogen source. The medium consisted of Na-Glutamate, 55.2 g/L; D-glucose, 40 g/L; $MgSO_4$, 2.0 mM; phosphate mix (NaH_2PO_4 and K_2HPO_4), 4.6 mM; supplemented minimal medium trace element solution SMM-TE (0.1 g/L of each of $ZnSO_4.7\ H_2O$, FeSO4.7 H2O, $MnCl_2.4$ H2O, CaCl2 and NaCl)[20], 8 mL/L; Trace Minimal Supplement 1 (TMS1: $FeSO_4.7\ H_2O$, 5 g/L; $CuSO_4.5\ H_2O$, 390 mg/L; $ZnSO_4.7\ H_2O$, 440 mg/L; $MnSO_4.H_2O$, 150 mg/L; $Na_2MoO_4.2\ H_2O$, 10 mg/L; $CoCl_2.6\ H_2O$, 20 mg/L, and HCl, 50 mL/L), 5.6 mL/L; Antifoam (Sigma 204), 1 ml/L. The pH was adjusted to pH 7 by adding NaOH 2M to the media. The cultures were grown in an incubator at 30°C and shaking at 250 rpm. Cultivations were performed in 250 mL baffled conical flasks (Simax), with a culture volume of 50ml, all the flasks have been coated (with SigmaCote) to reduce aggregation of the cells on the side of the flask. 10^8 of viable spores were inoculated per flask. The growth was estimated by OD_{450} measurements (3 measures per time point). The concentrations of coloured secondary

metabolites were spectrophotometrically measured by following a series of extraction and measures on a culture aliquot of 1mL for each time-point (*Supplementary Figure 5.5*), the amount of yCpk (yellow pigment) produced was estimated by measuring the absorption at 460nm in the supernatant after spinning down 1mL of culture (2 min at 12,000rpm). The Act (blue pigment) concentration was determined by measuring the absorption at 640nm after adding the same volume of 1M NaOH to the same volume of supernatant (0.7ml), an OD_{640} of 0.5 is equivalent to 60 µg/mL of Act [22]. In order to measure the concentration of RED, the mycelium pellet was re-suspended in 1 mL of HCl spun down (2 min at 8,000 rpm), the supernatant was discarded and the mycelium in the pellet re-suspended in 1 mL of 100% methanol and spun down (2 min at 12,000 rpm), the absorption in the supernatant was measured at 533 nm, an OD_{533} of 1 is equivalent to a concentration of 3.91 µg/mL of RED [22]. These conversion rates were used in the input of the tool. The dataset is summarised in *Supplementary Figure 5.6*.

5.4.3 Constraint-based metabolic modelling and ensemble modelling

The experimental data used for the constraint-based metabolic model came from a published chemostat dataset using the same media as the one described above [14]. This dataset contains measurements of the biomass (in gDW), phosphate (in mg/L), glucose (in g/L), glutamate (in g/L), actinorhodin (OD), and undecylprodigiosin (OD). This data was then analysed using the tool. The output uptake and secretion rates were used at each time point (at 24, 28, 30, 32, 34, 36, 38, 40, and 48 hours) to constrain the exchange fluxes in the genome-scale metabolic model. The genome-scale metabolic model used is the *i*AA1259 model of *S. coelicolor* A3(2) [15]. The model was analysed using Flux Balance Analysis (FBA) to predict optimal growth using the COBRA toolbox in

MATLAB and Python [23,24]. The constraints corresponding to the media composition were applied; the uptake fluxes for exometabolites not present were set to zero but were allowed to leave the metabolic system. The uptake rates are used to define constraints of the nutrient uptake for the model. The objective function maximized in the modelling was the growth rate. The constraints on glucose, phosphate, and glutamate uptake based on fermenter time-course data [14] enabled simulation of the growth across time. Finally, the predicted growth curve was compared with the experimental growth data.

An ensemble modelling approach was used by sampling exchange fluxes to constrain the model. Exchange constraints were sampled 500 times between the maximum and minimum value of the uptakes and export rates. The sampling method used divided the data range in 500 between the maximum and minimum flux calculated for each time point. The exchange fluxes were combined in sets of constraints in the order of maximum to minimum, this considers that if the strain reaches the max/min uptake of a given metabolite it will likely reach the max/min uptake or export for the other metabolites, as well as reaching the max/min at the next time point. These 500 sets of exchange fluxes were constrained on the model using COBRApy in a loop running for each time point and optimising the growth using FBA. The 500 predicted growth curves were plotted with the average predicted growth and the experimental data.

5.5 Results & Discussion

5.5.1 Exometabolic fluxes tool development

The tool was developed in the open-source statistical language *R*, the tool is based on an input/output structure, and the tool is in the format of an *R* Markdown to ensure reproducibility and reusability. The tool needs to be used for a single metabolite at a time. The outputs are displayed in the *R* notebook, as well as saved in a .csv and .png format in a folder. The quantitative experimental data is used as an input and it outputs a set of fluxes to constrain the genome-scale metabolic model (*Figure 5.1*). The tool structure is described below:

5.5.1.1 Input and conversions

One can input a data table in .csv format, containing time points (in hours), exometabolites concentrations (in mg/L, mmol/L, or in OD), and biomass measurements (in gram of dry weight or in OD). The user needs to input some key information at the start of the *R* script, such as the molecule molecular weight, the unit of inputs (e.g. OD or mg/L), culture volumes, uptake/secretion trend (i.e., linear or non-linear). The tool is available in two versions, for uptake fluxes and for secretion fluxes. Once the input is provided, the user just needs to "Run" the *R* Markdown notebook for the tool to generate the outputs.

The inputs are converted into suitable units using the *R* input (i.e., molecular weight, culture volume) to calculate fluxes and uptake/secretion rates. Metabolite concentrations in OD are converted to mg/L, and then converted to mmol/L. Biomass input in OD is converted to gram of dry weight (gDW). These conversions are necessary

to calculate fluxes in mmol/gDW/h. There are multiple quality control steps to visualize and validate the data analysis, for example, to verify that the equations fit the data.

5.5.1.2 Interpolation and calculations

To calculate fluxes at any time points the converted data are fitted with non-linear or linear equations. The growth curves were fitted using a logistic equation of growth (*Equation 5.1*) [17], where the growth parameters K (carrying capacity), N_0 (initial bacterial density), and r (growth rate) are estimated. Once the growth parameters are estimated, the equation is used to calculate the biomass in gram of dry weight at any time point.

The next step requires fitting of an equation to the concentrations measured across time. The equation fitted depends on the metabolite uptake trend: for a linear uptake the equation fitted is a linear equation (*Equation 5.2a*), and for a non-linear uptake it is a decaying logistic equation (*Equation 5.2b*). The equation fitted for the antibiotics secretion is an adapted version of the logistic population growth equation (*Equation 5.2c*). *Equation 5.3c* was derived from the logistic growth equation, two parameters were added, the first one that represents a constant flux (F in $mmol.gDW^{-1}.h^{-1}$) and another one delaying the production switch ON time (S in hours). As the flux is a derivative of concentration over biomass to time (*Equation 5.3a*), the biomass at a given time point is replaced by the logistic equation (with the previously estimated growth parameters). Then, at the given time point the flux is considered as constant (steady-state condition), it is multiplied by the logistic growth equation with the growth parameters values calculated earlier (*Equation 5.1*) and divided by the volume (V). The value S is subtracted to the time point t in the logistic equation (*Equation 5.1*);

$$G = \frac{K}{1 + \frac{(K - N0)}{N0} \times e^{-rt}}$$

Equation 5.1: Logistic equation for population growth

Logistic equation of population growth used to fit the biomass growth curve. G corresponds to the biomass (in gDW), K is the carrying capacity (maximum population size), N0 is the initial bacterial density, r is the growth rate (or division rate), and t is the time in hours. This equation is fitted to the growth curve using the "grofit" *R* package.

a) $$y = a + bx$$

b) $$y = \frac{a}{1 + b \times c^{x}}$$

c) $$y = F \times \frac{K}{1 + \frac{(K - N0)}{N0} \times e^{-r(x - S)}}$$

Equation 5.2: Equations fitted to the concentration measurements

a) Linear equation fitted to the concentration measurements with 2 parameters: a, and b. Y is the concentration and X is the time.
b) Non-linear equation fitted for metabolites uptake, with 3 parameters, a, b, and c. When c>1 it is a decaying logistic equation fitting the type of decay of limiting metabolites uptake like phosphate. Y is the concentration and X is the time.
c) Non-linear equation adapted from the logistic population growth model to fit the antibiotics secretion. The parameter S corresponds to the switch on of the antibiotics production. F is a parameter corresponding to a constant flux (in mmol/gDW/h). Y is the concentration and X is the time.

Note: for clarity the volume was omitted but the equations a, b, and c are divided by the volume to calculate the concentration in mmol/L.

this parameter represents the delayed "switch ON" time of the antibiotic production and export. Finally, the constant values F and S are parameters estimated to fit the *Equation 5.3c* to the antibiotics concentration data. The equations' parameters are estimated using linear and non-linear least square regression from the core *R*

a)

$$Flux = \frac{\mathrm{d}C \times V}{\mathrm{d}B \cdot \mathrm{d}t}$$

b)

$$Flux(t) = \frac{(C_t - C_{t-1}) \times V}{B_t \times t}$$

Equation 5.5: Flux calculation in mmol.gDW^{-1}.h^{-1}

a) Absolute flux calculation. C is the concentration in mmol/L, V is the culture volume in L, and B is the biomass in gDW, t is the time in hours.
b) Uptake and export rates function at a time t. This calculates the flux rates to use as constraints for the metabolic model. C is the concentration in mmol/L, V is the culture volume in L, and B is the biomass in gDW.

package [18]. Then, artificial data with noise was generated within a 0.1 standard deviation; this is used to re-estimate the parameters with a noisy dataset using the linear and non-linear least square regression [18] and compare the fitting of the equation with the experimental data and with the noisy artificial data. As the "nls" package cannot fit artificial data without noise, in case of a very small dataset (e.g. 3 or 4 points) the algorithm would not fit the data.

5.5.1.3 Uncertainty estimations

Once the parameters are calculated, the tool estimates the parameters uncertainty to take it into account when estimating the fluxes 95% confidence interval (CI). First, the tool calculates the 95% confidence interval for the parameters used in the equation fitting. These parameters are used to plot a first version of the fitted concentration with the upper and lower bounds of the confidence interval. Then, the residuals from the non-linear fit are analysed by generating plots of the residuals, to verify if the linear or non-linear equation fitted is appropriate for the dataset. The tool generates

multiple plots: one of the non-transformed residuals against the fitted values, another one of the standardised residuals against fitted values; a plot of the square root of the absolute value of the standardised residuals against fitted values, a plot of the auto-correlation of the residuals; a plot of the histogram for the non-transformed residuals, and a plot of the normal Q-Q plot for the standardised residuals [18,25]. The estimated parameters for the antibiotics production are further analysed, as the equation fitted is not a usual linear or non-linear equation (like *Equation a*, and *b*) but it contains an unusual parameter (S in *Equation c*) delaying the switch ON time. Beale's 95% confidence interval is calculated for each parameter; from this the tool analyses the residual sum of squares (RSS) contours of the two parameters to visualise the area of 95% confidence of the possible couples of parameters. Then, in all cases, the confidence band of the equation fit is estimated and plotted using an inverse estimation approach with the 95% confidence interval [19]. This is followed by the equation fitting to the maximum and minimum predicted values from the inverse estimation steps, hence defining the max and min equation parameters. The equations with the max and min parameters are used to predict the upper and lower values of the concentration over time at each time point, which is converted in the maximum and minimum fluxes.

5.5.1.4 Outputs

The predicted concentrations (in mmol/L) are calculated at the queried time points using the fitted equation as well as the maximum and minimum concentrations of the 95% CI. The biomass (in gDW) and the specific growth rate (in h^{-1}) are calculated at the

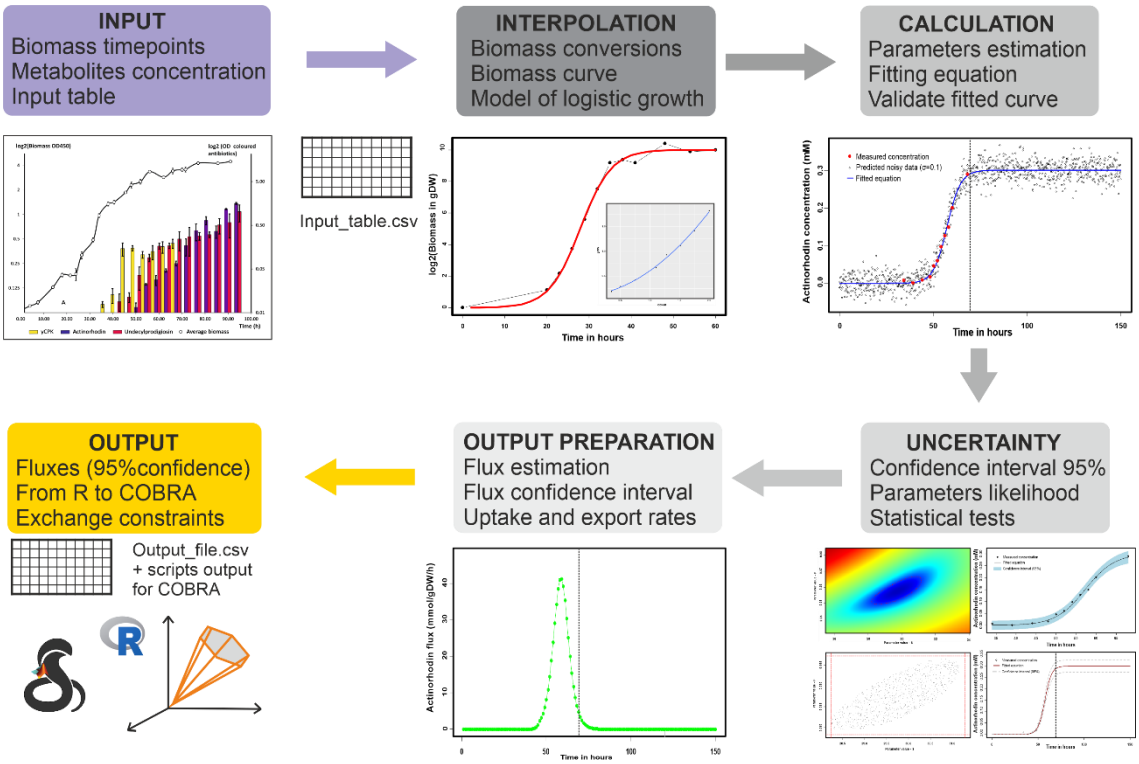


Figure 5.1: *R* tool pipeline structure to estimate exometabolic flux constraints

Description of the pipeline of different analysis of data processing applied to the input data to get an output with exometabolic fluxes to be used as constraints in the genome-scale metabolic model. Here the data for Actinorhodin was used as an input to generate the figures above.

- INPUT: data input as a table with biomass time series and metabolite concentrations. The data is then converted in the units of interest (e.g. concentration in mM).
- INTERPOLATION: biomass can be converted in gram of dry weight. The growth is fitted to a logistic growth curve interpolating growth between time points.
- CALCULATION: a linear or non-linear equation is fitted to the metabolite concentrations, the equations parameters are estimated and refined.
- UNCERTAINTY: the parameters 95% confidence interval (CI) is estimated, following multiple statistical tests. The 95% CI of the equation fitted is estimated.
- OUTPUT PREPARATION: the output is prepared by calculated different values at the queried time points, such as the exchange rates and absolute flux, the biomass, the growth rate, and the exchange rates with the upper and lower values based on the 95% CI.
- OUTPUT: the tool generates an output table with the fluxes and biomass. The tool also generates scripts for COBRA (Python and MATLAB) with the exchange constraints.

time points. The absolute fluxes are calculated by calculating the derivative of the concentration (multiplied by the volume) to the biomass to the time (Equation 5.3a). The reason for calculating the absolute flux is that it is interesting to see the flux pattern; for example, the maximum flux of some nutrient uptakes will be at the exponential phase and antibiotics production at transition and stationary phase. The constraint-based metabolic models use uptake and export rates (in $mmol.gDW^{-1}.h^{-1}$) as constraints so the tool also calculates these exchange rates. It takes the change in metabolites concentration (in mM) between the time point and the last time point queried, multiply it by the volume (in L), divide these by the biomass (in gDW) multiplied by the time (in hours) see *Equation 5.3b*. This is similar to the method used for the "conc2Rate" function in COBRA Metabotools [26]. The tool generates plots for the flux to time and flux to concentration to visualize the flux dynamics across time. Finally, the tool automatically writes scripts lines for COBRApy and COBRA MATLAB [23,24] to constrain the exchange fluxes at the time points queried with the mean flux, the maximum, and the minimum fluxes at the edge of the 95% CI. All the plots and results tables (i.e., converted data, predicted data, and flux outputs) are saved in the output folder.

The tool is automated, but it is critical to verify that the fitting methods and the different steps are appropriate for the data input; this is to avoid using an output when issues occur such as equations not fitting datasets with any decrease or increase pattern. That is why the tool exports multiple analyses of the fitting and uncertainty estimations to let the user judge if the data is properly analysed. The tool is modular and available in an *R* Markdown notebook format to ease modification, and tests of the script by chunks to adapt it to the data used when necessary.

5.5.2 Initial tool tests of antibiotics flux analysis

A first test of the tool consisted of validating the equation for antibiotics export (*Equation 5.3c*); to this end, the tool was tested using an experimental dataset with measurements of biomass OD, as well as actinorhodin and undecylprodigiosin OD measurements collected over time. The data was inputted in the tool, where the biomass was converted from OD to gDW using a conversion curve (*Supplementary Figure 5.1a)*. The logistic growth curve was then fitted to the biomass data (*Supplementary Figure 5.1b*). The OD for undecylprodigiosin was the input and converted to mM (as explained in the *Method* section). The tool fitted the equation by estimating the parameters S and F for the *Equation 5.3c*; these parameters were refined by introducing noise in the data. The resulting equation is in agreement with the measured concentrations (*Figure 5.2a*). The tool also estimated the 95% confidence interval for the concentrations predictions using the equation (*Figure 5.2b*). The tool analysed the uncertainty around parameter estimations, such as the 95% CI for parameters couples (*Figure 5.2c* and 5.*2d*). Further parameters analyses are available in *Supplementary Figure 5.2*. The probability to find the two parameters (95% CI) within a given range is limited to a tight window (dotted red line *in Figure 5.2d*). This suggests that early switch ON of the production/export necessitate a lower flux, whereas a late switch ON would necessitate a higher flux to reach the same concentrations. It indicates a trade-off between possible switch ON time parameters S for the secondary metabolites production and the maximum constant flux (F). The predicted concentrations using the estimated parameters are in good agreement with the measured concentrations of coloured antibiotics (*Figure 5.2e*). This is confirmed by a low residual sum of squares (RSS) for undecylprodigiosin fitting (RSS=$6.656*10^{-5}$,

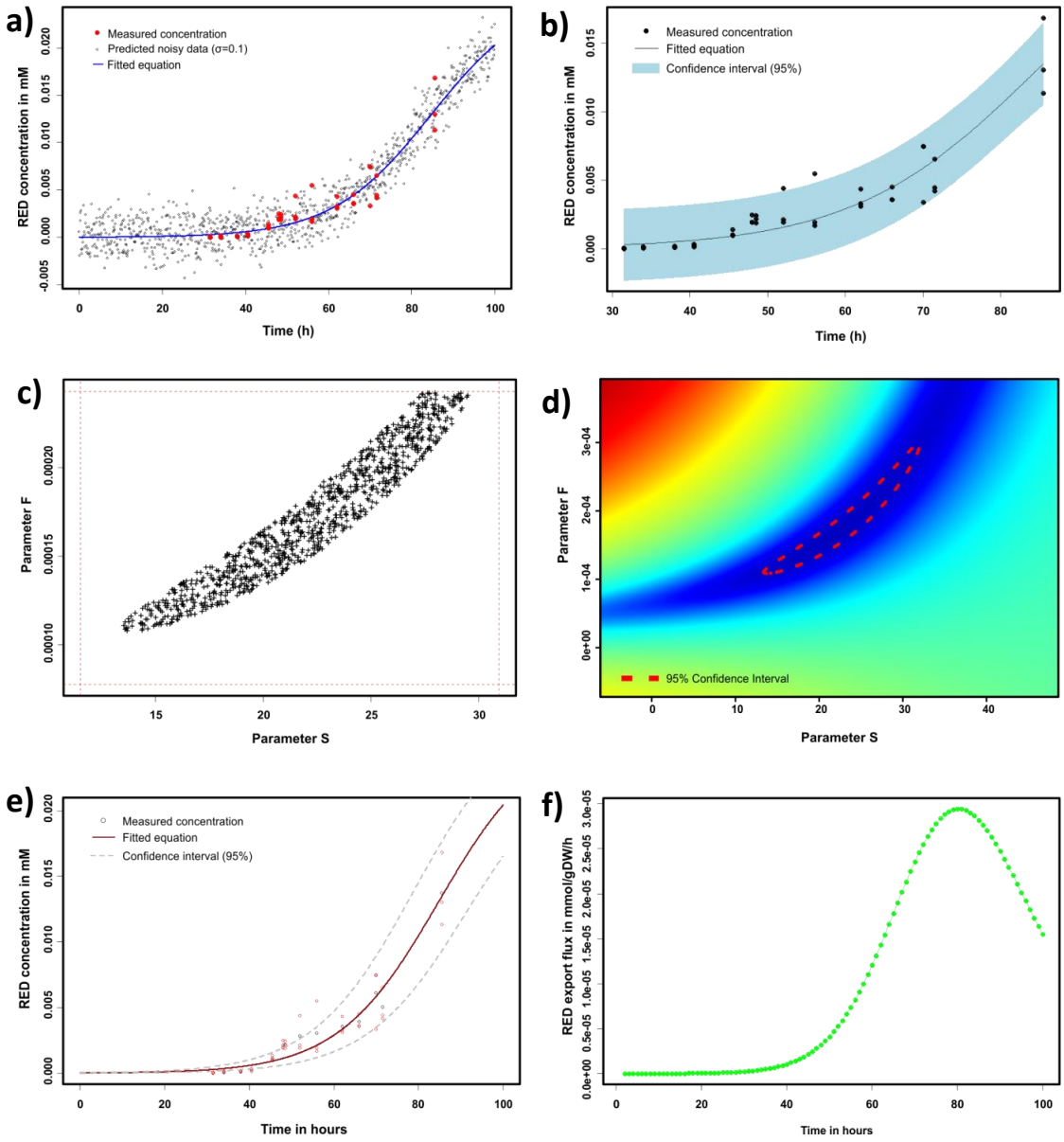


Figure 5.2: Outputs for the analysis of undecylprodigiosin by the tool

a) Fitting of the equation to the data with noisy simulated data around a sigma=0.1
b) Fitting of the equation to the data with the 95% CI band
c) Couple parameters Beale's 95% unlinearized confidence region
d) Couple parameters RSS contour with 95% CI
e) Fitted equation with upper and lower values of the 95% CI for the predicted concentration
f) Absolute undecylprodigiosin export flux over time

details in *Electronic Supplementary 5.1 - Notebook RED 1*). Finally, the tool calculated the export absolute fluxes and secretion rates (*Figure 5.2f*), the flux to time has a bell shape showing a maximal flux at 80h (corresponds with the start of stationary phase). For actinorhodin, the same plots are available in the *Supplementary Figure 5.3*. This is confirmed by a low residual sum of squares (RSS) for actinorhodin fitting (RSS=0.3437, details *in Electronic Supplementary - Notebook ACT 1*). The tool used the starting value for the parameter S (at 40h) as no solution was found for the fitting when the parameters S fitting was unconstrained; this appears on the parameters residuals analysis (*Supplementary Figure 5.4*).

5.5.3 Tool testing with constraint-based metabolic modelling

The tool needed further validation with a larger dataset, and a demonstration of the tool usage by using the tool outputs to constrain a constraint-based metabolic model. For this, a chemostat dataset [14] previously used to validate the *i*AA1259 model of *S. coelicolor* A3(2) [15] was used as an input for the tool. This dataset contains measurements of phosphate, glucose, and glutamate concentrations, as well as OD measurements of ACT and RED, and biomass measurements. The tool analysed these data: the phosphate uptake was a non-linear uptake, while glucose and glutamate were linear uptakes (*Figure 5.3*). For all three metabolites, the tool fittings and the 95% CI calculations were in good agreement with the experimental data (*Figure 5.3*a-i). The residuals analysis for all three metabolites showed that the fittings were reasonable (*Figure 5.3j-l*). Finally, the tool calculated and exported the uptake fluxes for all three metabolites (*Figure 5.3m-o*). The uptake fluxes for the phosphate peaks during the exponential phase dropped rapidly; this makes sense as phosphate is the limiting

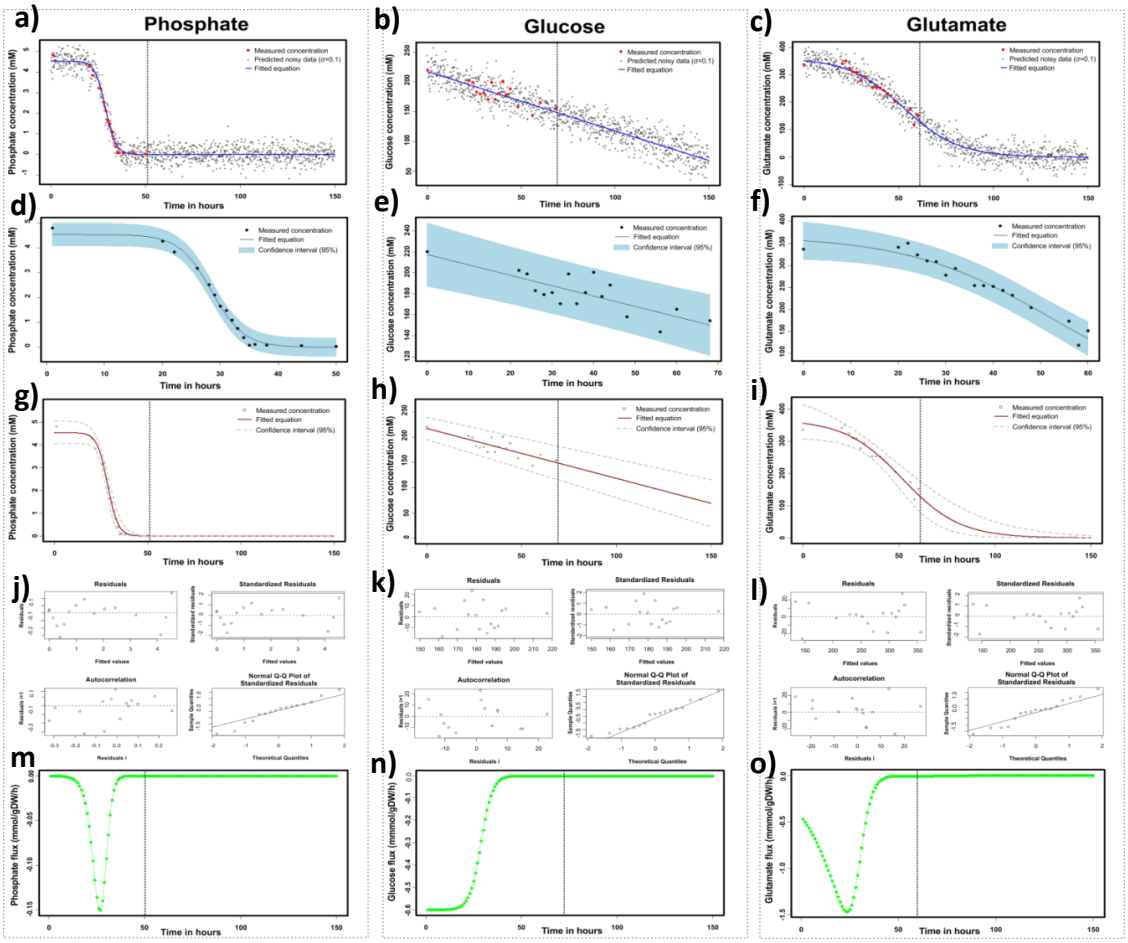


Figure 5.3: Analysis of exometabolome data for phosphate, glucose, and glutamate.

a) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (phosphate)
b) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glucose)
c) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glutamate)
d) Fitting of the equation to the data with the 95% CI band for the phosphate data
e) Fitting of the equation to the data with the 95% CI band for the glucose data
f) Fitting of the equation to the data with the 95% CI band for the glutamate data
g) Fitting of the equation to the data with the refined 95% CI band for the phosphate data
h) Fitting of the equation to the data with the refined 95% CI band for the glucose data
i) Fitting of the equation to the data with the refined 95% CI band for the glutamate data
j) Residuals analysis for the parameters estimation during equation fitting (phosphate)
k) Residuals analysis for the parameters estimation during equation fitting (glucose)
l) Residuals analysis for the parameters estimation during equation fitting (glutamate)
m) Uptake flux values across times for the phosphate
n) Uptake flux values across times for the glucose
o) Uptake flux values across times for the glutamate

metabolite in this chemostat experiment (*Figure 5.3m*). The uptake fluxes for the glutamate reaches its maximum at early exponential phase then decreases slowly until the stationary phase (*Figure 5.3o*). The glucose uptake fluxes are following the growth in an inverse trend, where uptake fluxes are maximal at the start then decreases following the exponential phase and reaching near 0 around stationary phase (*Figure 5.3n*). The data for the coloured antibiotics ACT and RED were also analysed using the tool (*Figure 5.4* and *5.5*). The fitting for actinorhodin and the re-estimated fittings with the 95% CI are in good agreement with the data and show a close fit (*Figure 5.4a-c*). The confidence interval for the parameters estimation showed a narrow window of possible parameter couples (*Figure 5.4e* and *5.4f*); in a similar fashion as in test on the ACT and RED data described previously (*Figure 5.2c* and *5.d*). The residuals analysis also confirms that the parameters used for the fitting are appropriate for the measured ACT concentrations (*Figure 5.4g*). Finally, the tool calculated the export fluxes for ACT (*Figure 5.4d*); the fluxes start at the transition phase, rapidly reach a maximum during stationary phase, and then drop again. This suggests that the peak of production of actinorhodin is in a short period of time (20 to 30 h long). The same analysis was carried on the undecylprodigiosin data; the fitting for the RED and the re-estimated fittings with the 95% CI are in agreement with the data, but the last data point seems slightly off the fitted trend, but it is still in the 95% CI (*Figure 5.5a-c*). This misfit point needs to be taken into account in the analysis as the RED concentration might be higher than predicted after 70h. The confidence interval for the parameter estimation showed a narrow window of possible parameter couples (*Figure 5.5e* and *5.5f*), in a similar fashion as in testing on the ACT and RED data described previously (*Figure 5.2c* and *5.2d*), as well as ACT in the same experiment

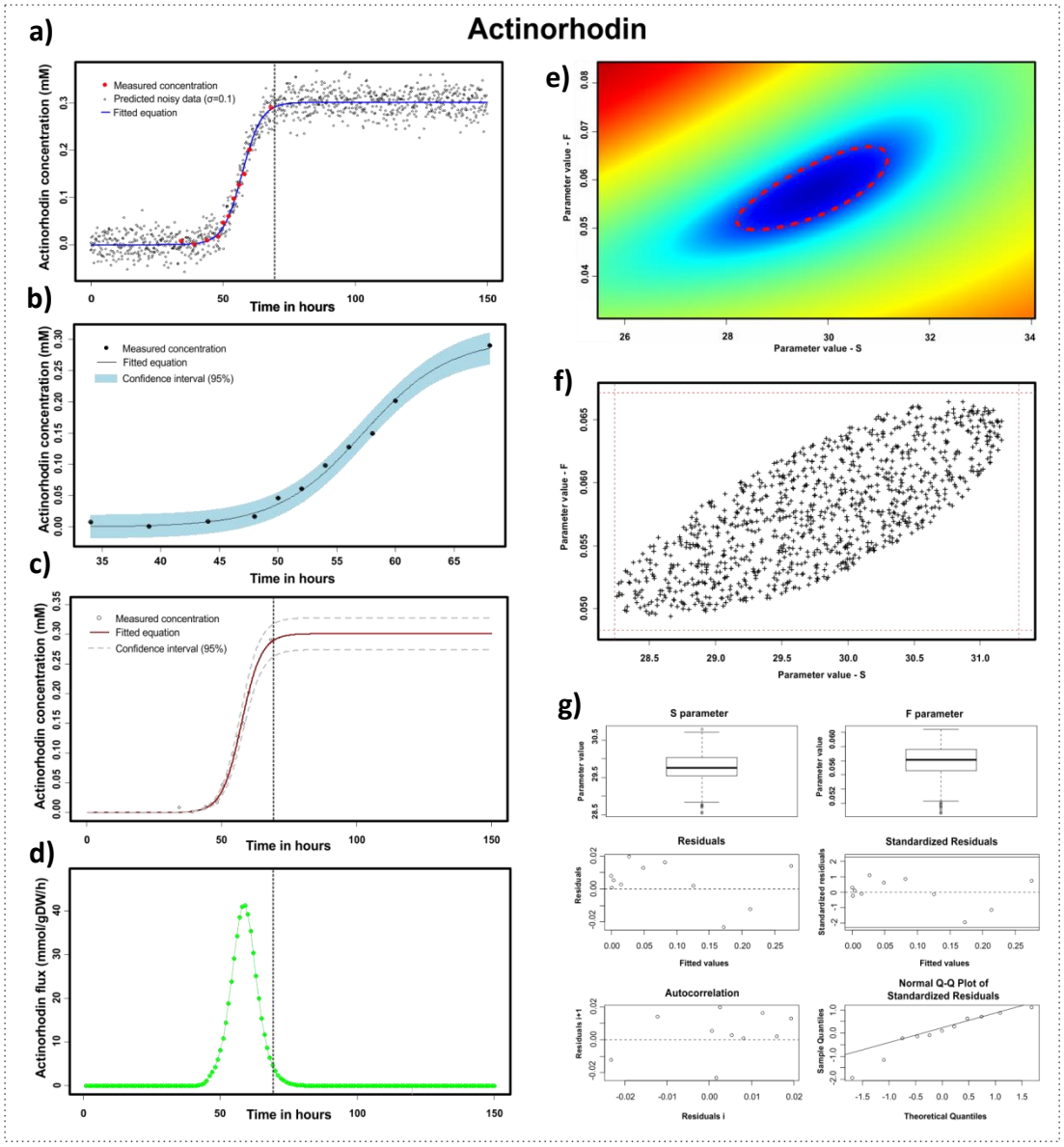


Figure 5.4: Analysis of Actinorhodin input data

a) Fitting of the equation to the concentration data with noisy simulated data sigma=0.1
b) Fitting of the equation to the ACT concentration data with the 95% CI band
c) Fitting of the equation to the ACT data with the refined 95% CI band
d) Calculated ACT secretion flux (in mmol/gDW/h)
e) Equation parameters couples probability (S&F) using RSS surface (95% CI in dotted red)
f) Equation parameters couples probability (S&F) using bootstrapping
g) Residuals analysis for the parameters estimation of the equation fitting

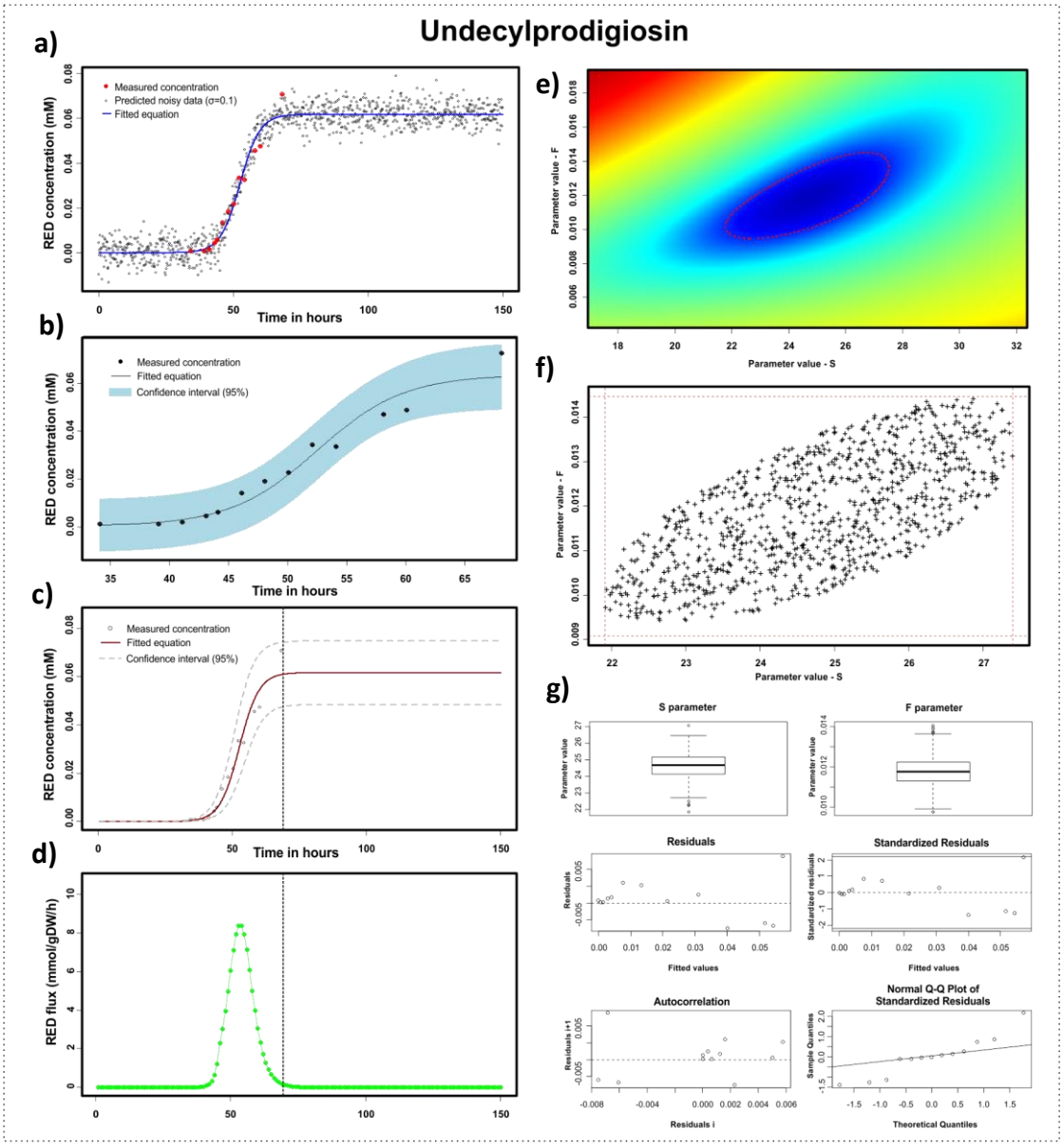


Figure 5.5: Analysis of Undecylprodigiosin input data

a) Fitting of the equation to the concentration data with noisy simulated data sigma=0.1
b) Fitting of the equation to the RED concentration data with the 95% CI band
c) Fitting of the equation to the RED data with the refined 95% CI band
d) Calculated RED secretion flux (in mmol/gDW/h)
e) Equation parameters couples probability (S&F) using RSS surface (95% CI in dotted red)
f) Equation parameters couples probability (S&F) using bootstrapping
g) Residuals analysis for the parameters estimation of the equation fitting

(*Figure 5.4e* and *5.4f*). The residuals analysis also confirms that the parameters used for the fitting seem appropriate for the measured RED concentrations (*Figure 5.5g*). However, the Q-Q plot (*Figure 5.5g*) shows that the last point is violating the assumption that the data have a normal distribution, but not the rest of the data that are on a straight line. The difference observed with this point might come from a noisy measurement, but it is unclear and should be taken into account when using the data after this time point as the concentration and flux values could be underestimated. Finally, the tool calculated the export fluxes for RED (*Figure 5.4d*); the fluxes starts at the transition phase to rapidly reach a maximum during stationary phase and drops back. This suggests that the peak of production of undecylprodigiosin similarly to actinorhodin is on a short period of time (20 to 30h long).

The output flux analysis using the Nieselt et al. chemostat data [14] indicates that the strain prefers consumption of glucose in lag phase and early exponential phase, once in early exponential, the strain increases its glutamate consumption. Then it rapidly increases its consumption of phosphate, the concentration of which drops quickly, triggering the strain to undergo a metabolic switch due to the phosphate depletion. Finally, the peak of production is reached after the transition phase for the two coloured antibiotics ACT and RED. This is in accordance with the metabolic measurement of *S. coelicolor A3*(2) in this phosphate-limited medium [14,27]. The metabolites' uptake or secretion concentrations from this dataset are summarized in *Figure 5.6*.

5.5.4 Application to ensemble constraint-based metabolic modelling

The final aim of this tool is to use the uptake and secretion rates calculated to constrain genome-scale metabolic models. The uptake and secretion fluxes calculated at the different time points by the tool were constrained in the *i*AA1259 metabolic model of *S. coelicolor* [15]. The metabolic model was constrained across the different time points of growth (*Figure 5.6a*). The phosphate, glutamate, and glucose uptake rates were also constrained (*Figure 5.6b*), as well as the ACT and RED production rates (*Figure 5.6c*). The uptakes rates were then automatically input in a script to predict the bacteria growth rate using FBA. The metabolic model biomass predictions were in good agreement with the experimental growth data (*Figure 5.7*). Finally, the tool output was also used to illustrate the application of an ensemble modelling approach, where at each time point 500 FBA simulations were done between the upper and lower values of the 95% CI for the flux constraints. Instead of only analysing the metabolic predictions with a single optimal result given by FBA, the ensemble modelling approach offers the opportunity to explore the uncertainty around the possible metabolic solutions and enables a more careful analysis of metabolic modelling results [13]. Here, the sampling method used is simplified (as described in *Method*) to illustrate the ensemble modelling approach. However, when applying an ensemble modelling approach it requires use of an appropriate sampling method taking into account the data distribution (normal or log-normal), the data autocorrelation at a time point with the next one, as well as how the exchange fluxes depends on each other (e.g., does high phosphate uptake implies a more likely pairing with high glucose uptake?). The 500 simulations predict a range of possible states for the metabolic model where the predicted growth is in agreement with the data for all

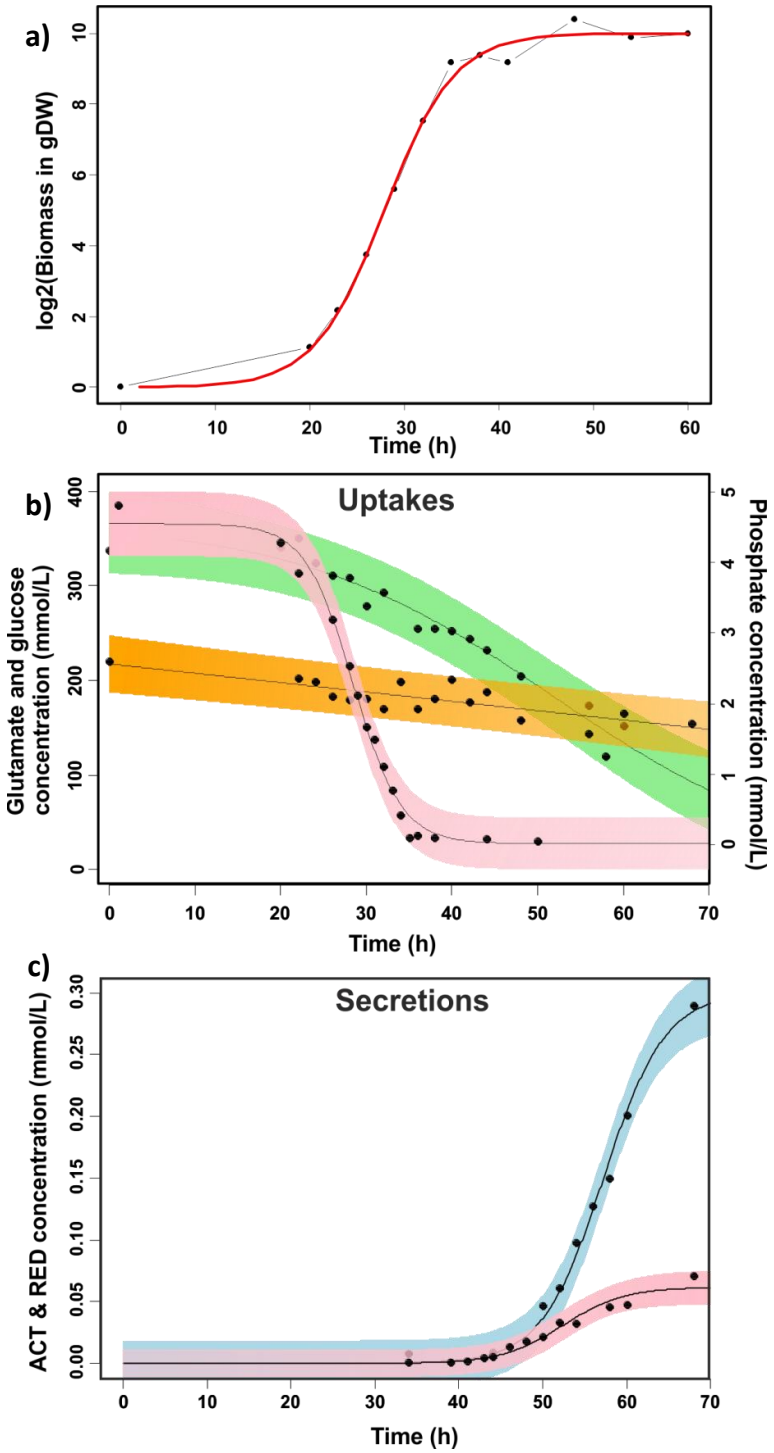


Figure 5.6: Summary of the tool predictions of biomass and concentrations by the tool for the input dataset

a) Growth curve for *S. coelicolor* in the chemostat experiment, in red the logistic equation fitted by the tool, in black the experimental data.

b) All the metabolites concentration (black points) with the fitted equation (black line), and the 95% CI band (green: glutamate, pink: phosphate, orange: glucose).

c) Antibiotics concentrations (black points) with the fitted equation (black line), and the 95% CI band (blue: actinorhodin, red: undecylprodigiosin).

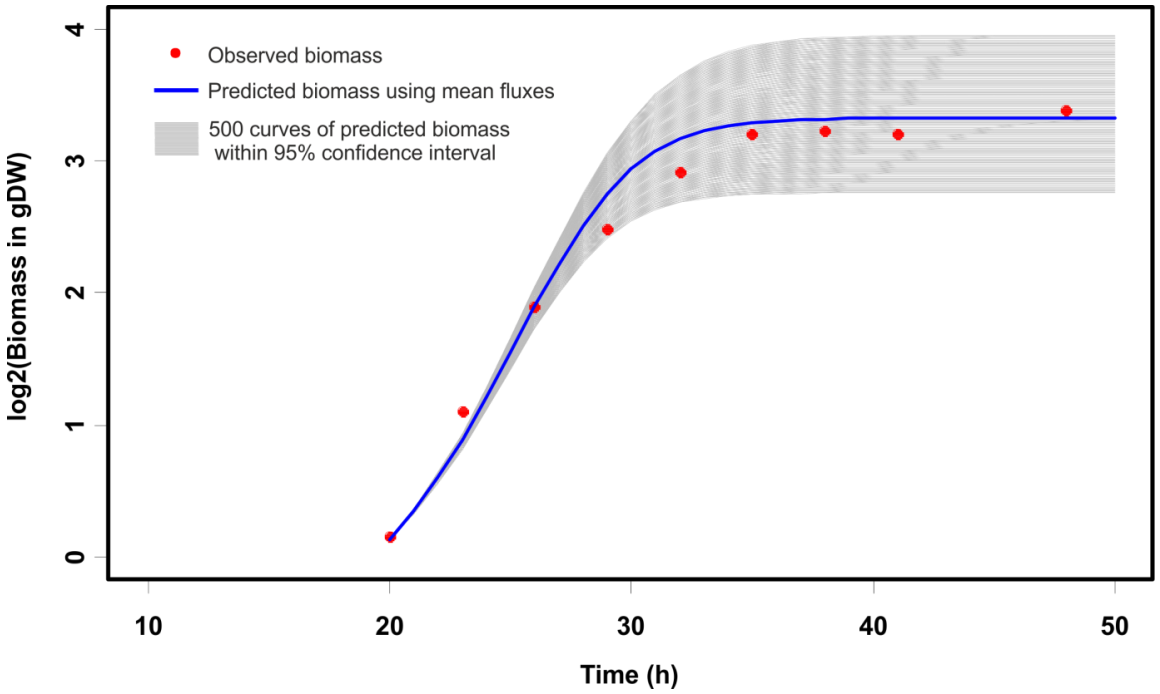


Figure 5.7: Ensemble modelling based on sampling of the exchange rates within the confidence interval

The main growth prediction using the mean uptake rates calculated by the model (in blue) is compared to the experiment growth data (red points). This shows that the predictions are in good agreement with the experimental. The 500 predictions using exchange (grey lines) rates sampled within the 95% CI of exchange fluxes from the tool are compared to the experimental data (red points). The 500 predicted growth curves form a range of possible growth states that includes most of the experimental points, with a similar optimum at exponential phase then more difference from transition to stationary phase suggesting more uncertainty for the metabolic predictions.

of these (*Figure 5.7*). There is an interesting feature to these predictions: in exponential phase the prediction range is very tight suggesting that the optimum is in a narrow zone during exponential growth (*Figure 5.7*). In contrast, in the transition and stationary phases the possible solutions are more diverse, suggesting that the uncertainty in determining the metabolic state at this point is higher. This analysis shows that the tool is adapted to analyse the input data and convert it into constraints for genome-scale metabolic models. Furthermore, this tool facilitates the application of ensemble modelling methods to popular constraint-based metabolic methods like FBA using COBRA.

5.6 Conclusion

Here an automated *R* tool to calculate exchange rates and their confidence interval for constraint-based metabolic modelling was developed and tested for time series data. The tool was tested with two different datasets, one based on our experimental measurements in a triplicate culture in flasks and one based on chemostat data [14]. The tool output was then used to constrain a *S. coelicolor* metabolic model [15] and the predictions were validated with the chemostat experimental data [14]. Finally, the output was also used for an ensemble modelling approach generating 500 metabolic predictions by sampling exchange fluxes within the confidence interval.

This tool helps to bridge the synthetic biology test and design phase by automatically using test data to integrate it in the metabolic modelling, to study the metabolic state of a strain following an engineering cycle. Nowadays synthetic biology is less limited by the pace of the building phase, but more by the ability to analyse engineered strains, learn from the observed phenotype and feed this information back into the design phase [7]. The tool is modular and open-source (using *R* Markdown) so anyone can adapt it to their datasets. The tool also helps to take into account the uncertainty inherent to biological systems by estimating the confidence intervals and including the tool outputs for use with ensemble modelling approach. Taking uncertainty into account in the modelling will help to increase the robustness of the predictions by allowing a more informed assessment of the confidence in individual results [13,28]. In the future, this type of tool would ideally have a user-friendly interface that can integrate test data into a constraint-based genome-scale metabolic model of the user's choice with ease. The development of more user-friendly tools would help the experimentalists to quickly test their engineered strain(s), analyse the effect of the

modification on metabolism, then predict new targets. Which would accelerate the DBTL cycle by putting into the hands of any synthetic biologist the power of constraint-based metabolic modelling.

5.7 References

(1) Lee, S. Y.; Kim, H. U.; Chae, T. U.; Cho, J. S.; Kim, J. W.; Shin, J. H.; Kim, D. I.; Ko, Y.-S.; Jang, W. D.; Jang, Y.-S. A Comprehensive Metabolic Map for Production of Bio-Based Chemicals. *Nat. Catal.* **2019**, *2* (1), 18–33.

(2) Chubukov, V.; Mukhopadhyay, A.; Petzold, C. J.; Keasling, J. D.; Martín, H. G. Synthetic and Systems Biology for Microbial Production of Commodity Chemicals. *npj Syst. Biol. Appl.* **2016**, *2* (1), 16009.

(3) Breitling, R.; Takano, E. Synthetic Biology Advances for Pharmaceutical Production. *Curr. Opin. Biotechnol.* **2015**, *35*, 46–51.

(4) Nielsen, J.; Keasling, J. D. Engineering Cellular Metabolism. *Cell* **2016**, *164* (6), 1185–1197.

(5) Medema, M. H.; Breitling, R.; Bovenberg, R.; Takano, E. Exploiting Plug-and-Play Synthetic Biology for Drug Discovery and Production in Microorganisms. *Nat. Rev. Microbiol.* **2011**, *9* (2), 131–137.

(6) Cardinale, S.; Arkin, A. P. Contextualizing Context for Synthetic Biology–Identifying Causes of Failure of Synthetic Biological Systems. *Biotechnol. J.* **2012**, *7* (7), 856–866.

(7) Cameron, D. E.; Bashor, C. J.; Collins, J. J. A Brief History of Synthetic Biology. *Nat. Rev. Microbiol.* **2014**, *12*, 381–390.

(8) Schellenberger, J.; Que, R.; Fleming, R. M. T.; Thiele, I.; Orth, J. D.; Feist, A. M.; Zielinski, D. C.; Bordbar, A.; Lewis, N. E.; Rahmanian, S.; et al. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6* (9), 1290–1307.

(9) O'Brien, E. J. J.; Monk, J. M. M.; Palsson, B. O. Ø. Using Genome-Scale Models to Predict Biological Capabilities. *Cell* **2015**, *161* (5), 971–987.

(10) Aurich, M. K.; Paglia, G.; Rolfsson, Ó.; Hrafnsdóttir, S.; Magnúsdóttir, M.; Stefaniak, M. M.; Palsson, B. Ø.; Fleming, R. M. T.; Thiele, I. Prediction of Intracellular Metabolic States from Extracellular Metabolomic Data. *Metabolomics* **2015**, *11* (3), 603–619.

(11) Cakir, T.; Efe, C.; Dikicioglu, D.; Hortacsu, A.; Kirdar, B.; Oliver, S. G. Flux Balance Analysis of a Genome-Scale Yeast Model Constrained by Exometabolomic Data Allows Metabolic System Identification of Genetically Different Strains. *Biotechnol Prog* **2007**, *23* (2), 320–326.

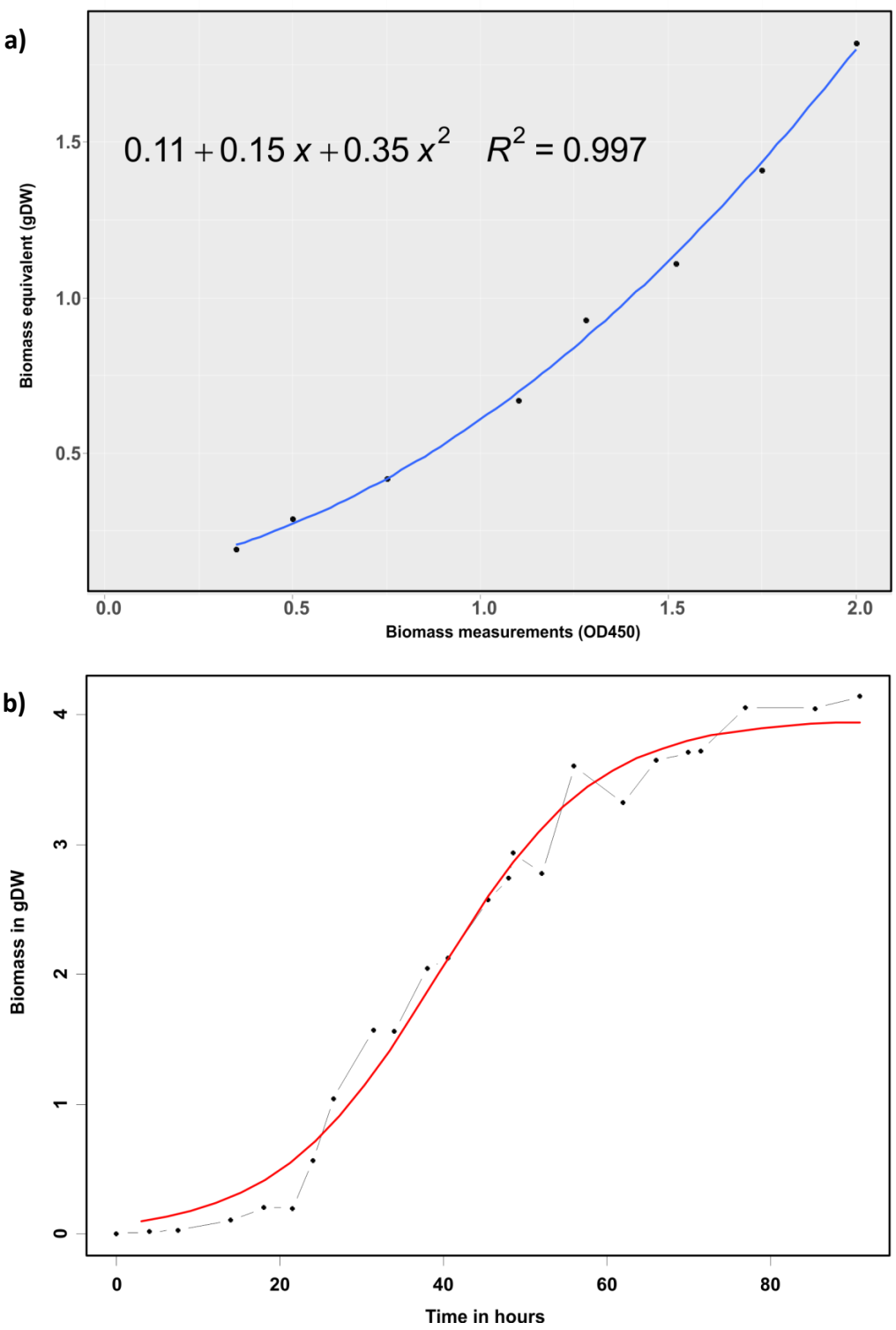
(12) Mo, M. L.; Palsson, B. Ø.; Herrgard, M. J. Connecting Extracellular Metabolomic Measurements to Intracellular Flux States in Yeast. *BMC Syst Biol* **2009**, *3*, 37.

(13) Tsigkinopoulou, A.; Baker, S. M.; Breitling, R. Respectful Modeling: Addressing Uncertainty in Dynamic System Models for Molecular Biology. *Trends Biotechnol.* **2017**, *35* (6), 518–529.

(14) Nieselt, K.; Battke, F.; Herbig, A.; Bruheim, P.; Wentzel, A.; Jakobsen, Ø. M.; Sletta, H.; Alam, M. T.; Merlo, M. E.; Moore, J. The Dynamic Architecture of the

Metabolic Switch in *Streptomyces Coelicolor*. *BMC Genomics* **2010**, *11* (1), 10.
(15) Amara, A.; Takano, E.; Breitling, R. Development and Validation of an Updated Computational Model of *Streptomyces Coelicolor* Primary and Secondary Metabolism. *BMC Genomics* **2018**, *19* (1), 519.
(16) Takano, E. PpGpp and Antibiotic Production in Streptomyces Coelicolor A3(2). **1993**.
(17) Kahm, M.; Hasenbrink, G.; Lichtenberg-Fraté, H.; Ludwig, J.; Kschischo, M. **Grofit** : Fitting Biological Growth Curves with *R*. *J. Stat. Softw.* **2010**, *33* (7), 1–21.
(18) Team, R. C. R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing. 2017.
(19) Greenwell, B. M.; Schubert Kabban, C. M. Investr: An R Package for Inverse Estimation. *R J.* **2014**, *6* (1), 90–100.
(20) Kieser, T. *Practical Streptomyces Genetics*; John Innes Foundation, 2000.
(21) Wentzel, A.; Bruheim, P.; Øverby, A.; Jakobsen, Ø. M.; Sletta, H.; Omara, W. A. M.; Hodgson, D. A.; Ellingsen, T. E. Optimized Submerged Batch Fermentation Strategy for Systems Scale Studies of Metabolic Switching in *Streptomyces Coelicolor* A3(2). *BMC Syst. Biol.* **2012**, *6* (1), 59.
(22) Bystrykh, L. V; Fernández-Moreno, M. A.; Herrema, J. K.; Malpartida, F.; Hopwood, D. A.; Dijkhuizen, L. Production of Actinorhodin-Related" Blue Pigments" by *Streptomyces Coelicolor* A3 (2). *J. Bacteriol.* **1996**, *178* (8), 2238–2244.
(23) Ebrahim, A.; Lerman, J. A.; Palsson, B. Ø.; Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7* (1), 74.
(24) Becker, S. A.; Feist, A. M.; Mo, M. L.; Hannum, G.; Palsson, B. Ø.; Herrgard, M. J. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox. *Nat. Protoc.* **2007**, *2* (3), 727–738.
(25) Bates, D. M.; Watts, D. G. *Nonlinear Regression Analysis and Its Applications*; Wiley, 1988.
(26) Aurich, M. K.; Fleming, R. M. T.; Thiele, I. MetaboTools: A Comprehensive Toolbox for Analysis of Genome-Scale Metabolic Models. *Front. Physiol.* **2016**, *7*, 327.
(27) Thomas, L.; Hodgson, D. A.; Wentzel, A.; Nieselt, K.; Ellingsen, T. E.; Moore, J.; Morrissey, E. R.; Legaie, R.; Wohlleben, W.; Rodriguez-Garcia, A.; et al. Metabolic Switches and Adaptations Deduced from the Proteomes of *Streptomyces Coelicolor* Wild Type and PhoP Mutant Grown in Batch Culture. *Mol. Cell. Proteomics* **2012**, *11* (2), M111.013797-M111.013797.
(28) Tsigkinopoulou, A.; Hawari, A.; Uttley, M.; Breitling, R. Defining Informative Priors for Ensemble Modeling in Systems Biology. *Nat. Protoc.* **2018**, *13* (11), 2643–2663.

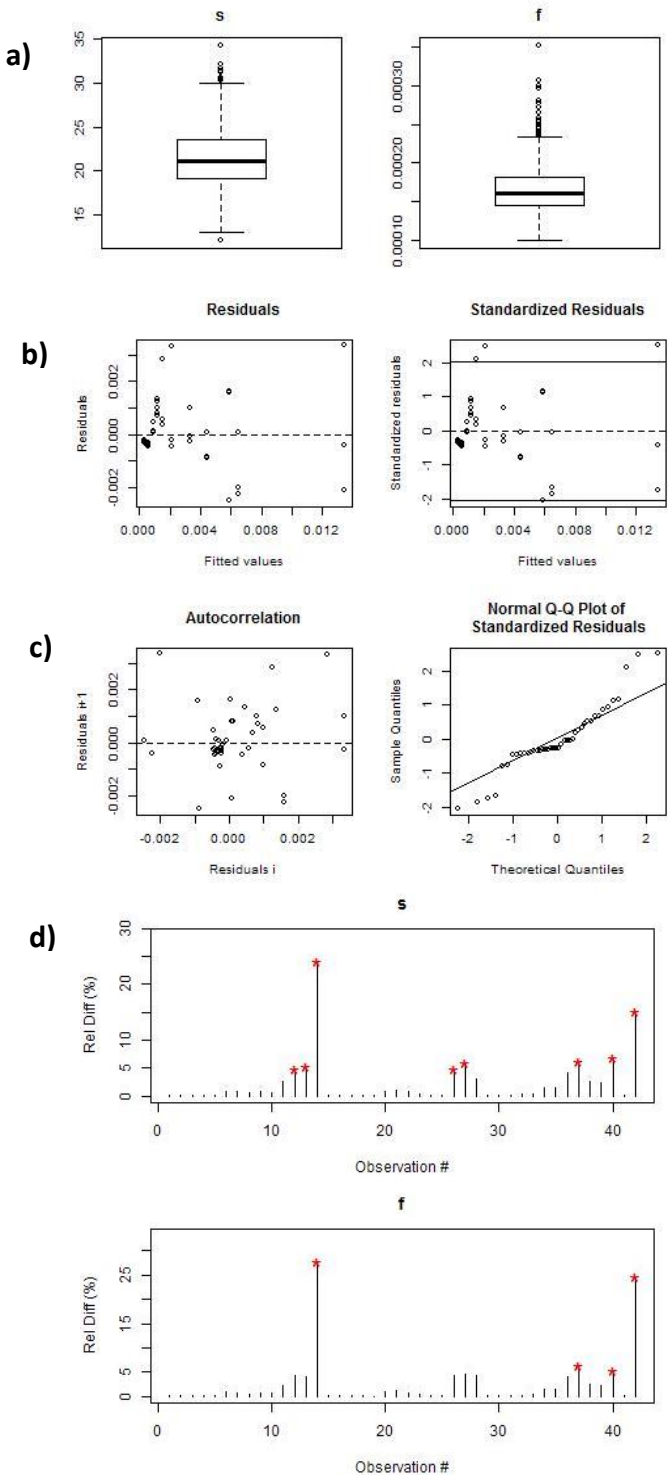
5.8 Supplementary Files



Supplementary Figure 5.1: Biomass conversion and fitting by the tool for the first dataset tested

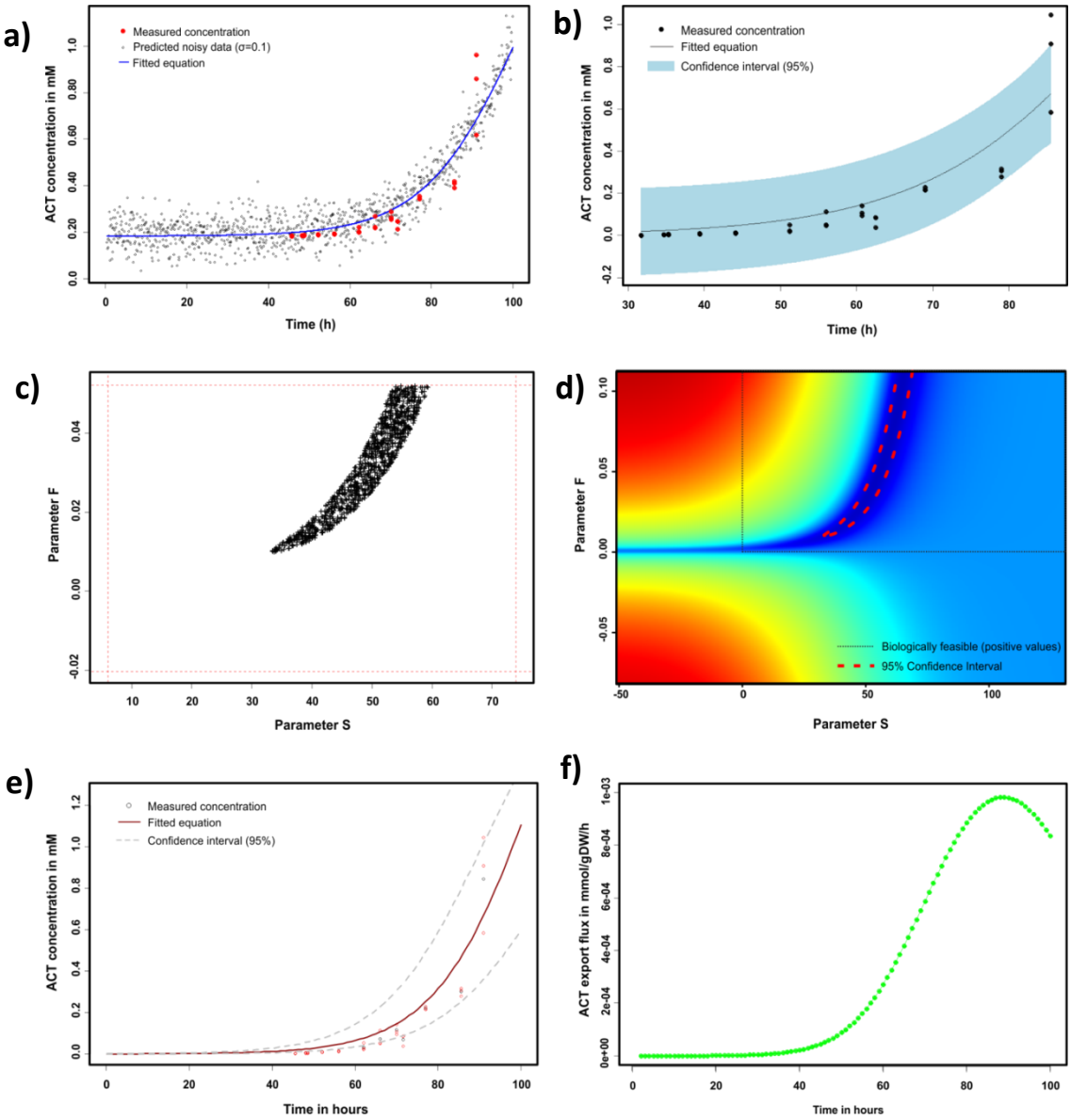
a) Curve of calibration of biomass OD450 to biomass in gDW. Blue line corresponds to the fitted equation.

b) Fitted logistic equation of growth to the converted biomass (in gDW). The red line corresponds to the fitted equation.



Supplementary Figure 5.2: Parameters analysis for RED

a) Parameters S and F boxplot of the parameter bootstrap resampling
b) The residuals and standardised residuals are plotted against the fitted values to see if a linear are non-linear equation is more adapted to the dataset.
c) The autocorrelation is analysed by plotting residuals at a time (i) against residuals at the next time point (i+1). The normal Q-Q plot of standardised residuals is plotted to verify if some data points are misfits in the equation fitting.
d) Jackknife resampling to further estimate the two parameters distribution showing the most probable couples.



Supplementary Figure 5.3: Outputs for the analysis of actinorhodin by the tool

a) Fitting of the equation to the data with noisy simulated data around a sigma=0.1

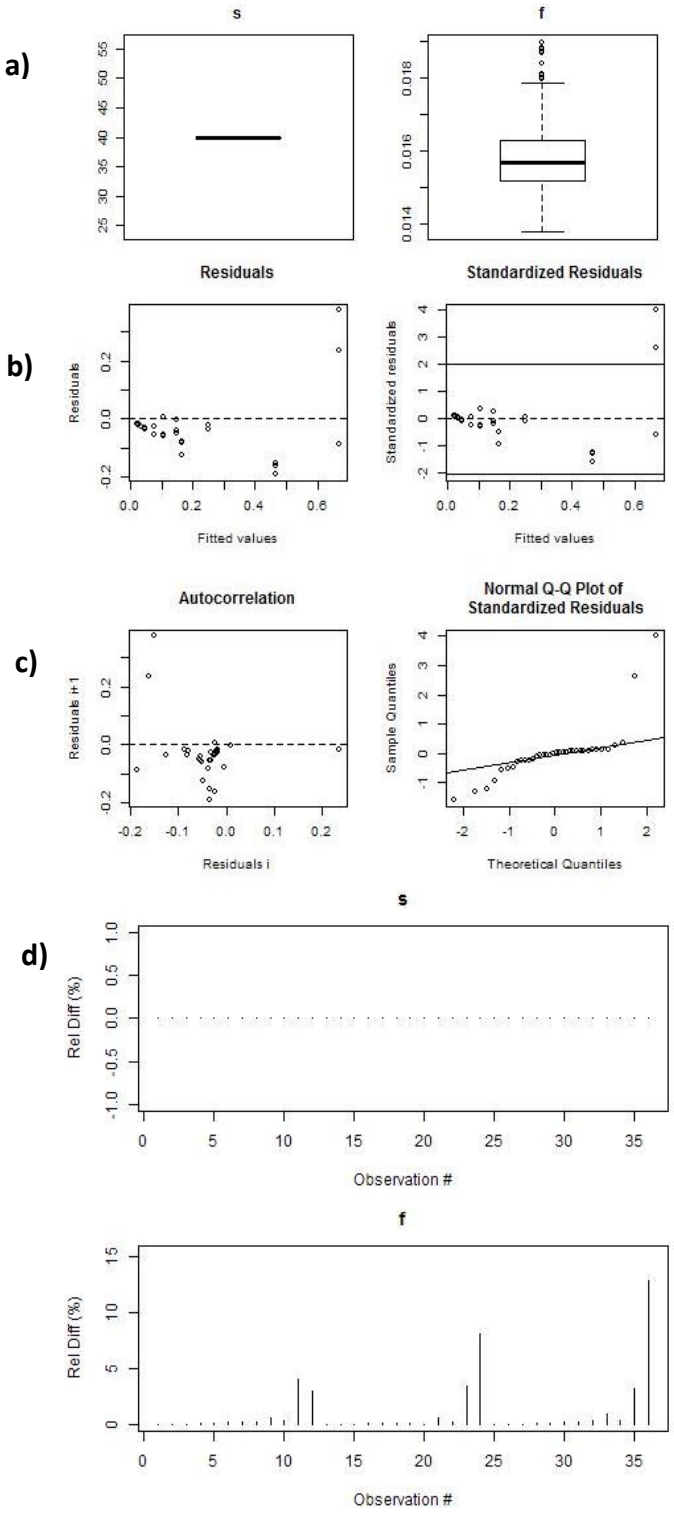
b) Fitting of the equation to the data with the 95% CI band

c) Couple parameters Beale's 95% unlinearized confidence region

d) Couple parameters RSS contour with 95% CI. Heatmap colours corresponds to the couple of parameters from the least fitting in red to the most fitting in dark blue.

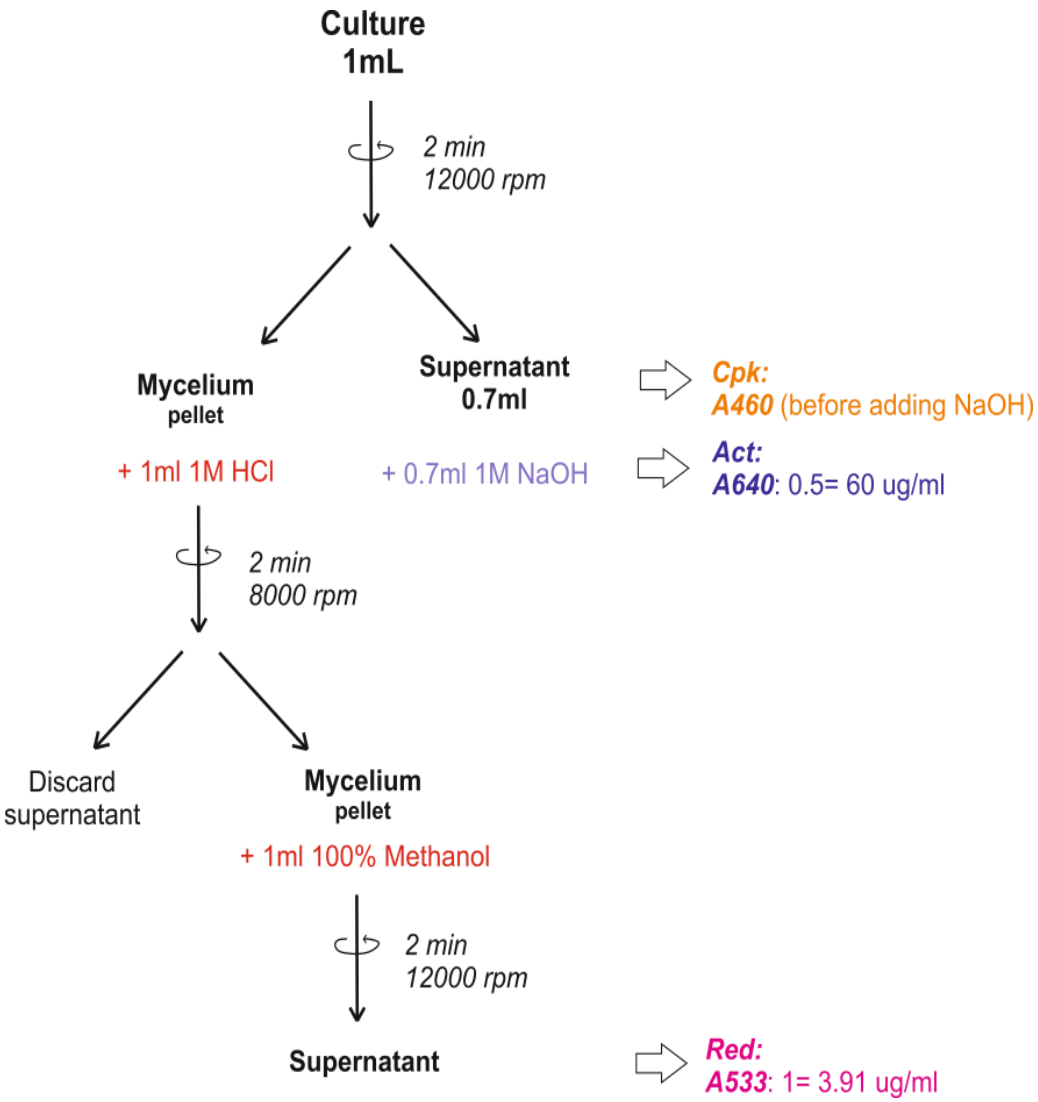
e) Fitted equation with upper and lower values of the 95% CI for the predicted concentration

f) Absolute undecylprodigiosin export flux over time



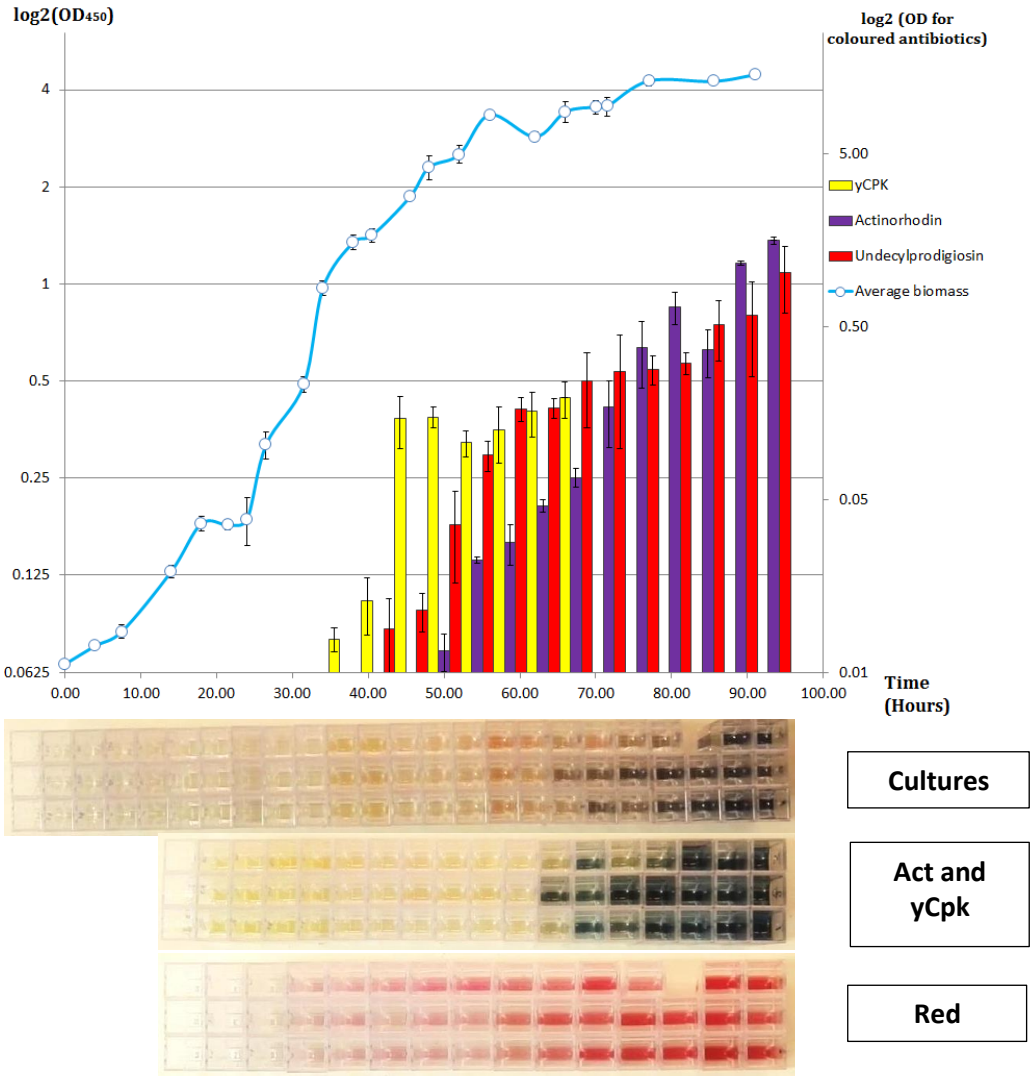
Supplementary Figure 5.4: Parameters analysis for ACT

a) Parameters S and F boxplot of the parameter bootstrap resampling
b) The residuals and standardised residuals are plotted against the fitted values to see if a linear are non-linear equation is more adapted to the dataset.
c) The autocorrelation is analysed by plotting residuals at a time (i) against residuals at the next time point (i+1). The normal Q-Q plot of standardised residuals is plotted to verify if some data points are misfits in the equation fitting.
d) Jackknife resampling to further estimate the two parameters distribution showing the most probable couples.



Supplementary Figure 5.5: Protocol for coloured secondary metabolites concentration measures

Experimental protocol used to measure the coloured antibiotics from the *Streptomyces coelicolor* flasks culture. yCpk (yellow pigment) was measured by absorption at 460nm in the supernatant after spinning down 1mL of culture (2 min at 12000rpm). The Act (blue pigment) concentration was determined by measuring the absorption at 640nm after adding the same volume of 1M NaOH to the same volume of supernatant (0.7ml), an OD_{640} of 0.5 is equivalent to 60 µg/mL of Act [25]. In order to measure the concentration of RED, the mycelium pellet was re-suspended in 1 mL of HCl spun down (2 min at 8000 rpm), the supernatant was discarded and the mycelium in the pellet re-suspended in 1 mL of 100% methanol and spun down (2 min at 12000 rpm), the absorption in the supernatant was measured at 533 nm, an OD_{533} of 1 is equivalent to a concentration of 3.91 µg/mL of RED [25].



Supplementary Figure 5.6: Growth curve of *S. coelicolor* M145 and coloured secondary metabolites concentration across time

Growth curve (in blue) in the minimal chemically defined media for metabolomics, the concentration of the coloured secondary metabolites Act, Red and yCpk are respectively displayed in purple, red and yellow.

Chapter VI

6. Metabolomics and metabolic model-driven *Streptomyces* strains design for antibiotics production

Adam Amara[1], Katsuaki Nitta [2], Sastia Putri [2], Eiichiro Fukusaki [2], Francesco Del Carratore [1], TOPCAPI consortium, Rainer Breitling [1], Eriko Takano [1] (*to be submitted*)

[1] *Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, School of Chemistry, Faculty of Science and Engineering, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK*
[2] *Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan*

6.1 Preface

In the first part of this chapter, about actinorhodin production in *Streptomyces coelicolor*, all the experimental data (metabolomics and growth data), and the processed metabolomics data was generated by Katsuaki Nitta in Prof. Fukusaki Lab (Osaka University). AA did the *in silico* analysis, such as metabolomics data integration, the metabolic modelling, and the gene target identification.

The second part of this chapter, focused on metabolic modelling of GE2270A production, this is part of the Thoroughly Optimised Production Chassis for Advanced Pharmaceutical Ingredients (TOPCAPI) project. AA carried out the metabolic modelling and target identification for GE2270A production. AA and Francesco Del Carratore jointly introduced and verified the GE2270A pathway in the model, ran the knockouts algorithm and the OptGene algorithm, and interpreted the OptGene output. The experimental data (CMAN media determination) was generated by the TOPCAPI

consortium (http://topcapi.eu/). This is aimed to be submitted for publication at a later date.

6.2 Abstract

The decreasing rate of new antibiotics discovery is failing to meet the need of new antibiotics to face the rising antibiotics resistance; this is partly due to economic barriers making antibiotic discovery and production unattractive to industry. Thus, the production of antibiotics at industrially relevant levels is essential to enable cost-effective production, and accelerate novel antibiotics production. The Actinobacteria phylum is a rich source of antibiotics and other secondary metabolites of clinical or industrial interest, and among this phylum the *Streptomyces* species are the most prominent group of antibiotic producing bacteria. Most antibiotics are produced in improved native producing strains or by heterologous production in a host organism. Application of omics' data analysis and metabolic modelling has helped to identify engineering strategies to improve production in both native and heterologous hosts. In order to better understand and increase the production of the actinorhodin antibiotic in the native host *Streptomyces coelicolor*, we studied *in vivo* and *in silico* the production of this antibiotic in four different mutants of *S. coelicolor*. We also studied, by metabolic modelling, the heterologous production of a thiopeptide antibiotic in clinical trial, GE2270A, in the host strain *S. coelicolor* M1146. This enabled identification of two promising gene knockout targets to favour metabolic production of GE2270A in the host strain.

6.3 Introduction

The Actinobacteria phylum is a prolific producer of antibiotics with two thirds of antibiotics on the market derived from the *Streptomyces* genus [1,2]. Many of the antibiotic producing strains are not producing high enough titres of compounds to enable cost-efficient production, or production at sufficient levels to be assayed (e.g., for clinical testing) [3,4]. To produce more antibiotics and other secondary metabolites, one can improve production in the native strain, or improve heterologous production of the compound in a chassis strain [5–7]. The main advantages of improving production in a native strain is that the metabolism is already wired to produce this compound (metabolic precursors are produced); so one can increase production of the compound by directly genetically manipulating the strain [7,8]. However, if the organism is genetically intractable, or difficult to cultivate under laboratory conditions then it is complicated to increase production because it becomes impossible to engineer the organism [6,7]. Here, heterologous production is a good solution to face this problem, by expressing the BGC into a chassis strain that is in general a well-studied organism, such as *Streptomyces coelicolor* for the Actinobacteria phylum. Heterologous production used to have major practical challenges, such as problems with genetic manipulation of the expression host or with the difficulty in cloning large sized BGCs. These barriers are partially overcome by the implementation of gene editing tools (e.g., CRISPR, or TALENs) [9], low-cost gene synthesis [10], and expression vectors [11] added to the synthetic biology toolbox [6,12,13]. This is particularly important for secondary metabolite production, where *Streptomyces* strains (e.g., *S. coelicolor*, or *Streptomyces albus*) are better hosts to produce BGCs discovered in Actinobacteria strains compared to other model organisms from different phylum (e.g., *Escherichia coli*, or *Saccharomyces*

cerevisiae) [13]. The *Streptomyces* strains are more likely to have the necessary metabolic precursors for the secondary metabolites pathway, and to have the machinery to express the BGCs without issues with the Actinobacteria gene sequences (e.g., avoiding issues with GC content) [6,7].

Synthetic biology has the potential to accelerate the engineering of these organisms for antibiotics production and other secondary metabolites [3,12]. The engineering for the production of high-value chemicals (e.g., antibiotics) in microorganisms (i.e., *Streptomyces*) using synthetic biology is driven by the design-build-test-learn (DBTL) cycle (see *Chapter I* for more details) [14,15]. There are two main approaches used to produce a known secondary metabolite: the production in a native producer strain or the heterologous production in a host strain [7]. The application of the synthetic biology DBTL cycle on a native strain starts by analysing the production in this organism (test phase), integrating the data (e.g., omics data) into a genome-scale metabolic model to analyse the organism metabolism (learn phase), then predict potential targets (e.g., gene knockouts and overexpression) to increase production (design phase), and finally execute the design (build phase) [6,14,16]. Then, repeat the cycle to iteratively improve the production. A similar method is applied for using synthetic biology approaches for heterologous production, starting with the selection of a chassis strain and the refactoring of the heterologous pathway (design phase), followed by the expression of the construct with the metabolic pathway in the chassis strain (build phase), the strain is tested (e.g., concentration, and omics data) to identify any bottleneck in the production (test phase), the data is then integrated in a metabolic model to analyse the organism metabolism (learn phase), to finally create new designs (e.g., genes knockouts) [5–7].

The native and heterologous production of antibiotics studied here are in the *S. coelicolor* strain as this strain has the highest quality genome-scale metabolic model available for the *Streptomyces* species [17,18] which was published as part of this thesis. Here, we studied the native and heterologous production of actinorhodin in *S. coelicolor* M145 and M1146 strains, as well as the heterologous production of GE2270A in the M1146 strain (see *Chapter I* for more details on M1146 and M145 strains). Actinorhodin (ACT) is a type II polyketide (PK), and a pH dependent blue-pigmented antibiotic [19,20]. This compound is produced in *S. coelicolor*, naturally, and its production is triggered by the depletion of nutrients such as phosphate or nitrogen [21,22]. GE2270A is a ribosomally synthesized and post-translationally modified peptide (RiPP), it is a thiopeptide antibiotics with derivatives of clinical interest [23,24]. GE2270A is natively produced by *Planobispora rosea*, but this strain is genetically intractable; so, to further study the compound biosynthesis, and to improve production of this compound, it was expressed in the well-studied *S. coelicolor* chassis strain M1146 [25]. *S. coelicolor* was also chosen as host because it is the closest model organism to the Actinobacteria species *P. rosea* [25].

In the first part, we present the study and design of the production of ACT in the WT and mutants *S. coelicolor* strains. The production of ACT was studied in four different strains, the WT M145, M145 with an ACT cosmid introduced, M1146, and M1146 with an ACT cosmid introduced. Metabolomics data and nutrient exchange were used to validate the *i*AA1259 [17] metabolic model for the four strains. The exometabolomics data was used to study the metabolic trade-offs of metabolic exports on growth and on ACT production. The exometabolomics data was also integrated in the metabolic model to predict the intracellular metabolic pathways

active in the different strains. The reactions with predicted metabolic fluxes correlated to high levels of ACT production were also identified. Finally, the metabolomics integration and the modelling were interpreted to identify potential gene knockouts or overexpression targets to redirect metabolic fluxes toward ACT biosynthetic pathway.

In the second part, we present the heterologous production of the RiPP, GE2270A, in *S. coelicolor* M1146. The GE2270A pathway was introduced in the *i*AA1259 model and tested *in silico* in a complex industrial media which was used to obtain the measurements. The antibiotic production trade-off with biomass in this media was analysed, followed by an analysis of the predicted metabolic exchange within the complex media. Then, a simulation was conducted to predict the single and double gene deletions which are non-essential to growth or to GE2270A production. Then, the genes predicted as non-essential were used as an input for the OptGene algorithm. OptGene is an evolutionary algorithm that was used to simulate multiple generations of *in silico* mutants to identify gene deletions that can give an advantage to GE2270A production. This led to the identification of 5 main possible gene deletions that were tested *in silico* by comparing the predicted metabolic fluxes from simulated mutant strains with the different possible combinations of knockouts. Following this analysis, two promising gene knockouts were selected (both associated to FMN reductases); these gene deletion were predicted to couple GE2270A production to biomass, hence forcing production of the compound.

6.4 Methods & Material

Part I: Production of Actinorhodin in *S. coelicolor*

6.4.1 Experimental growth conditions and metabolomics data acquisition

The laboratory experiments described here were carried out by Katsuaki Nitta at Osaka University. Multiple strains were tested, M145, M145+ACT, M1146 [26], M1146+ACT, a cosmid containing the actinorhodin BGC (+ACT) was expressed in the M145 and M1146 strains [26]. The strains were cultivated in a minimal fully-defined glutamate based media [22]. The medium consists of Na-Glutamate, 46.4 g/L; D-glucose, 40 g/L; $MgSO_4$, 2.0 mM; phosphate mix (NaH_2PO_4 and K_2HPO_4), 4.6 mM; supplemented minimal medium trace element solution SMM-TE (0.1 g/L of each of $ZnSO_4.7H_2O$, $FeSO_4.7H_2O$, $MnCl_2.4H_2O$, $CaCl_2$ and NaCl), 8 mL/L; Trace Minimal Supplement 1 (TMS1: $FeSO_4.7\ H_2O$, 5 g/L; $CuSO_4.5$ H2O, 390 mg/L; $ZnSO_4.7\ H_2O$, 440 mg/L; $MnSO_4.H_2O$, 150 mg/L; $Na_2MoO_4.2\ H_2O$, 10 mg/L; $CoCl_2.6\ H_2O$, 20 mg/L, and HCl, 50 mL/L), 5.6 mL/L; Antifoam, 1 ml/L. The pH was adjusted to pH 7 by adding NaOH 2M to the media. The cultures were grown in an incubator at 30°C and shaking at 250 rpm. Cultivations were performed in 250 mL baffled conical flasks containing springs, with a culture volume of 50 ml; all the flasks have been coated to reduce aggregation of the cells on the side of the flask. 10^9 of viable spores were inoculated per flask. The growth was estimated by OD_{450} measurements (3 measures per time point). In order to calculate growth (DCW/mL) of each strain for growth curve, the cultures were collected and separated to supernatant and cell pellet by centrifugation (10,000 rpm, 5 min, 4°C). The cell pellet were washed by twice amount of milliQ water and centrifuged (10,000 rpm, 5 min, 4°C) to obtain medium-free cell pellet and the cell pellet were frozen by liquid N_2 and followed by freeze drying. The supernatant was used for following secondary metabolites analysis and other primary extracellular metabolites. The concentrations of coloured secondary metabolites were spectrophotometrically measured by

following a series of extraction and measurements on a culture aliquot of 1 mL for each time-point. The amount of yCpk (yellow pigment) produced was estimated by measuring the absorption at 460nm in the supernatant after spinning down 1mL of culture (2 min at 12,000rpm). The Act (blue pigment) concentration was determined by measuring the absorption at 640 nm after adding the same volume of 1M NaOH to the same volume of supernatant (0.7 ml), an OD640 of 0.5 is equivalent to 60 µg/mL of Act [19]. In order to measure the concentration of RED, the mycelium pellet was re-suspended in 1 mL of HCl spun down (2 min at 8,000 rpm), the supernatant was discarded and the mycelium in the pellet re-suspended in 1 mL of 100% methanol and spun down (2 min at 12,000 rpm), the absorption in the supernatant was measured at 533 nm, an OD533 of 1 is equivalent to a concentration of 3.91 µg/mL of RED [19]. The glucose, glutamate, and phosphate levels in the media were measured using a F-kit-D-Glucose (J.K.International), a L-glutamate assay kit Yamasa NEO (Yamasa), and a PiBlue Phosphate Assay kit POPB-500 (Bio Assay Systems).

Cells for intracellular metabolome analysis were collected by fast-filtration with 5.0 µm pore size nylon membrane filter. The cells were washed with twice the amount of $NaHCO_3$ to eliminate medium component and the filter with cells are transferred to 15 mL falcon tube and the metabolism of cells was quenched by liquid N_2 immediately. The cells were kept at -80°C until extraction.

In order to extract intracellular metabolites, 4 mL of mixed-solvent solution (MeOH: $CHCl_3$: Water = 5:2:2, v/v/v) including 20 µg/L (+)-10 Camphorsulfonic acid and 20 µg/L ribitol as internal standards was added to quenched samples. Three cycle of freeze and thaw cycle as follows were conducted to improve metabolites extraction efficiency: (1) freezing at -80°C , (2) thawing at -30°C, (3) vortex for 5 seconds and

sonicating for 10 seconds. The extracts were centrifuged (10,000 rpm, 5 min, 4°C) and the supernatant was transferred to new tube. 1.0 mL ultrapure water was added to the 3 mL supernatant and vortexed well. The mixture was centrifuged (10,000 rpm, 5 min, 4°C) and the upper layer (polar phase) was filtered by 0.2 μm PTFE filter. 2.0 mL of the upper layer was used for LC-MS/MS analysis while 1.0 mL of the upper layer was used for GC-MS analysis. The upper layer samples were concentrated by spin dryer (at 4°C for 2 hours) and freeze dried for later analysis. Here pooled sample was obtained from all the samples and prepared as QC samples to check metabolome analysis stability during the analysis, especially untargeted GC-MS analysis. The lyophilized sample was kept at -80°C until analysis. On the other hand, for extracellular metabolites analysis, 50 μL of supernatant of cell culture was added to 1.8 mL mixed-solvent solution (MeOH: $CHCl_3$: Water = 5: 2: 2, v/v/v) and incubated at -30°C for 30 minutes. The extracts were centrifuged (10,000 rpm, 5 min) and 400 μL of ultrapure water was added to the supernatant. The mixture was vortexed well and centrifuged for two layers separation. The upper layer was filtered and 700 μL was transferred to new tube for concentrating by spin dryer and freeze drying. The freeze dried sample was kept at -80°C until analysis. Freeze dried intracellular and extracellular samples were reconstituted with 100 μL ultrapure water for LC-MS/MS analysis. The solution was centrifuged (10,000 rpm, 5 min) and 40 μL supernatant was transferred to LC vial. Intracellular and extracellular metabolites were analysed by two kinds of LC-MS/MS analysis platforms. One LC-MS/MS analysis platform is ion-pair LC-MS/MS with LCMS8030 Plus and L-column2 ODS (150 nm × 2.1 mm, particle size 3 μm). The mobile phase A is 10 mM tributylamine with 15 mM acetate in ultrapure water while the mobile phase B is MeOH. The analysis was performed with gradient mode, metabolites

were ionized by negative ion mode and the injection volume was 3 µL. The data file from LC-MS/MS was converted to abf format file by Abf file converter and analysed by MRM probes [27]. Another LC-MS/MS analysis platform is LC-MS/MS with LCMS8050 and Discovery HS F5-3 (150 nm × 2.1 mm, particle size 3 µm). The mobile phase A is 0.1% formic acid in ultrapure water while the mobile phase B is ACN. The analysis was performed with gradient mode, metabolites were ionized by positive ion mode and the injection volume is 3 µL. The data file from LC-MS/MS was analysed by Lab solution (SHIMADZU). MRM parameters are optimized from authentic standards in advance and all MS/MS analysis was performed by MRM mode. Each metabolite's annotation was performed by the MRM parameters and RT of authentic standards. On the other hand, samples for GC-MS untargeted intracellular metabolites analysis was derivatized by 100 µL of 20 mg/ml methoxyamine hydrochloride in infinity pure pyridine at 30°C, 1,200 rpm for 90 minutes and by 100 µL of MSTFA solution at 37°C, 1,200 rpm for 30 minutes. The derivatized sample was centrifuged (10,000 rpm, 5 min) and 40 µL of the supernatant was transferred to GC-vial. The analysis was performed by GC-ultra (single Q-MS) with InertCap 5MS-NP column (30 m × 0.25 mm I.D., df. = 0.25 µm). The carrier gas is He and the linear velocity was set to 39 cm/second. The injection volume was 1 µL and the split ratio was 1:25. The analysis was performed by scan mode with mass range from 85 to 500 and column oven was changed by gradient mode. The data file from GC-MS analysis was converted to abf file by Abf file converter and the Abf file was analysed by MS-DIAL [28]. Metabolites were annotated by similarity comparison of RI and mass fragmentation pattern between real data and online library. In whole analysis, extracellular data was normalized by IS while intracellular data was normalized by IS and cell amount.

6.4.2 Metabolic model experimental constraints and validation

The nutrient and antibiotic levels in the media for the 4 different *S. coelicolor* strains were converted into exchange rates using the *R* tool described in the *Chapter V*. The metabolite import tool version was used for glucose, glutamate, and phosphate, while the metabolite export tool version was used for ACT and RED. The tool was applied to ACT data only for M145, M145+ACT, and M1146+ACT, as only these strains were producing ACT. The tool was applied to RED data only for M145, and M145+ACT as only these strains were producing RED. However, the tool could not fit an equation to the RED data in M145+ACT due to noisy data and the overall trend with high fluctuations between time points (levels going up and down and up again); so the RED export rates were estimated manually at each time point. The uptake and export rates calculated for the different strains were used to constrain the *i*AA1259 model [17]; then the biomass was optimised using flux balance analysis (FBA) [27] in the COnstraint-Based Reconstruction and Analysis in Python (COBRApy) framework [28]. The growth rates (gDW/h) predicted was converted into accumulated biomass (in gram of dry weight) which was compared to the measured biomass for each of the four strains. An ensemble modelling approach was used to predict the biomass by sampling exchange fluxes to constrain the model (method described in more details in *Chapter V*). Exchange constraints were sampled 500 times between the maximum and minimum value of the uptakes and export rates. The sampling method used divided the data range in 500 between the maximum and minimum flux calculated for each time point. The exchange fluxes were combined in sets of constraints in the order of maximum to minimum. This considers that if the strain reaches the max/min uptake of a given metabolite it will likely reach the max/min uptake or export for the other metabolites,

as well as reaching the max/min at the next time point. These 500 sets of exchange fluxes were constrained on the model using COBRApy in a loop running for each time point and optimising the growth using FBA. Finally, the 500 predicted growth curves were visualised with the average predicted growth and the experimental data. All the program scripts, data inputs, data outputs, and visualisations from the tool are available in the *Electronic Supplementary 6.1*.

6.4.3 Exometabolomic data integration in the metabolic model

The ratio of metabolites for M145+ACT to M145 (and inversely) as well as M1146+ACT to M1145 (and inversely) were calculated from the normalised metabolomics data (based on ion intensity) to identify if metabolites were higher or lower in the different strains. The metabolomics data was processed and normalised by Katsuaki Nitta. Then, I used this dataset to visualise it on a heatmap and cluster it using Ward's hierarchical clustering from *R* [29]. The metabolites identified in the extracellular milieu of the different strains were checked in the optimised metabolic models to find out if the model predicted their export, but none of the exometabolites were predicted as exported by the models. Some of the metabolites to export did not have a transporter in the model nor an extracellular metabolite annotation. So the transporters and the extracellular metabolites were added in the models for citrulline, hypoxanthine, S-adenosyl-methionine, S-adenosyl-homocysteine, and urate.

All the metabolites identified in the media were used to simulate the metabolites export in the models for the different strains to determine the trade-off cost of exporting these metabolites for the organism. The metabolic models were constrained with the metabolic exchange fluxes of glucose, glutamate, phosphate, ACT, and RED (as described in the previous section) at exponential growth phase (30h)

and stationary phase (64h). Then, the metabolites export were set as objective one by one for the models and optimized using FBA [30] in COBRApy [28]; the optimum value was saved as the maximum possible export. The export fluxes were then individually constrained between 0 and the maximum possible export flux (using a range of 50 points in between), at each export flux value constrained (within the range) the biomass or the ACT production was optimized by FBA in COBRApy [28]. The constrained export fluxes values were plotted in the X-axis and the optimized biomass or ACT production rates in the Y-axis, to visualise the trade-off between the export and the optimal biomass or ACT production predicted. This trade-off curve represents the predicted cost of exporting these metabolites for the metabolic network under these conditions.

The metabolic models were first constrained using the media conditions and the exchange fluxes (as described in the previous section). The exometabolites were then used to integrate new constraints to the metabolic models by forcing export of these metabolites. However, as these metabolomics datasets are semi-quantitative, the metabolite export rates were constrained to simply force exports at the maximum level at which the constrained models could still optimise growth. The combination of the metabolite export rates cannot be forced to levels so high that the model cannot optimise growth at the same time. So, the maximum possible exports rates for all the metabolites at the same time were determined to be between 0.1 and 0.0001 mmol/gDW/h depending on the metabolite. The aim is to simply simulate the metabolic pathways activated by these exports even if the fluxes are not accurately predicted. Once the metabolite exports were constrained, the growth was optimised in the four metabolic models, at the 6 time points. Then, a Pearson correlation was

calculated between the predicted intracellular fluxes and ACT production across time, and the reactions were ranked from the most to the least correlated in each strain.

6.4.4 Identification of the metabolic differences due to metabolites secretion and actinorhodin production

The initial metabolite analysis was conducted by Katsuaki Nitta (University of Osaka). It focused on the isolation of the metabolic differences between M1146 and M1146+ACT. M1146 has a cleaner metabolic background [31] in comparison with M145 as the native ACT, CDA , CPK, and RED biosynthetic genes clusters are deleted.

Further analysis using the model was conducted by AA, the metabolic models for M1146 and M1146+ACT were constrained at the 6 time points with the corresponding media uptakes. Next, the metabolites exported were individually defined as the objective function and the models were optimised to predict optimal export and the intracellular metabolic fluxes under these conditions. The predicted fluxes for each optimised metabolites exports are then compared to the fluxes when growing only (M1146) or when growing and producing ACT (M1146+ACT). These flux comparisons at each time point were then compiled into a single matrix where all the reactions carrying fluxes when exporting a given metabolite are kept. In this analysis all the predicted fluxes inferior to 10^{-8} or inferior to -10^{-8} mmol/gDW/h were considered as null. So, this final matrix represents all the metabolic reactions predicted as active due to the metabolic exports only, as the metabolic fluxes associated to other functions such as growth or ACT production were filtered out. Here the optimisation was done using parsimonious FBA (pFBA), which improves intracellular fluxes distribution predictions but obtain the same optimal value for the objective compared to FBA [32,33] .

The metabolic model is constrained with the minimal media open constraints, where the uptakes and exports rates for glucose, glutamate, and phosphate were open to the maximum and minimum possible. The optimal ACT production value is predicted by optimising the model with ACT production as objective. Then, the objective function is set on biomass, while the ACT production is progressively constrained from 0 to the optimal value at 7 different points (0, 5, 10, 30, 50, 90, and 100% of the optimal value). Finally, the Pearson correlation is calculated between the predicted fluxes compared and the ACT production to identify the reactions increasing or decreasing with ACT production. Here the optimisation was done using pFBA [32] .

The metabolic reactions predicted to be correlated to ACT production, the reactions predicted to be activated by the metabolic exports, and the reactions carrying fluxes but not associated to the two previous groups are mapped on an Escher map [34] of central metabolism. The central metabolism map was built from the *E. coli* metabolic model [35], by adapting it to the *S. coelicolor* central metabolism from *i*AA1259 [17] using the Python version of Escher [34]. This was based on the BiGG identification for both models as the *i*AA1259 model is compliant with the BiGG standards [17,36].

<u>Part II</u>: Heterologous production of GE2270A in *S. coelicolor* chassis strain

6.4.5 GE2270A pathway introduction, complex media constraints, and production modelling

The GE2270A biosynthetic pathway [23,25] was introduced in the *S. coelicolor i*AA1259 metabolic model [17]. The pathway was divided in two reactions, the first corresponding to the production of the precursor peptide by the ribosome from the aminoacyl-tRNAs, and the second corresponding to the post-translational modifications, and the

cleavage of the leader peptides leading to the final GE2270A compound. The compound production was then entirely tested *in silico* as no experimental data was available. The growth was simulated in the complex Manchester (CMAN) media, which was developed by the Manchester team of the TOPCAPI consortium. The CMAN media is a complex media based on glucose (10g/L), soluble potato starch (5g/L), hydrolysed casein (5g/L), Yeast extract (8g/L), and calcium carbonate (2g/L). As the exact concentrations of most of the metabolites in the media were unknown, the metabolites from the different components of the media were openly constrained (uptake of -1000 mmol/gDW/h); unless the tryptophan and cysteine which are present in very low concentrations in yeast extract and in casein. Also, glucose was constrained to a maximum possible import of 55 mmol/gDW/h (as its concentration could be calculated). The CMAN media composition is available in the *Supplementary Table 6.1.*

The GE2270A production was set as objective function and optimized using FBA [30] in COBRApy [28]; the optimum value was used as the maximum possible production. Then, GE2270A production was constrained between 0 and the optimum GE2270A production rate at each point the biomass was optimised by FBA in COBRApy [28]. The optimum growth against the GE2270A constraint helped to generate a trade-off curve to visualise the cost of an increasing production on the growth.

The metabolic exchanges within the CMAN media were studied with growth as an objective, with GE2270A production as an objective, then with growth as an objective and GE2270A production constrained to 98% of the optimal growth. The metabolites consumed and exported were predicted by FBA [30] analysis under the three different conditions(growth, GE2270A production, and growth and GE2270A production at the same time), the metabolites only exchanged under a particular

condition were were highlighted. The metabolic exports and imports increasing or decreasing between two conditions (GE2270A production, and growth) were normalised to the optimum growth to identify if the flux increase or decrease is associated to biomass, and the same with GE2270A optimum.

6.4.6 Gene knockout targets for GE2270A production

The metabolic model was used to identify single gene knockouts essential to growth and the ones essential to GE2270A in the CMAN media (developed by the TOPCAPI consortium). The single knockouts were simulated using the knockout tool in the COBRApy [28] toolbox; a gene is essential if its deletion leads to an impossible optimisation of the objective function (growth or GE2270A production) by FBA [27]. The optimum value obtained for the gene knockouts for growth and for GE2270A production were matched based on the genes locus ID, and plotted to identify the genes that lead to a decrease in growth and/or to GE2270A production. The exact same method was applied for double knockouts, as double knockouts are more likely to impact significantly the metabolism by cutting multiple reactions at once [37].

The metabolic model constrained in CMAN medium was considered by the evolutionary algorithm OptGene [38] to identify potential gene knockouts to improve GE2270A production. Glucose was used as the main carbon source in the algorithm, while the minimum growth allowed was 0.1 gDW/h, and the objective to optimise was GE2270A production. The space of possible genes knockouts for the algorithm to consider was reduced to the non-essential genes for biomass and GE2270A production. The OptGene algorithm considers GE2270A as the fitness trait (objective) with a fitness score (objective optimum). 1) The algorithm generates a population of 500 individuals, by introducing random mutations to the initial model (the rate of

mutations is adjusted to a minimum of 1 mutation and maximum of 10). 2) OptGene performs an FBA analysis on each of the models maximising growth. 3) All the individuals are ranked according to their fitness score (GE2270A production). 4) The best "individuals" are selected for mating based on their fitness score, and subsequently crossed to produce a new "generation" of 500 individuals. 5) OptGene introduces new random mutations in the populations (based on the rate of mutations of 1 to 10). 6) The algorithm goes back to step 2 in a loop. The algorithm stops when the best individual of a "generation" satisfy a specific criteria, or when it reaches the maximum number of generation set (10,000 here) [38]. The OptGene outputs are then visualised, such as the fitness value across generations and the mutations histogram showing the number of knockouts per individuals in the final selected population. The final generation of 500 *in silico* mutants was studied in more detail by comparing the genes most frequently mutated to identify a list of the genes most frequently deleted by the algorithm. This revealed 6 genes that were very frequently knocked out (>70% frequency) in the final population, but 5 gene deletions out of 6 led to a reaction knockout. The 5 genes were knocked out in the model in all possible combinations, leading to 30 *in silico* mutant strains; the genes were also knocked out *in silico* using the COBRApy gene knockout function [28]. The 30 mutant strain models and a WT model (non-producer) were optimised for growth using pFBA [33]. The predicted metabolic fluxes were then compared between all the strains, and divided into groups based on the sum of all fluxes to identify the mutants that have the same metabolic phenotype as the WT, and the mutants that were producing GE2270A without constraining it. The predicted fluxes higher or lower in GE2270A producers compared to non-producers

were analysed to identify the metabolic fluxes predicted as increasing or decreasing in the GE2270A producing strains.

6.5 Results & Discussion

Part I: Production of Actinorhodin in *S. coelicolor*

The *S. coelicolor* strain is a native producer of the blue-coloured type II PKS antibiotic actinorhodin (ACT) [19,20]. The study of ACT production in *S. coelicolor* is used as a model for polyketides antibiotic production in *Streptomyces* species [19,21,39]. Here, the ACT production was studied in the wild-type (WT) reference strain M145 and in the *S. coelicolor* chassis strain M1146.

6.5.1 Model constraints and validations

The production of ACT was studied under four different conditions, in the *S. coelicolor* WT strain (M145), in the WT with a cosmid containing the ACT BGC (M145+ACT), in a *S. coelicolor* strain with the four active antibiotics (ACT, CDA, CPK, RED) BGCs deleted (M1146), and in the M1146 strain containing the ACT cosmid (M1146+ACT) (*Table 6.1*). This cosmid integrates a single copy of the entire ACT BGC into the host strain genome [26]. So, the M145 strain contains one native copy of ACT cluster, the M145+ACT contains two copies of the ACT cluster, the M1146 contains no ACT cluster, and M1146+ACT contain one introduced ACT cluster (in a different location compared to M145). Collaborators from the Osaka University acquired time-point data (20, 30, 40, 50, 64, and 72 hours) in the 4 different strains, for biomass, glutamate, glucose, phosphate, RED, and ACT concentrations in a minimum glutamate medium [22,40]. They also acquired targeted intracellular and extracellular metabolomics data at all the time points.

Strains	*Streptomyces coelicolor* **M145**	*Streptomyces coelicolor* **M145 + ACT**	*Streptomyces coelicolor* **M1146**	*Streptomyces coelicolor* **M1146 + ACT**
Characteristics	Plasmid-free WT strain *S. coelicolor A3(2)*. Plasmids SCP1$^-$ and SCP2$^-$ curated.	M145 strain containing a cosmid with the actinorhodin BGC	Derived from the M145 strain, with the main BGCs deleted. Δ*act* Δ*red* Δ*cpk* Δ*cda*	M1146 strain containing a cosmid with the actinorhodin BGC

Table 6.1: The four different *Streptomyces coelicolor* strains used in the experiments

The strain was grown in a minimal fully-defined media with phosphate as a limiting nutrient, once the phosphate is depleted, it triggers the entry into the stationary growth phase and secondary metabolites production [40]. The media conditions, nutrients, antibiotics, and growth data measurements were used to apply specific-constraints on the *S. coelicolor* metabolic model *i*AA1259 [17] for the four strains (*Figure 6.1*). The exchange rates of nutrients (glucose, glutamate, and phosphate) and antibiotics (ACT, and RED) of the cultures for the four strains were calculated and automatically constrained using the flux tool described in the *Chapter V* (*Figure 6.1*). The constraints applied generated by the flux tool for the M145 data are shown in *Figure 6.1*, while the data for all four strains are available in the *Supplementary Figures 6.1 to 6.9*. In all four strains, the phosphate consumption decreased following a decreasing logistic curve (*Figure 6.1a*, *Supplementary Figures 6.1g, 6.1h, 6.4g, 6.4h, 6.7g, 6.7h, 6.8g, and 6.8h*). The resulting uptake flux of phosphate has a bell shape peaking during exponential (*Figure 6.1b, Supplementary Figures 6.1i, 6.4i, 6.7i, and 6.8i*), the point after the peak is phosphate depletion leading to stationary phase. The glucose and glutamate levels are decreasing linearly in the media (*Figure 6.1c, 6.1e*, *Supplementary Figures 6.1a, 6.1b, 6.1e, 6.1f, 6.4a, 6.4b, 6.4e, 6.4f, 6.7a, 6.7b, 6.7e, 6.7f, 6.8a, 6.8b, 6.8e, and 6.8f*). The uptake fluxes are proportional to the growth

trends with a maximal uptake during exponential growth and a near zero uptake during stationary phase (*Figure 6.1d, 6.1f*, *Supplementary Figures 6.1c, 6.1g, 6.4c, 6.4g, 6.7c, 6.7g, 6.8c,and 6.8g*). RED is only produced in M145 and M145+ACT as the cluster is deleted from the M1146 strain, RED production exponentially increases after stationary phase (*Figure 6.1g, 6.1h, Supplementary Figures 6.3, and 6.6)*. However, the tool does not reliably analyse RED levels and fluxes in M145+ACT as the data is very noisy and do not follow a linear, exponential, or logistic trend (*Supplementary Figure 6.6*). This is likely due to the introduction of the extra ACT cluster that must perturb RED regulation or production. ACT is only produced in M145, M145+ACT, and M1146+ACT, with a rapidly increasing production after stationary phase (*Figure 6.1i*, *Supplementary Figures 6.2, 6.5, and 6.9*). The uptake and secretion fluxes estimated by the tool (*Figure 6.1b, d, f, and h*) are used to constraint the metabolic models at the 6 time points.

The constrained metabolic models for the four strains were validated by predicting the growth rate at the 6 different time points (20, 30, 40, 50, 64, and 72 hours); the growth predictions were compared to the experimental growth curves (*Figure 6.2*). The predicted growth curves for all four conditions are in good agreement with the experimental data (*Figure 6.2a, b, c, and d*). An ensemble of 500 different constraints sets within the 95% confidence interval of uptake and secretion fluxes was generated by the flux tool (see *Chapter V* for more details), then an ensemble of 500 models at the 6 time points were optimised to predict 500 different growth curves (*Figure 6.2a, b, c, and d*). The ensemble of different growth predictions created a band of possible growth curves similar to a confidence interval of growth predictions (*Figure 6.2a, b, c, and d*).

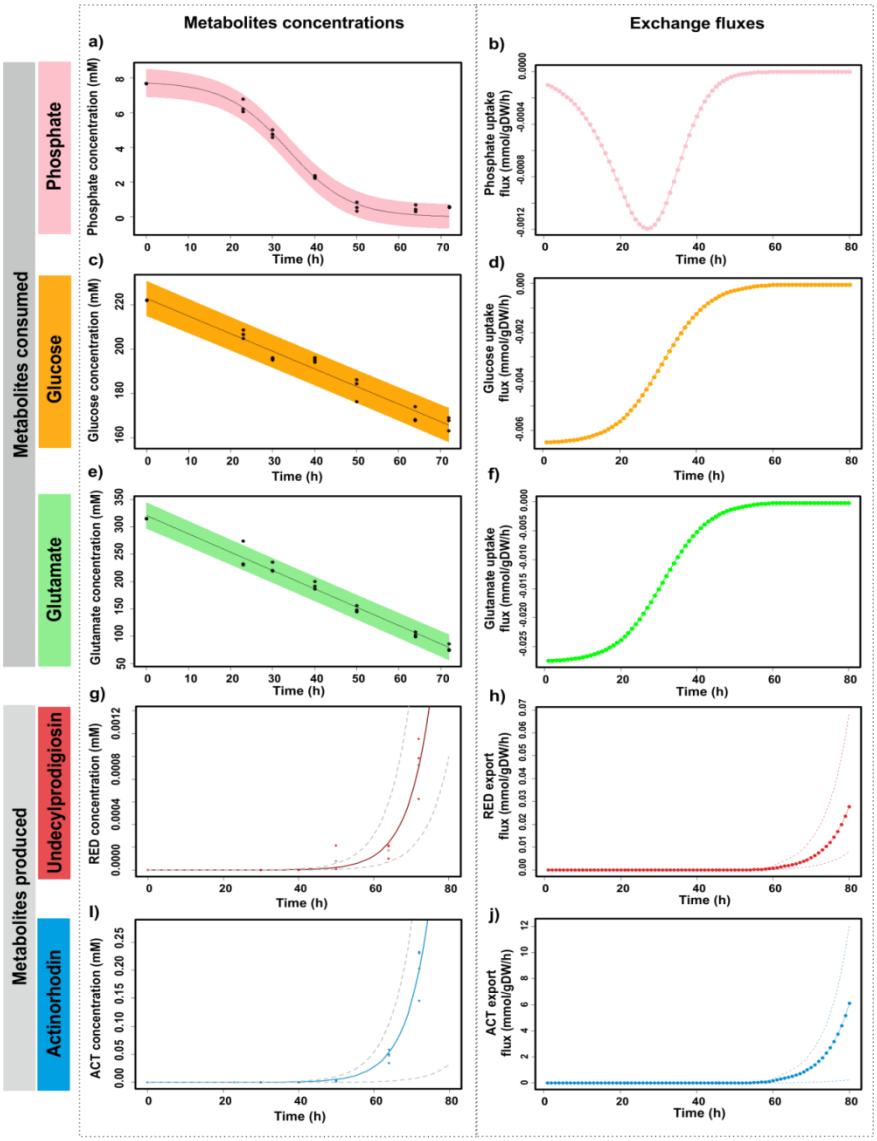


Figure 6.1: Example of metabolites exchange constraints automatically applied on the metabolic models

The data presented here is for the M145 culture, the detailed data for all four strains is available in ***Supplementary File 1*** (***Supplementary Figures 1 to 9***).

a) Fitting of a decreasing logistic equation to the phosphate concentration data with the 95% confidence interval band. The black points correspond to experimental data.
b) Uptake flux values across times for the phosphate
c) Fitting of a linear equation to the glucose concentration data with the 95% confidence interval band. The black points correspond to experimental data.
d) Uptake flux values across times for the glucose
e) Fitting of a linear equation to the glutamate concentration data with the 95% confidence interval band. The black points correspond to experimental data
f) Uptake flux values across times for the glutamate
g) Fitting of a logistic equation to the undecylprodigiosin concentration data with the 95% confidence interval band. The red points correspond to experimental data.
h) Secretion flux values across time for the undecylprodigiosin
i) Fitting of a logistic equation to the actinorhodin concentration data with the 95% confidence interval band. The blue points correspond to experimental data.
j) Secretion flux values across time for the actinorhodin

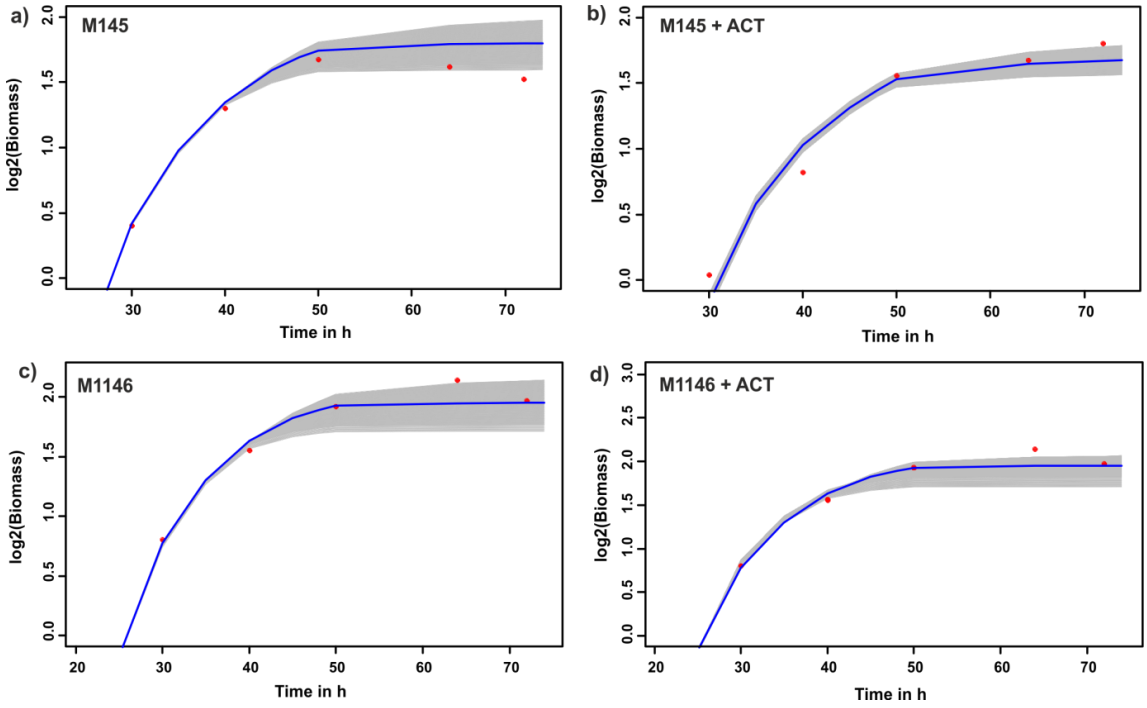


Figure 6.2: Validation of the metabolic model *i*AA1259 growth predictions under the four different conditions

a) Predicted growth by the constrained metabolic model (blue line) and 500 ensemble predictions (grey band) compared to the experimental data (red points) for the *S. coelicolor* M145 strain.

b) Predicted growth by the constrained metabolic model (blue line) and 500 ensemble predictions (grey band) compared to the experimental data (red points) for the *S. coelicolor* M145 strain with the ACT cosmid (M145+ACT).

c) Predicted growth by the constrained metabolic model (blue line) and 500 ensemble predictions (grey band) compared to the experimental data (red points) for the *S. coelicolor* M1146 strain (with no antibiotic gene clusters).

d) Predicted growth by the constrained metabolic model (blue line) and 500 ensemble predictions (grey band) compared to the experimental data (red points) for the *S. coelicolor* M1146 strain with the ACT cosmid (M1146+ACT).

6.5.2 Exometabolome integrative analysis

Microorganisms consume and secrete metabolites in the media during growth, which are a direct result of the intracellular metabolism. The metabolites taken up are consumed to feed the cellular processes, whereas the metabolites secreted have many different functions, such as secondary metabolites, overflow primary metabolism intermediates, waste by-products, or cell lysis [41,42]. So the analysis of extracellular

metabolites can give an insight into the metabolic state of a microorganism [41]. Uptake fluxes are already widely used to constrain the metabolic models used to predict the physiological state of an organism, as described in *Chapter V*. However, secreted metabolites are less commonly used as constraints but it was shown to generate better predictions of intracellular metabolic states when integrated in metabolic models [27,43,44]. Thus, the exometabolome data acquired for the four strains at the 6 time-points was analysed and integrated in the metabolic models already constrained with the nutrients uptake and antibiotics export.

First, the exometabolome was processed and normalised by Katsuaki Nitta, I then analysed this processed data to identify the metabolites with higher or lower levels in the media for the high ACT producers: M145+ACT compared to M145 and M1146 compared to M1146+ACT. However, this dataset is semi-quantitative, so the data was not converted into metabolite concentrations, which would allow calculating the exchange rates to constrain the models. So only the trends were analysed (increasing or decreasing) and the differences of metabolite levels between strains (higher or lower). The metabolites are clustered in groups based on the data trends and on the ratio values; so these groups correspond to metabolites globally higher, lower, or with no difference between the two strains compared (*Supplementary Figure 6.10 and 6.11*). The strain with an extra ACT cluster M145+ACT, secreted multiple metabolites associated with glycolysis and TCA cycle (FBP, DHAP, GAP, PEP, iso-/citrate, pyruvate) at higher levels than M145, a few with purines & pyrimidines metabolism (urate, CDP, UDP, orotate, thymine, cystine, cAMP), pentose phosphate pathway (R5P, Ru5P/Xu5P), and with amino-acid metabolism (glutamine, aspartate, serine, ornithine) (*Supplementary Figure 6.10*). Whereas the M145 strain secreted

higher levels of many metabolites associated with purines & pyrimidines metabolism (uridine, uracil, xanthine, guanosine, guanine, hypoxanthine, inosine, thymidine, UMP, xanthosine, adenine, cytosine, adenosine), and a few metabolites associated to central and amino-acid metabolism (6PGA, UDP-glucose, SAH, glycine) (*Supplementary Figure 6.10*). The strain with no ACT production (M1146) secreted higher levels than M1146+ACT of a few metabolites associated with amino-acids and central metabolism (citrulline, ornithine, glutamine, 4-aminobutyrate, lactate, SAH, and gluconate), with purines & pyrimidines metabolism (cytosine, and urate) (*Supplementary Figure 6.11*). In contrast, M1146+ACT strain excreted higher levels than M1146 of metabolites associated with purines & pyrimidines metabolism (uracil, thymine, xanthine, adenine, orotate, xanthosine, guanosine, adenosine, uridine, thymidine, cytidine, hypoxanthine, guanine), and with central and amino-acids metabolism (SAM, iso-/citrate, pyruvate, 2-oxoglutarate, cAMP, aspartate) (*Supplementary Figure 6.11*). The metabolite cAMP was identified by our collaborators as correlated with ACT production, as the cAMP levels in the media increase at the same time as ACT extracellular concentrations (*Supplementary Figure 6.12*). This metabolite is thought to have an important role in *Streptomyces* germination [45], secondary metabolism regulation [46], and central metabolism regulation [46,47]. Further transcriptomics experiments by our collaborators are currently being carried on genes expression levels to identify the impact of cAMP on ACT production. The cAMP role will be further discussed with the integration of the metabolomics data.

It is also important to make sure that these extracellular metabolites are not the result of cell lysis or damage [41]. For example, if multiple primary metabolites are measured as increasing in the extracellular milieu during low growth rates this is

unlikely to be caused by overflow metabolism, but rather by lysis. One of the possible reason for cell damage is the overproduction of antibiotics compounds such as Actinorhodin that may lead to autolysis but this has not been observed here. The data analysed here did not seem to show major patterns related to cell lysis, such as primary metabolites higher during stationary. However, to make sure there is no significant cell lysis occuring the cultures need to be tested for live and dead cell count. In the metabolic models constrained and optimised, none of the metabolites measured in the media were predicted as exported because only a few growth-coupled waste metabolites secretions contribute to maximise biomass (e.g., CO_2). The metabolic model, when optimised using FBA, only predicts balanced fluxes complying with the constraints or objective optimisation [27,30]. So, constraining these metabolite exports is necessary to predict the metabolism more accurately.

The metabolite exports have a cost for the organism metabolism, and these exports may compete for metabolic resources used for biomass and ACT production. So, studying the impact and the cost of metabolic secretions on the metabolism could help to identify metabolites involved in non-essential metabolic processes, which could be deleted to redirect metabolic resources toward ACT production. All the exports of metabolites (identified in the exometabolome) were individually studied *in silico* for the four strains during exponential phase (at 30h) and during stationary phase (at 64h) (see *Electronic Supplementary – Additional File 2*). The metabolites exported were divided in two groups the ones in direct and the ones in indirect competition with biomass. For example, in all four strains, increasing exports of cAMP lead to a direct decrease of biomass (*Figure 6.3a*). Whereas an increasing export of 2-oxoglutarate did not impact biomass growth until a “breaking” point in high cAMP export rates, where

biomass growth rates dropped to near 0 values (*Figure 6.3b*). The metabolites predicted as in direct competition with growth are different between the M1146 and the M145 derived strains (*Electronic Supplementary 6.1 – Additional File 2*). In M145 and M145+ACT, the metabolites exported predicted as competing with biomass are involved in glycolysis and the pentose phosphate pathway (G6P, DHAP, F6P, FBP, G3P, PEP, R5P, S7P, UDP-glucose, cAMP, PEP), and in the purines and pyrimidines metabolism (CDP, UDP, UMP) (*Electronic Supplementary 6.1 – Additional File 2*). In M1146 and M1146+ACT, the exported metabolites predicted as competing with biomass are involved in amino-acid metabolism (glycine, arginine, asparagine, and methionine), in purine and pyrimidine metabolism (xanthosine, hypoxanthine) and central metabolism (pyruvate, 2-oxoglutarate, lactate, and cAMP) (*Electronic Supplementary 6.1 – Additional File 2*). The same strategy was applied to ACT production to identify metabolite exports predicted as directly competing with the antibiotic production (*Electronic Supplementary 6.1 – Additional File 2*). However, in the case of ACT production for all four strains at both exponential (30h) and at stationary phase (64h) the metabolites exported all directly competed with ACT production, under all circumstances. For example, cAMP is already in direct competition with biomass (*Figure 6.3a*) and also with ACT production (*Figure 6.3c*), whereas 2-oxoglutarate did not compete with biomass (*Figure 6.3b*) but it is competing with ACT production (*Figure 6.3d*). Some of these metabolites cost more than ACT production to the metabolism, for example an increasing cAMP export leads to a faster predicted decrease of biomass than ACT does at exponential phase (32h) (*Supplementary Figure 6.13a*) and even more at transition phase (64h) where nutrients are depleted (*Supplementary Figure 6.13b*). If all the metabolites exported are really in

direct competition with ACT production as predicted by the model, this suggests that even small export rates of these metabolites are using metabolic resources that could be used for ACT production.

The extracellular metabolites detected in the culture media of the different strains were used to further constrain the metabolic models by forcing export of the metabolites. The growth was then optimised in the four metabolic models, at the 6 time points. This allowed to predict the intracellular metabolic fluxes of the organisms at the different time points [43,44]. The correlation between the predicted intracellular fluxes and the ACT production in each strain was calculated to identify metabolic pathways and reactions that are directly dependent on ACT production (*Electronic Supplementary 6.1- Additional File 1*). In M145 and M145+ACT all the reactions highly correlated with ACT production (Pearson correlation r>0.5) are involved in glycolysis (pyruvate kinase, phosphoglycerate mutase, and enolase) and in the pentose phosphate pathway (ribulose 5-phosphate 3-epimerase) (*Supplementary Figure 6.14a*). Whereas in M1146 and M1146+ACT, the reactions highly correlated with ACT production (r>0.5) are the cAMP export reaction, and an NAD transhydrogenase (*Supplementary Figure 6.14b*). There are no reactions correlated with ACT production in common between the three ACT producers (M145, M145+ACT, and M1146+ACT). The 4 reactions correlated with ACT production in the two M145 strains are involved in glycolysis and the PPP which are involved in the production of the ACT precursors acetyl-CoA and malonyl-CoA [48]. However, these reactions do not show correlation with ACT production in M1146+ACT, only the NAD transhydrogenase and cAMP export correlated with ACT production.

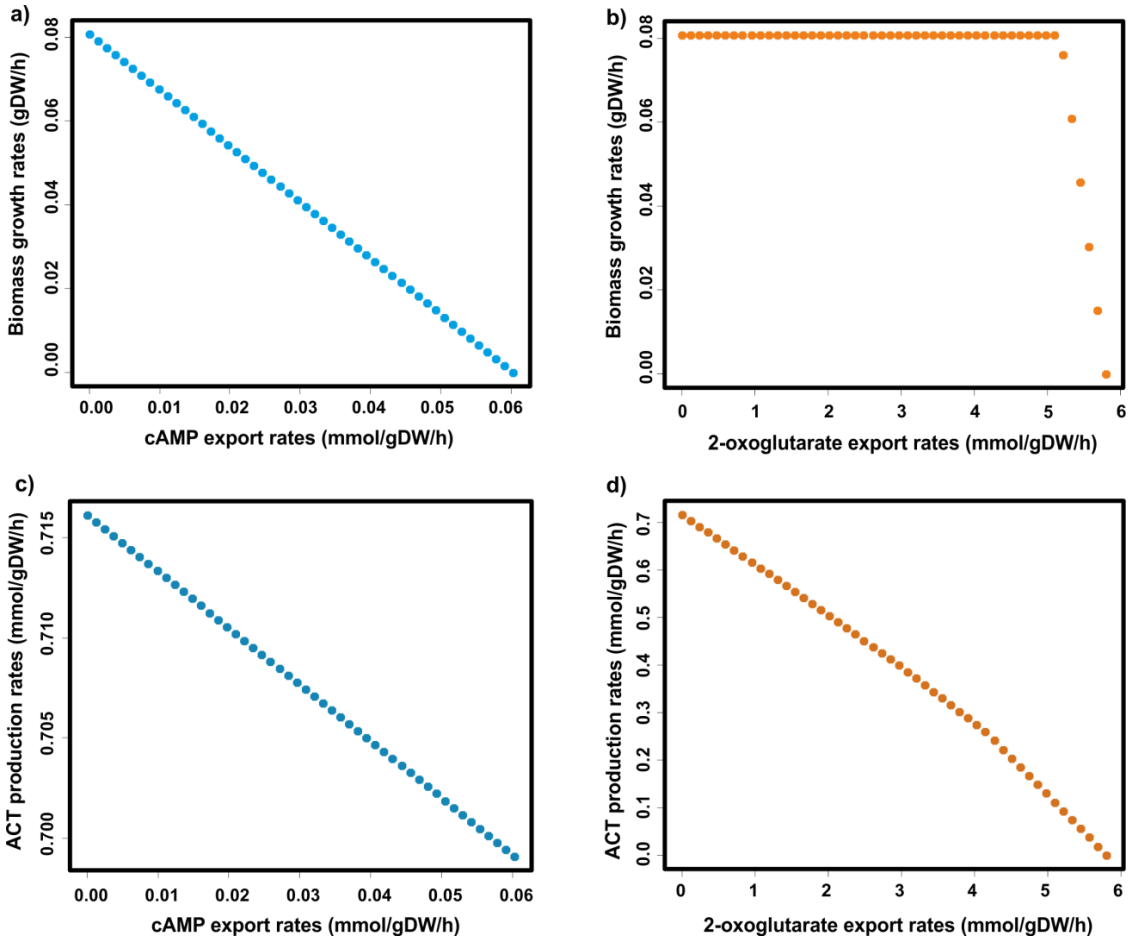


Figure 6.3: Cost of metabolites export to biomass growth and to ACT production rates

Data shown here is a selection of two metabolites exported by M145, these two metabolites are exported in a similar trend (different values) by all four strains, and the full dataset is available in *Supplementary File 6.1 – Additional File 2*.

a) Biomass growth rate to cAMP export rates: direct competition between metabolite export and biomass production. Increasing cAMP export lead to a direct decrease of biomass.
b) Biomass growth rates to 2-oxoglutarate export rates: non-direct competition between metabolite export and biomass production. 2-oxoglutarate export does not impact biomass until a "breaking" point where high export rates lead to a decrease in predicted biomass.
c) ACT production rates to cAMP export rates: direct competition between metabolite export and ACT production. Increasing cAMP export lead to a direct decrease of ACT production.
d) ACT production rates to 2-oxoglutarate export rates: direct competition between metabolite export and ACT production. Increasing 2-oxoglutarate export lead to a direct decrease of ACT production rates.

6.5.3 Metabolic differences associated to metabolites secretions and actinorhodin production

An *in silico* analysis was carried out to identify primary metabolism reactions and pathways needed for ACT production under these conditions; in parallel, an analysis was carried out to predict the metabolic pathways activated by the metabolic exports observed (based on the exometabolome). This analysis aims at determining the metabolic states of ACT producers compared to non-producers, and identify potential metabolic engineering targets (e.g., gene overexpression or knockout). The analysis was focused on the M1146 and M1146+ACT strains as these do not produce other secondary metabolites that M145 produces, so the M1146 has a "cleaner" metabolic background [31]. Furthermore, this enables an easier comparison between an ACT producer (M1146+Act) and a non-producer (M1146). The constrained metabolic models for M1146 and M1146+ACT were used to individually produce the metabolites identified in the M1146 and M1146+ACT exometabolome (*Supplementary Figure 6.11*). The metabolite exports were set as unique objective (no biomass and no ACT production) to identify the metabolic reactions involved in the production of the exometabolites. The comparison of all the predicted intracellular fluxes for the metabolic exports, the growth, and the ACT production helped to identify all the intracellular metabolic reactions that were activated only by the metabolites exported. There were only 60 predicted active metabolic reactions when exporting these metabolites (*Figure 6.4*). Most of the reactions were involved in nucleotide salvage pathways (41% of the reaction). A significant portion of the predicted reactions were associated with pathways involved in carbon central metabolism (32% in total) such as the pentose phosphate pathway (12%), alternate carbon metabolism (10%), anaplerotic reactions (7%), glycolysis and gluconeogenesis (3%); as well as amino-acid

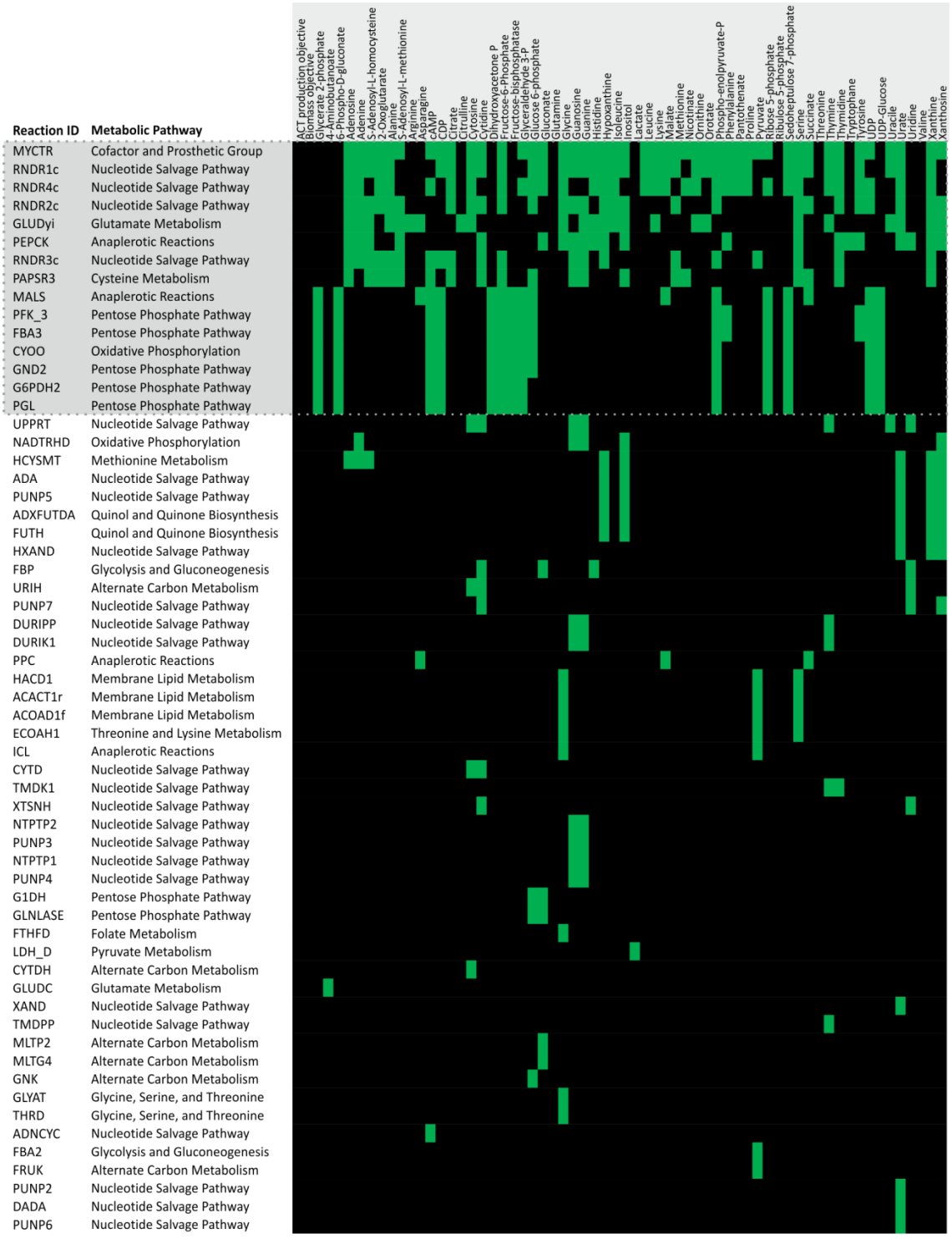


Figure 6.4: Metabolic reactions with fluxes predicted as activated by exports

The metabolic objectives are in the column, such as the ACT production, biomass, and the individual metabolic exports. The metabolic reactions exclusively switched ON by the metabolic exports and not carrying fluxes when producing biomass or ACT are highlighted in green with associated metabolite exported that activated the reaction. If the reaction did not carry any flux the column is in black. A group of reactions were frequently (> 15 times) predicted as switched ON by the different metabolic exports these are shown by the dashed grey box.

metabolism (12%) (*Figure 6.4,* and *Supplementary Figure 6.15*). The activation of the nucleotide salvage pathways and carbon central metabolism reactions make sense when taking the measured data into account as the metabolites exported by M1146 and M1146+ACT were largely nucleotides (33%) of all the exported metabolites, while central carbon metabolites represent about 15%. Interestingly, amino acids represent 42% of the metabolites exported but only 12% of the metabolic reactions were predicted to be activated. Only a few reactions associated to amino acids metabolism are predicted as activated by the exports (despite many amino acids identified in the media), because the main amino acid metabolism reactions are already carrying fluxes for growth; so these reactions are not activated but may carry higher fluxes than predicted. The predicted most active metabolic pathways (most frequently switched ON) due to the metabolic exports are involved in the nucleotide salvage pathway and the PPP (*Figure 6.4*). These two pathways are connected by the PPP intermediate ribose-5-phosphate that is converted into 5-phosphoribosyl-1-pyrophoshate (PRPP) to enter the *de novo* biosynthesis of nucleotides [49].

An analysis was carried out to understand if the predicted metabolic pathways activated by the exports are linked to ACT production. First, the analysis focused on identifying metabolic reactions directly associated to ACT production. The ACT production in the model was progressively increased from 0 to the maximal value possible (optimum), the fluxes were predicted (*Electronic Supplementary 6.1 – Additional File 3*), then the reactions with fluxes highly correlated (Pearson correlation, r>0.7) with production were isolated (*Supplementary Figure 6.16*). The model predicted that under these conditions, most of the metabolic reactions correlated to ACT production are involved in central metabolism (53%), because glycolysis is the

main pathway converting glucose in the media leading to the acetyl-CoA precursor needed for ACT biosynthesis. The other reactions are associated to exchange and transport reactions, oxidative phosphorylation, and glutamate degradation (*Supplementary Figure 6.16*). The predicted metabolic reactions activated by exports and the predicted metabolic reactions associated to ACT biosynthesis were visualised onto metabolic maps of the central metabolism and of the nucleotides metabolism (*Figure 6.5, and Supplementary Figures 6.17*). Parts of the central and nucleotides metabolisms are only predicted as active when the metabolic model is optimised for growth, so these reactions were also added to the metabolic maps. While other parts of metabolism are only activated by the metabolic exports or are highly correlated to ACT production.

In nucleotide metabolism, the metabolite exports led to the predicted activation of parts of the adenine, uridine, cytosine, guanine, and thymine metabolism, but did not activate the inosine metabolism which was already active for biomass production (*Supplementary Figures 6.17*). The activation of nucleotide metabolism is either due to a degradation of DNA and RNA (autophagy) by the organism bringing nucleotides into the nucleotide salvage pathway [49], or to *de novo* production of purine and pyrimidines and derivatives from the PPP [49]. The ACT producers' metabolomics data showed higher levels of nucleotides and derivatives (*Supplementary Figure 6.10 and 6.11*), such as increased cAMP levels in the media and highly correlated with ACT production (*Supplementary Figure 6.12a and 6.12b*), and PRPP levels are also higher in the producer strain (*Supplementary Figure 6.18a and 6.18b*). The purine and pyrimidines metabolism is important for antibiotic production in *Streptomyces*, for example the accumulating levels of ppGpp (coming from the guanine metabolism) are

critical in triggering antibiotic biosynthesis [50,51]. This mechanism is triggered by nutrient starvation such as phosphate or nitrogen depletion, leading to accumulation of ppGpp that binds to the *act*II-ORF4 which is the pathway transcriptional activator of the ACT cluster (present in the native strain and in the ACT cosmid) [52]. Interestingly, the cAMP binding protein EshA is involved in ppGpp synthesis, and has been shown to be important for ACT production, as its knockout lead to lower ppGpp levels and no ACT production [53]. Some studies observed higher levels of cAMP after transition phase suggesting that cAMP extracellular levels may have a role in secondary metabolism [45,53,54]; which may explain the strong correlation observed between extracellular cAMP and ACT production levels. Another metabolites with a signalling role for stationary growth and secondary metabolism is the cyclic diGMP; which is produced from GTP, like ppGpp; cyclic diGMP plays a vital role in the control of developmental processes such as sporulation and vegetative growth [51]. During nutrient stress, when ATP levels are decreasing, the depletion of cyclic diGMP leads to morphological differentiation with the accumulation of ppGpp triggering antibiotics production [51]. Here, the metabolic fluxes from the purines salvage pathway generate GTP which is used to produce ppGpp rather than cyclic diGMP. Also, the large activation predicted in purine and pyrimidine metabolism due to the extracellular purine and pyrimidine associated metabolites (which are mostly higher in the ACT producer) suggests that this metabolism is correlated to ACT production.

Multiple parts of the central metabolism predicted have metabolic fluxes that are highly correlated to ACT production, such as the glycolysis and glutamate catabolism (green reactions in the *Figure 6.5*). Glycolysis consumes glucose in the media to generate the ACT precursor acetyl-CoA (then converted in malonyl-CoA), and

the glutamate is degraded to enter the tricarboxylic acid (TCA) cycle through 2-oxoglutarate. The model optimises the consumption of glucose and glutamate to produce ACT. In the constrained model (without exometabolome exports constraints) the optimisation of biomass predicts fluxes through other reactions in the TCA cycle, and to produce glycogen (blue reactions in the *Figure 6.5*). These are not predicted as correlated to ACT production but are still essential to generate metabolic intermediates essential for central metabolism (e.g., TCA cycle intermediate). The metabolite exports activated multiple reactions such as a lactate dehydrogenase, anaplerotic reactions (phosphoenolpyruvate carboxykinase and carboxylase), reactions of the glyoxylate cycle, and reactions of the PPP (red reactions in the *Figure 6.5*). The PPP activation is related to higher nucleotide metabolism, either to produce *de novo* nucleotides through the PRPP or from the degradation of DNA and RNA [49,51]. The observation of lactate in the media suggests that under these conditions the organism may have a fermentative growth [55,56], despite growing in an O_2 rich environment. It was previously suggested that *S. coelicolor* may have a fermentative metabolism as it has a lactate dehydrogenase annotated in its genome [56]. Furthermore, lactic acid production was previously observed in *Streptomyces griseus* [55] suggesting a fermentative metabolism. It was previously shown that *Streptomyces* cultures have a tendency to aggregate in liquid cultures leading to a depletion of O_2 for the cells in the centre of the aggregate [58], which may explain the observation of lactate despite growing in O_2 rich conditions. The activation of the anaplerotic pathways and of the glyoxylate pathway suggests that fluxes from glycolysis are redirected to replenishing the TCA cycle, which may reduce fluxes toward acetyl-CoA (for ACT production) by going toward the TCA cycle (*Figure 6.5*).

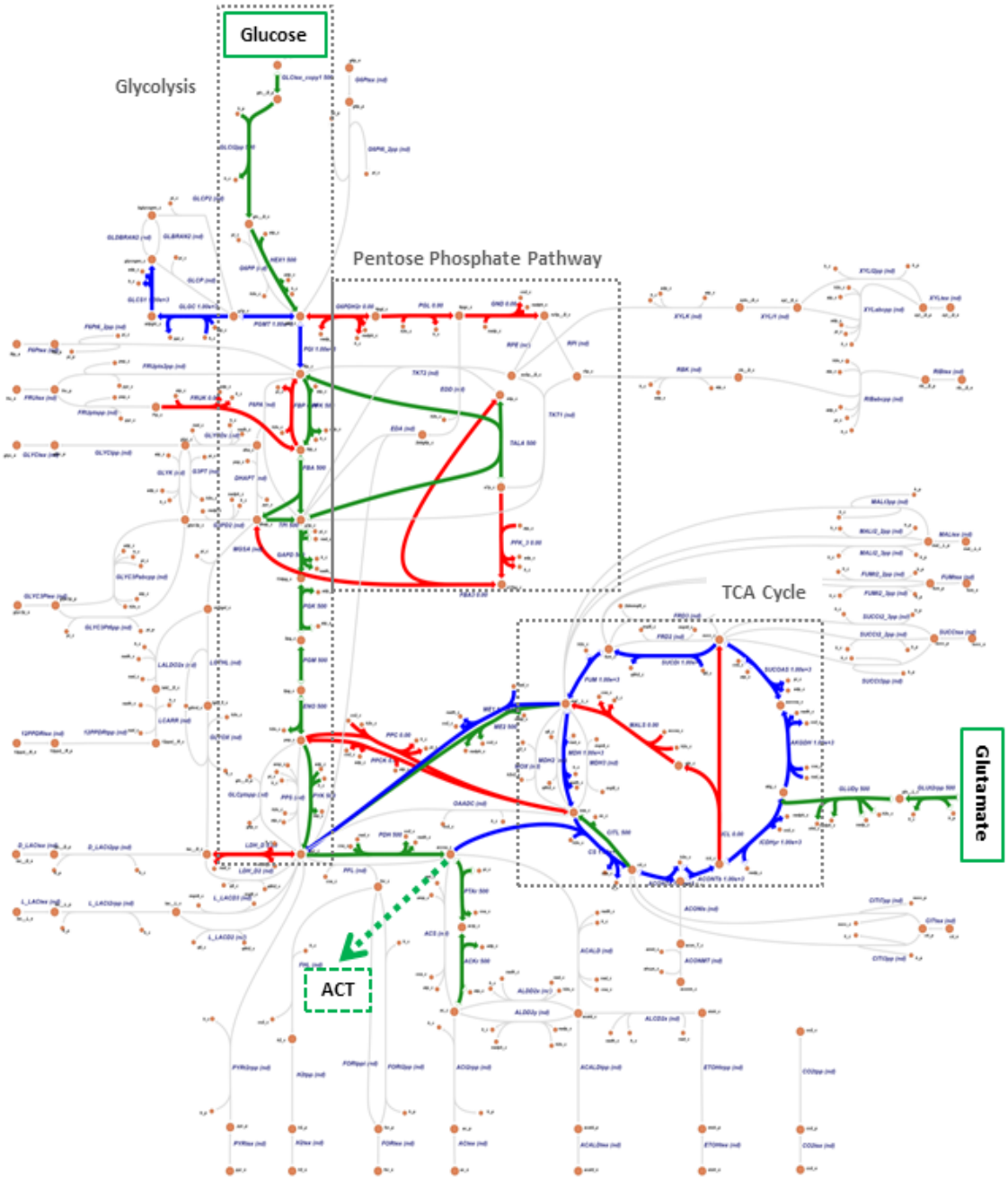


Figure 6.5: Central metabolism map with fluxes associated to ACT production and with metabolic exports

The predicted metabolic reactions highly correlated with ACT production (Pearson correlation, r>0.7) are highlighted in green, with most of the reaction associated to glycolysis and with generating acetyl-CoA for ACT production. The other reactions predicted as carrying fluxes are highlighted in blue, these reactions are associated with the TCA cycle, the anaplerotic reactions, and the glycogenesis. The metabolic reactions predicted as switched ON are highlighted in red; these reactions are associated with the PPP, the glyoxylate bypass metabolism, and some anaplerotic reactions.

A high-resolution version is available in *Supplementary Electronic 6.1 – Additional File 4*.

6.5.4 Metabolomics data integration to metabolic modelling for rational strain design

The metabolic reactions identified as highly correlated with ACT production may highlight relevant overexpression targets to increase genes expression levels and push metabolic fluxes toward ACT production. While the metabolic reactions activated by exports highlight potential knockouts or downregulation targets to reduce metabolic waste or overexpression targets to redirect these intermediates toward ACT production. Some of the reactions identified as activated by metabolites exports or correlated with ACT in the central metabolism (*Figure 6.5*) were previously experimentally tested in *S. coelicolor* to increase ACT production or study *S. coelicolor* metabolism [48,59–62].

A large part of glycolysis was highly correlated to ACT production, and the PPP was largely active due to the metabolic exports constrained in M1146+ACT and M1146 (*Figure 6.6*). A study in M145 reduced glycolytic fluxes by deleting a phosphofructokinase isoenzyme gene (*pfkA2*) to increase fluxes toward the phosphoketolase from the PPP increasing acetyl-CoA production, which lead to a significant increase in ACT and RED production [62]. In another study in M145, the fluxes through the PPP were reduced by knocking-out the glucose-6-phosphate dehydrogenase (*zwf*), which led to a decreased yield in ACT [59]. Furthermore, the PPP generates NADPH which is an essential cofactor for ACT production, the importance of NADPH generated by the PPP for antibiotics production was previously highlighted in *S. coelicolor* [62,63] and *S. lividans* [48,64]. Interestingly, the NADP+ levels were measured as higher in the M1146+ACT than in M1146 (*Supplementary Figure 6.19a*), so the producing strain may benefit from higher NADPH fluxes if the balance of NADPH/NADP+ was low (no data about NADPH levels). Also, a carbon-flux analysis

previously showed that M1146 strain had higher PPP fluxes than M145, as well as higher levels of NADPH due to oxidative stress [65].

The metabolic model predicted that the transaldolase (TALA) fluxes were highly correlated with ACT production (*Figure 6.5*). The transaldolase reversibly converts sedoheptulose-7-phosphate (S7P) and glycerate-3-P (G3P) into erythrose-4-phosphate (E4P) and fructose-6-phosphate (F6P). The model also predicted a conversion of E4P and F6P into S7P and G3P by the TALA, which makes glycolysis more efficient by avoiding the conversion of F6P into fructose-1,6-biphosphate (FBP) and into G3P and DHAP (saving 1 ATP). Bypassing the preparatory phase of glycolysis by reducing the conversion of the F6P into FBP (*pfkA2* KO) [62] reduces the use of ATP, while redirecting fluxes toward the PPP generating more NADPH [63] and directly leading to G3P with the TALA to enter the oxidative part of glycolysis generating more ATP and continue generating pyruvate and acetyl-CoA, as well as going toward more acetyl-CoA with the phosphate acetyl-transferase (PAT) from the transketolase pathway [62]. The FBP and ATP are strong allosteric inhibitors of the PFK, so these two metabolites can inhibit the glycolysis and reduce glycolytic metabolic fluxes. So, bypassing the PFK enzymatic step may have increase the glycolytic fluxes going toward acetyl-CoA. The gene associated with the TALA enzyme (SCO6662) was identified as a strongly activated by glucose in *S. coelicolor*, so this enzyme is likely to have a role with glucose degradation [66,67]. Hence, a combined increase of fluxes in specific part of the PPP (TALA and PAT) as well as decreased fluxes in the preparatory phase of glycolysis (PFK) would increase ACT production by increasing fluxes toward acetyl-CoA and NADPH.

The anaplerotic reactions are important to replenish the TCA cycle intermediates and are involved in carbon metabolism response to nutrient depletion

(switch between catabolism and anabolism) [61,68]. The malic enzyme (*me2*) generating NADPH and CO_2 while converting malate into pyruvate is predicted as highly correlated with ACT production, whereas the other malic enzyme (*me1*) is predicted as carrying fluxes for biomass production but generates NADH (*Figure 6.5* and *6.6*). While the phosphopenolpyruvate carboxylase (PEPC) and the phosphoenolpyruvate carboxykinase (PEPCK) were activated by the metabolite exports (*Figure 6.5* and *6.6*). Overexpression of the malic enzyme genes (*me1* and *me2*) could help ACT production by increasing the fluxes consuming TCA cycle intermediates to generate pyruvate and afterward acetyl-CoA. For PEPCK and PEPC, these enzymes are heavily regulated as these have an anaplerotic role to replenish oxaloacetate (OAA) from phosphoenolpyruvate (PEP) or to generate more PEP from OAA. The PPC converts PEP into OAA while consuming CO_2 and H_2O and producing Pi, whereas PEPCK converts OAA into PEP while consuming ATP and producing ADP and CO_2 [68]. The malic enzymes, the PEPC, and the PEPCK are associated to the PEP-pyruvate-OAA node involved in the switch of carbon fluxes between catabolism and anabolism [68]. The deletion of the malic enzymes (*me1* and *me2*) led to a strong reduction in ACT production and an accumulation of TAG in *S. coelicolor* [69]. These deletions seems to trigger an accumulation of TCA cycle intermediates disturbing the carbon flux balance, which may impact regulation of the ACT cluster as the ActII-ORF4 transcription decreased in the mutant [69]. The regulation of these enzymes is still unclear. PEPCK is used to generate OAA from PEP for gluconeogenesis [66,70,71]; however, this enzyme has been reported as operating in the opposite direction as well [67,72]. So, the overexpression of PEPCK in *Streptomyces lividans* led to an anaplerotic activity of PEPCK rather than a PEP production from OAA [71]. So, deleting these enzymes would be a better strategy

than overexpression to ensure that their anaplerotic activity is absent and keep the PEP usage for pyruvate formation.

The metabolic model predicted a high correlation between ACT production and the citrate lyase (CITL) which can convert citrate into OAA and produce an acetate byproduct that can be transformed into acetyl-CoA (via the acetyl-phosphate pathway). Another enzyme, the citrate synthase (CS), converts OAA and acetyl-CoA into citrate. The CITL overexpression has never been tested, while its overexpression could lead to increased level of acetate convertible into acetyl-CoA. The metabolic exports also activated fluxes in the glyoxylate bypass metabolism, where the malate synthase (MALS) consumes acetyl-CoA and glyoxylate to produce malate. But, the deletion of a citrate synthase (*citA*) in *S. coelicolor* made it a glutamate auxotroph in minimal media, and lead to defective antibiotic biosynthesis [60,73]. This pathway is heavily regulated [74,75] and is important in oxidative response which may impact ACT production [76]. Multiple TCA cycle intermediates were measured as higher in M1146+ACT than in M1146 (citrate, isocitrate, 2-oxoglutarate, succinyl-CoA) as well as some amino-acids derived from the TCA intermediates (isoleucine, arginine, methionine, S-adenosylmethionine). Reducing the fluxes toward the TCA cycle (anaplerotic reactions deletion) could reduce amino-acids biosynthesis and consumption of acetyl-CoA (glyoxylate pathway deletion). Furthermore, as discussed in *Chapter III*, this bypass is absent from the *S. lividans* strain and may have a role in differences of ACT production between *S. lividans* and *S. coelicolor*, however, the impact of the bypass absence is still unknown. Another enzyme was highly correlated to ACT production, the acetyl-CoA carboxylase (ACC) responsible for the conversion of acetyl-CoA into malonyl-CoA (*Supplementary Figure 6.16*). The overexpression of the

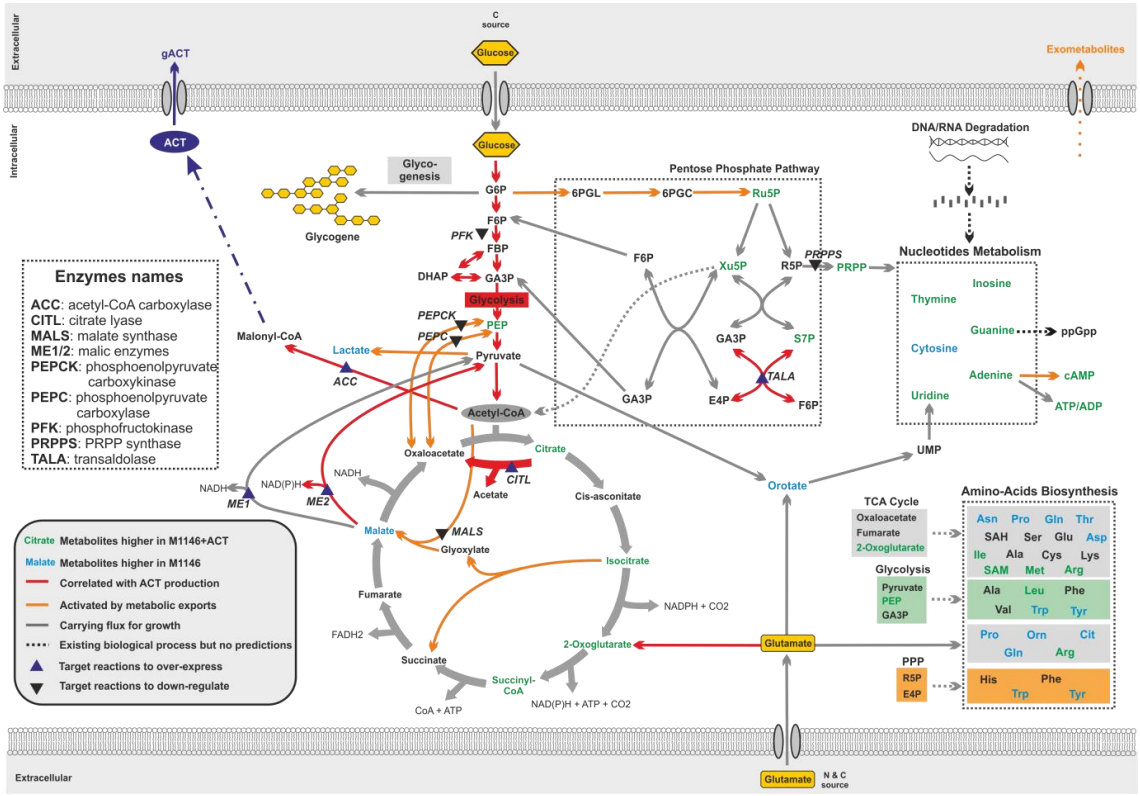


Figure 6.6: Summary metabolic map of ACT producer versus non-producer

The non-producer M1146 and the ACT producer M1146+ACT strains were compared using the metabolomics data and the metabolic predictions. The comparison is focused on the central metabolism connection to the pentose phosphate pathway and the purines and pyrimidines metabolism. The metabolic predictions for the reaction associated to ACT production (reactions in green) are showed with the metabolic predictions resulting from the exometabolome integration (reactions orange) and the prediction of biomass (reactions in grey). Based on the predictions and literature a set of enzymatic reactions that could carry higher fluxes to increase ACT production (e.g., by gene overproduction) are represented (purple up arrow). As well as reactions to block (e.g., via gene) are represented (black down arrow).

Metabolites abbreviations: G6P: glucose-6phosphate; F6P: fructose-6-phosphate; FBP: fructose-1,6-biphosphate; GA3P: glyceraldehyde-3-phosphate; DHAP: dihydroxyacetone phosphate; PEP: phosphoenolpyruvate; 6PGL: 6-phospho-D-glucono-1,5-lactone; 6PGC: 6-Phospho-D-gluconate; Ru5P: D-Ribulose 5-phosphate; Xu5P: D-Xylulose 5-phosphate; R5P: Ribose 5-phosphate; PRPP: phosphoribosyl pyrophosphate; S7P: sedoheptulose-7-phosphate; E4P: D-Erythrose 4-phosphate.

Enzymes abbreviations: PFK: phosphofructokinase; PEPCK: phosphoenolpyruvate carboxykinase; PEPC: phosphoenolpyruvate carboxylase; CITL: citrate lyase; MALS: malate synthase; ME1/2: malic enzymes; ACC: acetyl-CoA carboxylase; PRPPS: PRPP synthase; TALA: transaldolase.

ACC showed an increased ACT yield as it increased acetyl-CoA conversion into malonyl-CoA (ACT main precursor) [59]. So, a combination of knockouts to reduce fluxes toward the TCA cycle consuming acetyl-CoA with an overexpression of the known ACC target would force fluxes toward conversion of acetyl-CoA for ACT production instead of production of TCA intermediates.

Finally, the higher levels of purines and pyrimidines (except cytosine) exported by M1146+ACT, with the higher level of intracellular PRPP, with the higher predicted fluxes in the PPP reactions and nucleotide salvage suggest that nucleotide metabolism has an important role in ACT production. However, its precise metabolic role in ACT production is still unclear; there are no ACT correlated reactions associated with nucleotide metabolism. But nucleotide levels have been shown to have an important regulatory role on ACT production (e.g., ppGpp, ATP/ADP, cAMP) [51]. The increase in nucleotide metabolism may come from DNA/RNA degradation through the nucleotide salvage pathway, or via d*e novo* biosynthesis of nucleotides. If the *de novo* biosynthesis is predominant during ACT production it would be interesting to reduce the PRPP synthase (PRPPS) activity that consumes the R5P from PPP to generate PRPP, which is then used for nucleotides biosynthesis. It seems like the PRPP synthase has never been deleted in a *Streptomyces* strain before, so it may also help to learn more on the impact of the nucleotide metabolism on antibiotics production. However, it would require to first experimentally confirm that the observed increase in nucleotides metabolism originate from *de novo* biosynthesis and not from degradation of DNA and/or RNA.

The identification of single knockouts in primary metabolism presents a major challenge as primary metabolism, particularly the central carbon metabolism, is heavily

regulated and has multiple routes for similar functions to ensure robustness and adaptation of the organism [77]. However, a more holistic approach to rewire the central metabolism has proved successful in other model organisms such as *E. coli*, *C. glutamicum*, and *S. cerevisiae* to respectively increase bioproduction of poly(3-hydroxybutyrate) ,methyl-3-hydroxybutyrate, and ethanol by combining multiple gene KOs and overexpression [61]. Here, based on the integration of the metabolomics data and metabolic modelling, combined with results from published experimental studies, a set of 5 enzymes to KO or downregulate and 5 enzymes to overexpress were identified (*Figure 6.6*). In total, 4 out of the 5 knockout targets and 4 out of the 5 overexpression targets suggested have not been tested experimentally in *S. coelicolor* before.

The phosphofructokinase has three different associated genes *pfkA1-3* (respectively SCO2119, SCO5426, and SCO1214), where the *pfkA2* deletion was shown to increase ACT production [62]. Interestingly, the *pfkA2* gene (SCO2119) in M145 in the glutamate minimal media is expressed during exponential and transition phase while the two other genes *pfkA1* and *pfkA3* have a higher expression during exponential phase than in transition phase (ACT production starting) [22] (*Supplementary Figure 6.20a*). This may explain why the ACT production increase was only observed with *pfkA2* KO and not with the two other genes [62]. The PRPP synthase KO has never been tested *in vivo* in *Streptomyces* species, but PRPP synthase *in silico* KO is predicted as essential in the minimal media. In *Escherichia coli*, the PRPP synthase KO made it a NAD, guanosine, uridine, histidine, and tryptophan auxotroph [78]; *S. coelicolor* predicted growth was recovered when importing these metabolites. So, to only reduce fluxes through the PRPP synthase (without KO the reaction) deleting one of the two genes

associated with this reaction (SCO0782 and SCO3123) within a complex media (containing NAD, guanosine, uridine, histidine, and tryptophan) would be an option. One of the two genes SCO3123 expression decreases after transition phase whereas SCO0782 expression increases after transition phase in *S. coelicolor* M145 under the same media conditions [22] (*Supplementary Figure 6.20b*). Thus, deleting SCO0782 might help to avoid *de novo* production of purines and pyrimidines from PPP intermediates after transition phase without reducing the growth potential during exponential phase. The anaplerotic reactions PEPCK and PEPC are respectively associated to SCO4979 and SCO3127. The PEPC gene SCO3127 heterologous overexpression in *S. lividans* led to a decrease in growth rate and the authors suggested that this was due to a sequestration of PEP that also led to a delayed onset of secondary metabolism [79]. The overexpression of PEPCK in *S. lividans* led to a growth decrease and to a higher yield of amino-acids and protein [71], as well as a decrease in pyruvate levels. So the deletion of the PEPC and/or PEPCK could help decrease the TCA cycle intermediates replenishing from the PEP, and increase the available pool of PEP for pyruvate generation. Interestingly, PEPCK and PEPC genes expression in M145 have an opposite pattern, PEPC (SCO3127) expression is higher during exponential phase and the opposite for PEPCK (SCO4979) (*Supplementary Figure 6.20c and 6.20d*). In theory, the sole deletion of PEPCK gene (SCO4979) would have less impact than PEPC on growth as it is mostly expressed during stationary phase. The malate synthase (MALS) deletion has not been tested in *Streptomyces* species. Its deletion may reduce the use of acetyl-CoA, while helping to understand its possible role in ACT production (see *Chapter III*) [80]. The glyoxylate bypass was shown to have an important role for glucose and glutamate catabolism regulation and oxidative metabolism in bacteria [74,81], so the positive or

negative impact of its deletion on ACT production is hard to predict as it may cause large regulatory changes in the organism. This enzyme has two genes associated to it (SCO0983 and SCO6243) both genes do not show any distinct expression pattern in M145 [22] (*Supplementary Figure 6.20e*), so no particular is prioritised for deletion.

The overexpression targets for the two malic enzymes NADH (*me1*) and NADPH (*me2*) dependent are respectively associated to SCO2951 and SCO5261. The deletion of these two genes in *S. coelicolor* abolished ACT production [69]. The expression of both genes is higher during exponential phase then decreases after transition phase (*Supplementary Figure 6.21a*) [22]. Overexpression of these two genes may redirect fluxes from the TCA cycle to pyruvate to produce more acetyl-CoA during exponential and transition phase. Another target, the acetyl-CoA carboxylase is encoded by multiple genes *accA1* (SCO2777) or *accA2* (SCO4921), *accC* (SCO2445) or *accB* (SCO5535) and *accE* (SCO5536). The ACC was previously overexpressed by heterologous expression of *accA2*, *accB*, and *accE* leading to an increased ACT production [59]; all three genes are the only ones mostly expressed at early exponential phase (*Supplementary Figure 6.21b*). The citrate lyase was not studied in *Streptomyces* before, but overexpression of CITL in yeast to increase fluxes toward acetyl-CoA for terpenoids production revealed that in this strain the acetyl-CoA synthase was a limiting step to convert the acetate into acetyl-CoA [82]. In *Streptomyces*, the acetate is converted into acetyl-phosphate, then into acetyl-CoA via an acetate kinase and a phosphotransacetylase; so there is a risk that the enzymes degrading the acetate into acetyl-CoA may be limiting. The two genes associated to the citrate lyase are expressed across all growth phases (*Supplementary Figure 6.21c*). The transaldolase (TALA) has not been studied yet in *Streptomyces* species. The two genes associated to

TALA (SCO1936 and SCO6662) have an expression higher during the exponential phase than during stationary phase (*Supplementary Figure 6.21d*), so their expression during stationary phase may help to increase the fluxes from the PPP toward the second part of glycolysis to generate acetyl-CoA. These targets are summarised in *Table 6.2* and in the metabolic map *Figure 6.6*.

Reaction ID	Reaction Name	Pathway	Genes associated	Essential Reaction	Possible Strategy	Published Approach in *Streptomyces*
	Deletion or downregulation targets					
PFK	Phosphofructokinase	Glycolysis	SCO2119 (pfkA1) or SCO5426 (*pfkA2*) or SCO1214 (*pfkA3*)	NO	Isoenzyme deletion *pfkA2* (SCO5426)	SCO5426 deletion increased ACT production
PRPPS	PRPP synthase	Nucleotides Biosynthesis	SCO0782 or SCO3123	YES	Deletion of SCO0782 and/or SCO3123 in a complex media containing NAD, guanosine, uridine, histidine, and tryptophan	NA
PEPCK	Phosphoenolpyruvate carboxykinase	Anaplerotic Reaction	SCO4979	NO	PEPCK and PEPC double knockout SCO4979 & SCO3127 Or PEPCK (SCO4979) single KO	Overexpression decrease secondary metabolism
PEPC	Phosphoenolpyruvate carboxylase	Anaplerotic Reaction	SCO3127	NO		Overexpression increase amino-acids production
MALS	Malate synthase	Glyoxylate Bypass	SCO0983 or SCO6243	NO	Double knockout of SCO0983 & SCO6243	NA N.B.: Absent in *S. lividans*
	Overexpression targets					
ME1	Malic enzyme (NADH dependent)	Anaplerotic Reaction	SCO2951	NO	Overexpression of SCO2951 and/or SCO5261	Deletion of both genes led to abolition of ACT production
ME2	Malic enzyme (NADPH dependent)	Anaplerotic Reaction	SCO5261	NO		
ACC	Acetyl-CoA carboxylase	Precursor Biosynthesis	(SCO2445 or (SCO5535 and SCO5536)) and (SCO2777 or SCO4921)	YES	Overexpression of couple of isoenzyme	Overexpression of SCO4921, SCO2445, and SCO5536 increased ACT production
CITL	Citrate lyase	TCA Cycle	SCO2033 or SCO6471	NO	Overexpression of one of the genes	NA
TALA	Transaldolase	Pentose Phosphate Pathway	SCO1936 or SCO6662	NO	Overexpression of one of the genes	NA

Table 6.2: Summary of potential genes targets to delete or overexpress

Part II: Heterologous production of GE2270A in *S. coelicolor* chassis strain

GE2270A is a thiopeptide RiPP antibiotic natively produced in *P. rosea* and heterologously expressed in the *S. coelicolor* M1146 [23,24]. Here, the heterologous production of GE2270A was studied *in silico* in the *S. coelicolor* metabolic model to identify potential gene knockout targets to increase its production.

6.5.5 GE2270A production in the *S. coelicolor* metabolic model

The GE2270A is a RiPP compound (more details in *Chapter I*). The first step of the biosynthetic pathway is the production of a precursor peptide (containing a leader peptide and the final peptide) by the ribosome, where the main metabolic precursors are the aminoacyl-tRNAs used by the ribosome. The precursor peptide is then modified by multiple enzymes, first by serine and threonine dehydratases alongside a thiazoline dehydrogenase and a cyclohydratase to modify the amino-acids in the peptide. A cycloaddition leads to a cyclic precursor peptide, and then the leader peptide is cleaved generating the final compound GE2270A [25,83]. The enzymes responsible for the post-translational modifications on the peptide use ATP and FMN cofactors, so these two metabolites are also important for GE2270A production. This biosynthetic pathway was introduced in the *S. coelicolor* genome-scale metabolic model *i*AA1259 [17]. As no experimental data was available the metabolic model was constrained with an estimated composition of the complex media CMAN (see *Methods*). First, the metabolic cost of producing GE2270A was analysed, when GE2270A production rates increase, the growth rates decrease (*Figure 6.7*). So GE2270A production is in direct competition with biomass. This is expected as GE2270A's main precursors are aminoacyl-tRNAs that are also needed for biomass, so

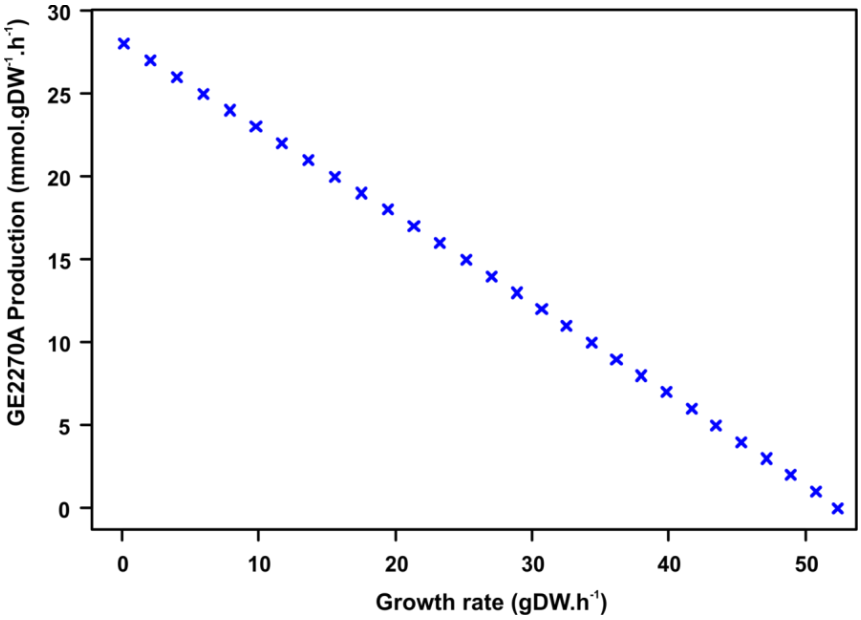


Figure 6.7: Trade-off of GE2270A production to biomass growth

The GE2270A production is in direct competition with biomass, as the growth rate decreases when the GE2270A production rate increases. So the two functions are in competition for metabolic resources.

the two functions compete for resources such as amino-acids and ATP (needed to load amino-acids on the tRNA).

The native strains start producing antibiotics at transition and stationary phase when the growth slows down or stops, hence the biomass and antibiotics cannot really be considered as in competition under these conditions. However, this is relevant here for a heterologous overproducing strain, where production is triggered early in the growth phase. In the native strains, antibiotics biosynthesis may rather help to cope with excess of metabolites, such as ATP and reducing metabolites, which are generated by an excess of catabolism over anabolism during slow or no growth.

The predicted metabolic exchanges with the constrained media are a good way to study the nutrients needed for the different functions (i.e., biomass and GE2270 production) and identify important media components. The metabolic exchanges with

the CMAN media predicted by the model were analysed for optimal production of GE2270A (*Supplementary Figure 6.22*), for optimal biomass production (*Supplementary Figure 6.23*), and for optimal production of biomass with a constrained GE2270A production (*Figure 6.8*). All of the exometabolites needed for GE2270A production are also needed for growth; this partly explains the direct competition between the two functions (*Figure 6.7*). The import of all amino acids (except glutamate) increased for GE2270A production compared to biomass, highlighting their importance for the secondary metabolite production. So, serine, histidine, asparagine, methionine, glycine, leucine, proline, phenylalanine, threonine, and cysteine are important component of the media to produce GE2270A (and biomass). While glutamate, maltohexose, and phosphate imports decreased because these are mostly used for biomass production, and the GE2270A constraint led to a decrease in biomass compared to the optimal biomass production levels so these imports decreased (*Figure 6.7*). Many exometabolites imported are exclusively used for biomass production and not for GE2270A production, including amino-acids, vitamins, and metals ions. The associated metabolism also produces exported by-products, for example, the only export unique to GE2270A production is the GE2270A compound and the cleaved leader peptide. However, the production of GE2270A led to an increase in exported dialuric acid, formate, and 5-dehydro-D-gluconate, which are associated to metabolic pathways in amino-acids biosynthesis, pentose phosphate pathway, and cofactors recycling. The alanine exports decreased due to GE2270A production, because the alanine biosynthesis fluxes were partially redirected to tRNA-alanine production for GE2270A biosynthesis (*Figure 6.7*). However, amino-acid imports do not reach the maximum import constraint, and the limiting resource is the

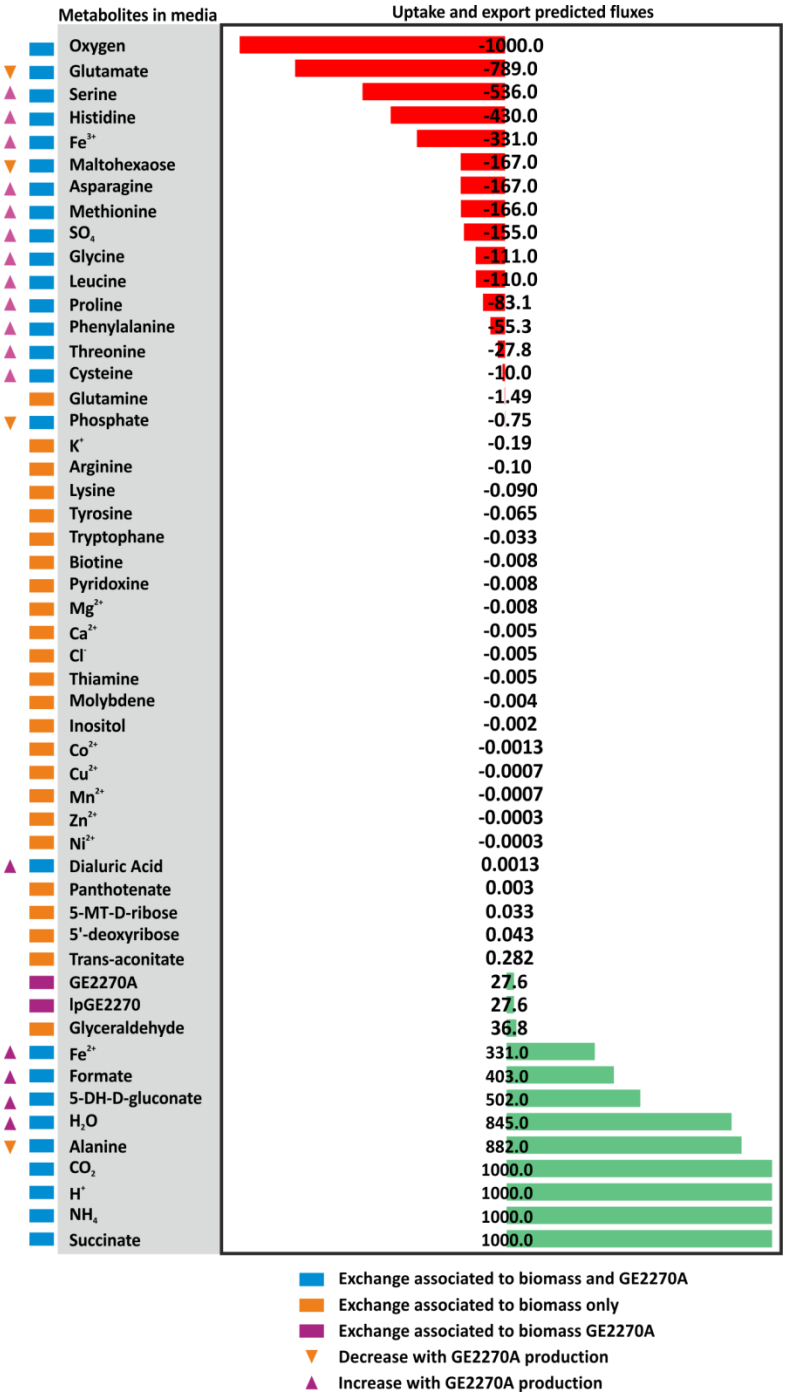


Figure 6.8: Predicted metabolite imports and exports from media for *S. coelicolor* growth and GE2270A production

The predicted exchange fluxes are represented here for *S. coelicolor* M1146 in a CMAN media producing GE2270A and biomass, with the import fluxes (red bars) and export fluxes (green bars). Many metabolites were only imported or exported to allow growth (metabolites with an orange square), while GE2270A and its leader peptide (lpGE2270A) were only produced for GE2270A production (purple square), and many metabolites were exchanged for both biomass and GE2270A production (blue square). The production of GE2270A led to some import and export of some metabolites to increase (purple triangle) and some to decrease (orange triangle).

O_2 needed (reaching the maximum import) for oxidative phosphorylation to generate the ATP needed to convert the imported amino-acids into aminoacyl-tRNAs (*Figure 6.6*). Thus, the cost of ATP for growth and for GE2270A production is responsible for the competition. Here, it is important to have enough amino acids available in the media for both biomass and GE2270A production; such as the following amino acids: serine, histidine, asparagine, methionine, glycine, leucine, proline, phenylalanine, threonine, and cysteine.

6.5.6 Identification of gene knockouts targets to improve GE2270A production

The metabolic model producing GE2270A and constrained with the CMAN media is used to determine the essential and non-essential single and double gene knockouts. The gene KOs identified as non-essential for GE2270A and for growth in the CMAN media will be used to identify a list of potential gene KO to increase GE2270A production. The majority of the single gene KO had little to no impact on growth or on GE2270A production (*Figure 6.9a*). Most of the predicted single gene deletions were non-essential with 986 genes (more than 78% of all genes) and only 52 genes reduced growth by less than 10% , whereas 152 genes (12%) were essential for growth, and only 12 genes deletions were predicted as reducing growth by more than 10% (*Supplementary Table 6.2*). For GE2270A production, the very large majority of the single gene deletions were non-essential for production, with 1123 genes (more than 89% of all genes), and only 38 genes were predicted as reducing production by less than 10%, whereas only 14 genes (12%) were predicted as essential for production and only 27 genes reduced production by more than 10% (*Supplementary Table 6.3*). The vast majority of the single gene KO have no or a small impact (<10% reduction of the

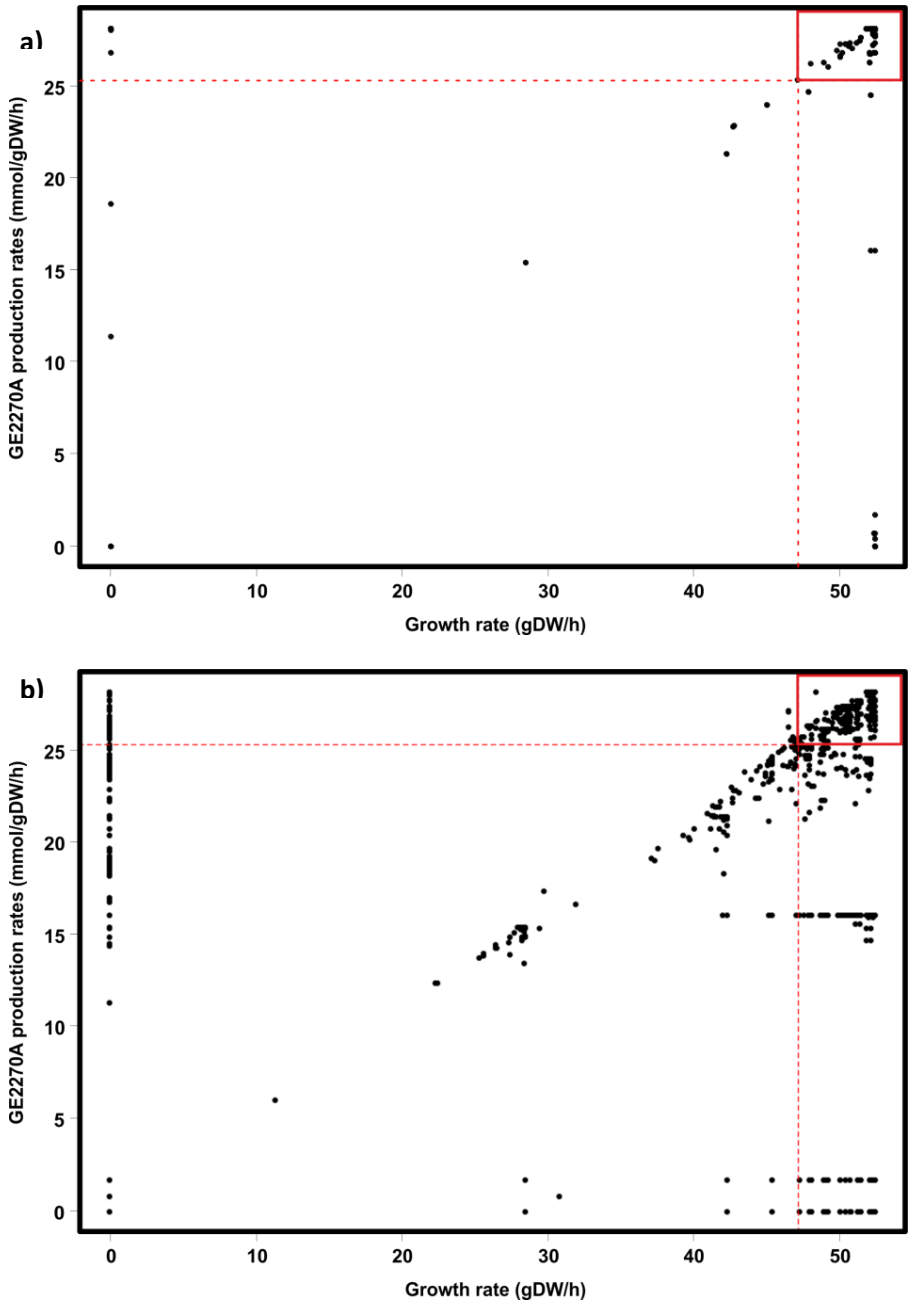


Figure 6.9: Growth rate to GE2270A production for single and double gene knockouts

a) Single-gene knockout effects on biomass and GE2270A production optimisation. The red lines correspond to the optimum value -10%, the red square at the intersection corresponding to the single knockouts with maintained fluxes toward both biomass and GE2270A. The majority of the single knockouts have a limited impact on both functions, hence the clustering in the top right corner (red square).

b) Double-gene knockouts effect on biomass and GE2270A production optimisation. The red lines correspond to the optimum value -10%, the red square at the intersection corresponding to the double knockouts with maintained fluxes toward both biomass and GE2270A. For the double knockouts as well, the majority of the knockouts have a limited impact on both functions, hence the clustering in the top right corner (red square).

objective) on biomass and GE2270A. This group of genes that have no or a very small impact are clustered together (*Figure 6.9a*) and will be used to identify potential gene KO to increase GE2270A production without significantly decreasing growth. The same analysis was applied with double KO (*Figure 6.9b*). The double KO showed a similar tendency where most of the double gene KO were non-essential for growth (55%) or for GE2270A production (88%), while a small portion of the double KO were essential for growth (19%) or for GE2270A production (0.02%) (*Supplementary Table 6.4* and *6.5*). In the single and double KO none led to an increase in optimal production of biomass or GE2270A; because the metabolic model already optimises growth or production so the flux distribution is already optimised and avoids metabolic pathways or reactions that could lead to a lower optimum.

The list of non-essential genes significantly decreases the space of potential gene KO targets, which were used as input for the OptGene algorithm [38]. From a list of non-essential genes, with an objective function to optimise (i.e., GE2270A production), a minimal growth rate constraint, and media constraints, the OptGene algorithm determines the best genes KO combination to optimise the fitness function (GE2270A flux). OptGene is an evolutionary algorithm that simulate the propagation of beneficial mutations to optimise the fitness function. It generated a population of 500 individuals with random mutations. The model growth is optimised by FBA and the individuals were ranked based on their fitness score (GE2270A production). The best individuals are selected and crossed with 500 other individuals from the next generation, the steps were repeated over 10,000 generations. In total, the algorithm tested 500,000 *in silico* mutants to identify potentially beneficial mutations to GE2270A production (*Supplementary Figure 6.24*). The algorithm reached a near optimal fitness score after

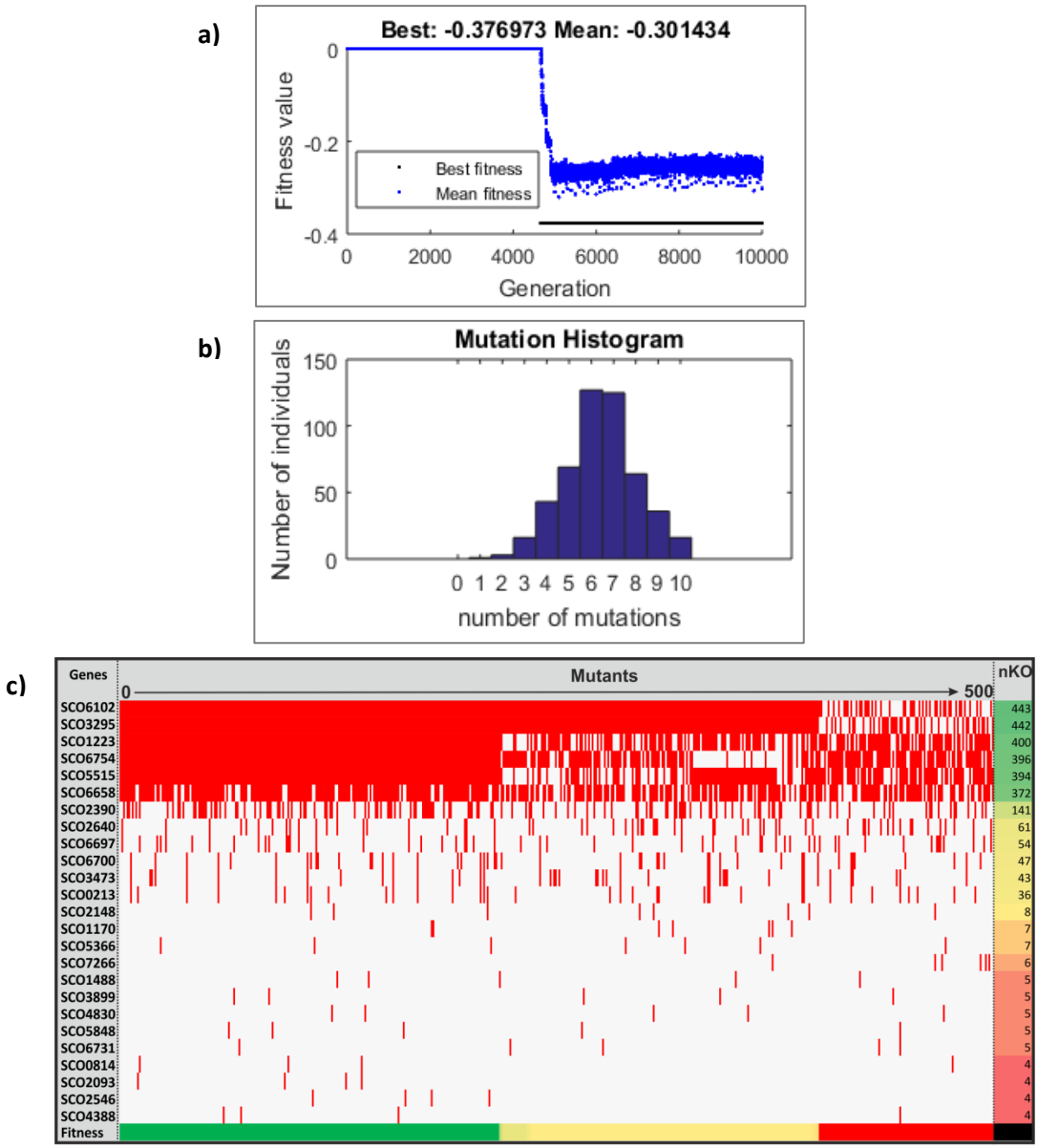


Figure 6.10: Summary of the OptGene output for the production of GE2270A in M1146

a) Diagram of the *in silico* mutants fitness score over 10,000 generations. The fitness score is determined by the objective function (GE2270A production), the algorithm tries to maximise the fitness score at each generation of mutants. Here the fitness score reaches its maximum around 5,000 generations.

b) Histogram of the number of genes knockouts in the final population of 500 individuals. The optimal number of mutations to maximise the fitness score seems to be from 6 to 7 gene knockouts.

c) Top 25 genes present in the final population after 10,000 generations. In the 500 individuals there are 6 genes that are frequently knocked out in the population between 443 and 372 number of knockouts (nKO). In these 6 knockouts, only two are always present in the individuals with the two highest fitness score (green and yellow bar), SCO3102 and SCO3295.

5,000 generations of mutants (*Figure 6.10a*). The algorithm managed to increase and stabilise the fitness score which means it identified a potential combination of gene deletions providing an advantage for GE2270A production [38]. The optimal number of knockouts in the final population of 500 individuals (after 10,000 generations of mutations) is from 6 to 7 genes knockouts per individual (*Figure 6.10b*). In the final population only 25 genes are present in 4 (0.01% of the mutants) to 443 (89% of the mutants) mutants (*Figure 6.10c*, and *Supplementary Table 6.6*). The genes KO with the highest frequency in the population are probably the ones confering the fitness advantage for GE2270A production. Out of the 25 genes, 6 genes are very frequently knockedout (between 74 to 89% of the time), which should contain the genes that are responsible for this predicted fitness advantage (*Figure 6.10c*). The 6 genes, SCO1223, SCO6102, SCO3295, SCO6754, SCO5515, and SCO6658 are involved in different metabolic pathways from amino-acid metabolism to cofactors metabolism (*Table 6.3*). However, the SCO6658 is not a functional knockout as it is a redundant gene, so its deletion do not have any impact on the metabolism as it does not delete any reactions (*Table 6.3*).

In the final population of 500 mutants, there are three different fitness score values ranked from the highest to the lowest (*Figure 6.10c*). But, only 2 gene knockouts are systematically present across the two groups with the highest fitness scores, SCO6102 and SCO3295, both associated to reactions involved in cofactor regeneration. The five functional genes knockout were simulated in the different possible knockouts combinations to identify their impact on the metabolism and on GE2270A production (*Supplementary Figure 6.25*). The multiple combinations of these 5 genes knockouts lead to 31 strains tested *in silico* that were compared to each other

and to a “WT” reference model not producing GE2270A (M1146 strain). The metabolic models of the different mutants were optimised and ranked based on the sum of fluxes. This created four groups of strains, the ones with the same flux distribution as the reference strain, two other groups that do not produce GE2270A but have slightly different flux distributions, and one group producing GE2270A (*Supplementary Figure 26*). The group producing GE2270A is producing the compound without any constraint on production, only the knockouts forced production. The GE2270A producing strains have multiple metabolic reactions predicted as carrying higher fluxes, most of these reactions are associated to amino-acids biosynthesis and aminoacyl-tRNA charging for GE2270A production (79% of the reactions), the rest are involved in central carbon metabolism needed for *de novo* amino-acids and for ATP regeneration (needed for tRNA-charging and GE2270A production) (*Supplementary Figure 26*). Some reactions were lower for GE2270A producers, such as the pentose phosphate pathway and lactate production, which are due to a higher growth in the non-producers (*Supplementary Figure 26*). The mutants predicted as forcing production of GE2270A all have SCO3295 and SCO6102 knocked out (*Supplementary Figure 25 and 26*). While none of the other knockouts seem to have a significant impact on the metabolism or on GE2270A production.

Finally, the two genes KO responsible for the predicted production of GE2270A (SCO3295 and SCO6102) in the models were further studied to identify their role in the GE2270A production. The SCO3295 gene is associated to a NADPH-dependent FMN reductase and a NADH-dependent FMN reductase, both reactions convert FMN into $FMNH_2$. The SCO6102 gene is also associated to the NADH-dependent FMN reductase,

Genes KO	Reactions associated	Function
SCO6102	FMNRx2, FADRx2, FLVR, SULRi	Cofactor nitrite or sulphite reductase: Reduce NADH or NADPH to generate FMNH2, FADH2, Riboflavin, or H2S
SCO3295	FMNRx, FMNRx2	
SCO1223	ORNTA	Ornithine transaminase: Use alpha-ketoglutarate to transform ornithine in Glutamate
SCO6754	GLYCDx	Glycerol dehydrogenase
SCO5515	PGCD	Phosphoglycerate dehydrogenase
SCO6658	GND*	Phosphogluconate dehydrogenase (redundant gene)

Table 6.3: Most frequently predicted genes knockouts in the final mutants population

as well as to a FAD reductase (NADPH-dependent), a flavin reductase (NADPH-dependent), and a sulfite reductase (NADPH-dependent). So, it appears that the models with these genes KO predict GE2270A production because the production becomes essential for the organism growth by using the GE2270A pathway to produce $FMNH_2$. In a WT strain, the FMN reductases produce the $FMNH_2$ (which is used for growth) by converting a FMN while using an NADPH or NADH cofactor (*Figure 6.11a*). So, in the case of a single SCO6102 deletion this may reduce fluxes producing $FMNH_2$ through the NADPH dependent FMN reductase and force the use of the GE2270A pathway to produce more $FMNH_2$ (*Figure 6.11b*). A single SCO3295 deletion could result in flux decreases through both FMN reductases, hence forcing production of $FMNH_2$ through the GE2270A pathway (*Figure 6.11c*). Finally, a double knockout of SCO3295 and SCO6102 would make the strain completely reliant on GE2270A for $FMNH_2$ production necessary for biomass, so by coupling GE2270A production to biomass the strain would have to produce GE2270A making it an essential by-product (*Figure 6.11d*). The SCO3295 gene has never been studied before in *Streptomyces*, but the SCO6102 (*sirA*) has been deleted from *S. coelicolor* before of its enzyme sulfite

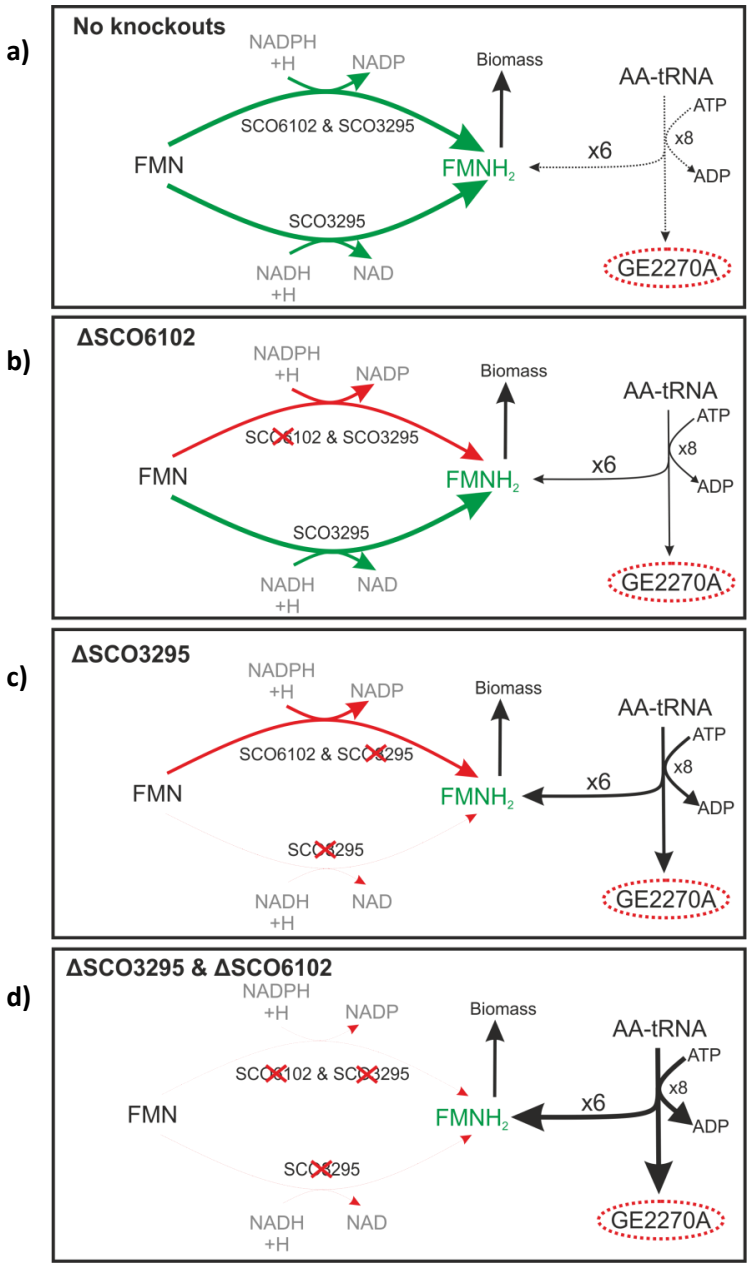


Figure 6.11: Scenarios of the metabolic impact of the SCO3295 and SCO6102 knockouts

a) Predicted state without knockouts

Without knockout FMNH2 is normally produced from a NADPH and/or a NADH FMN reductase. GE2270A is not predicted as produced as it is not necessary metabolically to produce biomass.

b) SCO6102 knockout effect on GE2270A

Knockout of SCO6102 (*SirA*) may reduce the NADPH dependent FMN reductase fluxes but not abolish it. If not enough $FMNH_2$ is produced this may force some GE2270A production.

c) SCO3295 knockout effect on GE2270A

Knockout of SCO3295 would abolish the NADH dependent FMN reductase activity and potentially reduce NADPH dependent FMN reductase fluxes. If $FMNH_2$ is not sufficient for growth this may force GE2270A production.

d) SCO3295 and SCO6102 double knockout effect on GE2270A

The double knockout would abolish production of $FMNH_2$ and force production of GE2270A to continue production of $FMNH_2$ for growth.

reductase activity [84]. The strain with a SCO6102 KO was not able to assimilate sulfate so it needed cysteine or methionine in the media to grow [84]. Here, the cysteine and methionine amino-acids are already present in the CMAN media, hence the ability to grow despite the SCO6102 knockout. Both genes are expressed in *S. coelicolor* M145 growing in minimal media [22] (*Supplementary Figure 27a*) and in R2YE media [85] (*Supplementary Figure 27b*), with SCO6102 showing higher expression levels. So the deletion of these two genes *in vivo* should have an impact on metabolism and flux redistribution without being essential for growth under the conditions used, while potentially forcing production of GE2270A. However, this would mainly work if the FMN is used as a free pool of cofactor by the GE2270A biosynthetic enzymes making it available for the metabolism. But the enzymatic mechanism of the FMN-dependent oxidoreductase from the GE2270A pathway is still unknown. Also, this is based on the metabolic model genes annotations, so if there are other $FMNH_2$ producing enzyme this prediction would be compromised. However, the other genes predicted in the *S. coelicolor* genome as FMN-binding are all associated to reactions consuming $FMNH_2$ for biosynthesis of diverse compounds such as desferrioxamine B or thymidylate (*Supplementary Table 6.7*). So, SCO3102 and SCO3295 are likely the only genes encoding enzymes capable of converting FMN into $FMNH_2$.

Finally, this stepwise method reduced the number of candidate knockouts from 1259 in the model to 2 genes knockouts. A major challenge here was that the GE2270A biosynthesis has a limited dependence over the rest of the metabolism as the main precursors are the aminoacyl-tRNA and the FMN and ATP cofactors needed for production. So the knockouts identified here could help to force the *S. coelicolor* strain to produce the compound by coupling it to biomass production. Other parameters may

be more important for GE2270A production in the *S. coelicolor* host such as gene expression levels, translation efficiency by the ribosome, ribosome availability, and compound toxicity. But identifying a metabolic approach to force production of these compounds would help to push the production levels further.

6.6 Conclusion

Here, we presented the integrated analysis of metabolomics data and metabolic modelling of native and heterologous actinorhodin production in four different *S. coelicolor* strains. The exometabolomics data helped to identify major metabolic pathways activated during ACT production, such as those involved in nucleotide metabolism (*Figure 6.6*, and *Supplementary Figure 6.15*). The modelling and metabolomics data analysis were used to predict 5 potential gene knockouts and 5 gene overexpression targets in the primary metabolism for increased ACT production, for example, bypassing the first part of glycolysis and increasing the pentose phosphate pathway (see *Table 6.2*) [48,62].

The measured cAMP levels correlated with increase in ACT production in the exometabolomics dataset; this metabolite has been reported to be important for antibiotic production and morphology changes in *Streptomyces* [45,46,54], however its role in regulating ACT production is still unclear. Further transcriptomics experiments on the impact of ACT production are ongoing.

Another important point is to experimentally determinate if the increase in nucleotide metabolism is the result of *de novo* production of nucleotides or of RNA/DNA degradation. Also the potential role of purines and pyrimidines metabolism in antibiotics production was highlighted, as this metabolism is still understudied in

Streptomyces species; despite molecules such as ppGpp, GTP, cAMP, or ATP that were shown to be directly or indirectly involved in antibiotic regulation [51].

The heterologous production of the GE2270A thiopeptide antibiotic in *S. coelicolor* M1146 was modelled in a complex industrial media. A stepwise method was used to reduce the potential gene knockouts from 1259 genes in the model to 2 gene knockout targets. The modelling predicted that the double knockout of genes SCO6102 (*sirA*) and SCO3295 (FMN reductase gene) associated to NADH and NADPH dependents FMN reductases would force production of GE2270A in M1146. These knockouts could force GE2270A production from the start of growth by making it essential for the organism to produce the antibiotics, and increasing production of this compound. This is currently being tested *in vivo* to identify the impact of these deletions and determine if the predictions were accurate and forced GE2270A production.

Constraint-based genome-scale metabolic modelling combined with metabolomics showed its versatility as a tool to interpret omics data, and to help identifying engineering targets for native and heterologous production of antibiotics in *Streptomyces* species. Due to the robustness and high adaptability of metabolisms [86,87] it becomes increasingly necessary to apply a combination of engineering targets such as knockouts and overexpression to remodel metabolism [61]. Here, the application of metabolic modelling with omics data helped to highlight these potential targets; now, more experimental data is needed to validate the predictions.

6.7 References

(1) Hopwood, D. A. Streptomyces *in Nature and Medicine : The Antibiotic Makers*; Oxford University Press, 2007.

(2) Watve, M. G.; Tickoo, R.; Jog, M. M.; Bhole, B. D. How Many Antibiotics Are Produced

by the Genus Streptomyces? *Arch. Microbiol.* **2001**, *176* (5), 386–390.
(3) Rokem, J. S.; Lantz, A. E.; Nielsen, J. Systems Biology of Antibiotic Production by Microorganisms. *Nat. Prod. Rep.* **2007**, *24* (6), 1262–1287.
(4) Nathan, C.; Goldberg, F. M. The Profit Problem in Antibiotic R&D. *Nat. Rev. Drug Discov.* **2005**, *4* (11), 887–891.
(5) Medema, M. H.; Breitling, R.; Bovenberg, R.; Takano, E. Exploiting Plug-and-Play Synthetic Biology for Drug Discovery and Production in Microorganisms. *Nat. Rev. Microbiol.* **2011**, *9* (2), 131–137.
(6) Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. Synthetic Biology to Access and Expand Nature's Chemical Diversity. *Nat. Rev. Microbiol.* **2016**, *14* (3), 135–149.
(7) Zhang, M. M.; Wang, Y.; Ang, E. L.; Zhao, H. Engineering Microbial Hosts for Production of Bacterial Natural Products. *Nat. Prod. Rep.* **2016**, *33* (8), 963–987.
(8) Myronovskyi, M.; Luzhetskyy, A. Native and Engineered Promoters in Natural Product Discovery. *Nat. Prod. Rep.* **2016**, *33* (8), 1006–1019.
(9) Tong, Y.; Weber, T.; Lee, S. Y. CRISPR/Cas-Based Genome Engineering in Natural Product Discovery. *Nat. Prod. Rep.* **2019**.
(10) Kosuri, S.; Church, G. M. Large-Scale de Novo DNA Synthesis: Technologies and Applications. *Nat. Methods* **2014**, *11* (5), 499–507.
(11) Phelan, R. M.; Sachs, D.; Petkiewicz, S. J.; Barajas, J. F.; Blake-Hedges, J. M.; Thompson, M. G.; Reider Apel, A.; Rasor, B. J.; Katz, L.; Keasling, J. D. Development of Next Generation Synthetic Biology Tools for Use in *Streptomyces Venezuelae*. *ACS Synth. Biol.* **2017**, *6* (1), 159–166.
(12) Breitling, R.; Takano, E. Synthetic Biology of Natural Products. *Cold Spring Harb. Perspect. Biol.* **2016**, *8* (10), 357–369.
(13) Medema, M. H.; Breitling, R.; Takano, E. Synthetic Biology in Streptomyces Bacteria. *Methods Enzym.* **2011**, *497*, 485–502.
(14) Carbonell, P.; Jervis, A. J.; Robinson, C. J.; Yan, C.; Dunstan, M.; Swainston, N.; Vinaixa, M.; Hollywood, K. A.; Currin, A.; Rattray, N. J. W.; et al. An Automated Design-Build-Test-Learn Pipeline for Enhanced Microbial Production of Fine Chemicals. *Commun. Biol.* **2018**, *1* (1), 66.
(15) Carbonell, P.; Currin, A.; Jervis, A. J.; Rattray, N. J. W.; Swainston, N.; Yan, C.; Takano, E.; Breitling, R. Bioinformatics for the Synthetic Biology of Natural Products: Integrating across the Design–Build–Test Cycle. *Nat. Prod. Rep.* **2016**, *33* (8), 925–932.
(16) Cameron, D. E.; Bashor, C. J.; Collins, J. J. A Brief History of Synthetic Biology. *Nat. Rev. Microbiol.* **2014**, *12*, 381–390.
(17) Amara, A.; Takano, E.; Breitling, R. Development and Validation of an Updated Computational Model of *Streptomyces Coelicolor* Primary and Secondary Metabolism. *BMC Genomics* **2018**, *19* (1), 519.
(18) Mohite, O. S.; Weber, T.; Kim, H. U.; Lee, S. Y. Genome-Scale Metabolic Reconstruction of Actinomycetes for Antibiotics Production. *Biotechnol. J.* **2019**, *14* (1), 1800377.
(19) Bystrykh, L. V; Fernández-Moreno, M. A.; Herrema, J. K.; Malpartida, F.; Hopwood, D. A.; Dijkhuizen, L. Production of Actinorhodin-Related" Blue Pigments" by *Streptomyces Coelicolor* A3 (2). *J. Bacteriol.* **1996**, *178* (8), 2238–2244.
(20) Mak, S.; Nodwell, J. R. Actinorhodin Is a Redox-Active Antibiotic with a Complex Mode of Action against Gram-Positive Cells. *Mol. Microbiol.* **2017**, *106* (4), 597–613.
(21) Doull, J. L.; Vining, L. C. Nutritional Control of Actinorhodin Production by Streptomyces Coelicolor A3(2): Suppressive Effects of Nitrogen and Phosphate. *Appl Microbiol Biotechnol* **1990**, *32* (4), 449–454.
(22) Nieselt, K.; Battke, F.; Herbig, A.; Bruheim, P.; Wentzel, A.; Jakobsen, Ø. M.; Sletta, H.; Alam, M. T.; Merlo, M. E.; Moore, J. The Dynamic Architecture of the Metabolic Switch in *Streptomyces Coelicolor*. *BMC Genomics* **2010**, *11* (1), 10.

(23) Fabbretti, A.; He, C.-G.; Gaspari, E.; Maffioli, S.; Brandi, L.; Spurio, R.; Sosio, M.; Jabes, D.; Donadio, S. A Derivative of the Thiopeptide GE2270A Highly Selective against Propionibacterium Acnes. *Antimicrob. Agents Chemother.* **2015**, *59* (8), 4560–4568.
(24) LaMarche, M. J.; Leeds, J. A.; Amaral, A.; Brewer, J. T.; Bushell, S. M.; Deng, G.; Dewhurst, J. M.; Ding, J.; Dzink-Fox, J.; Gamber, G.; et al. Discovery of LFF571: An Investigational Agent for *Clostridium Difficile* Infection. *J. Med. Chem.* **2012**, *55* (5), 2376–2387.
(25) Flinspach, K.; Kapitzke, C.; Tocchetti, A.; Sosio, M.; Apel, A. K. Heterologous Expression of the Thiopeptide Antibiotic GE2270 from Planobispora Rosea ATCC 53733 in Streptomyces Coelicolor Requires Deletion of Ribosomal Genes from the Expression Construct. *PLoS One* **2014**, *9* (3), e90499.
(26) Gomez-Escribano, J. P.; Bibb, M. J. Engineering Streptomyces Coelicolor for Heterologous Expression of Secondary Metabolite Gene Clusters. *Microb. Biotechnol.* **2011**, *4* (2), 207–215.
(27) Schellenberger, J.; Que, R.; Fleming, R. M. T.; Thiele, I.; Orth, J. D.; Feist, A. M.; Zielinski, D. C.; Bordbar, A.; Lewis, N. E.; Rahmanian, S.; et al. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6* (9), 1290–1307.
(28) Ebrahim, A.; Lerman, J. A.; Palsson, B. Ø.; Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7* (1), 74.
(29) Team, R. C. R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing. 2017.
(30) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat. Biotechnol.* **2010**, *28* (3), 245–248.
(31) Gomez-Escribano, J. P.; Bibb, M. J. Heterologous Expression of Natural Product Biosynthetic Gene Clusters in Streptomyces Coelicolor: From Genome Mining to Manipulation of Biosynthetic Pathways. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 425–431.
(32) Lewis, N. E.; Hixson, K. K.; Conrad, T. M.; Lerman, J. A.; Charusanti, P.; Polpitiya, A. D.; Adkins, J. N.; Schramm, G.; Purvine, S. O.; Lopez-Ferrer, D.; et al. Omic Data from Evolved *E. Coli* Are Consistent with Computed Optimal Growth from Genome-Scale Models. *Mol. Syst. Biol.* **2010**, *6*, 390.
(33) Monk, J.; Nogales, J.; Palsson, B. O. Optimizing Genome-Scale Network Reconstructions. *Nat. Biotechnol.* **2014**, *32* (5), 447–452.
(34) King, Z. A.; Dräger, A.; Ebrahim, A.; Sonnenschein, N.; Lewis, N. E.; Palsson, B. O. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLOS Comput. Biol.* **2015**, *11* (8), e1004321.
(35) Orth, J. D.; Conrad, T. M.; Na, J.; Lerman, J. A.; Nam, H.; Feist, A. M.; Palsson, B. O. A Comprehensive Genome-Scale Reconstruction of Escherichia Coli Metabolism--2011. *Mol. Syst. Biol.* **2014**, *7* (1), 535–535.
(36) King, Z. A.; Lu, J.; Drager, A.; Miller, P.; Federowicz, S.; Lerman, J. A.; Ebrahim, A.; Palsson, B. Ø.; Lewis, N. E.; J., H. BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models. *Nucleic Acids Res.* **2016**, *44* (D1), D515–D522.
(37) Goldstein, Y. A.; Bockmayr, A. Double and Multiple Knockout Simulations for Genome-Scale Metabolic Network Reconstructions. *Algorithms Mol. Biol.* **2015**, *10* (1), 1.
(38) Patil, K.; Rocha, I.; Förster, J.; Nielsen, J. Evolutionary Programming as a Platform for in Silico Metabolic Engineering. *BMC Bioinformatics* **2005**, *6* (1), 308.
(39) Melzoch, K.; De Mattos, M. J. T.; Neijssel, O. M. Production of Actinorhodin by Streptomyces Coelicolor A3(2) Grown in Chemostat Culture. *Biotechnol. Bioeng.* **1997**, *54* (6), 577–582.
(40) Wentzel, A.; Bruheim, P.; Øverby, A.; Jakobsen, Ø. M.; Sletta, H.; Omara, W. A. M.; Hodgson, D. A.; Ellingsen, T. E. Optimized Submerged Batch Fermentation Strategy for

Systems Scale Studies of Metabolic Switching in *Streptomyces Coelicolor* A3(2). *BMC Syst. Biol.* **2012**, *6* (1), 59.
(41) Pinu, F.; Villas-Boas, S. Extracellular Microbial Metabolomics: The State of the Art. *Metabolites* **2017**, *7* (3), 43.
(42) Pinu, F. R.; Granucci, N.; Daniell, J.; Han, T.-L.; Carneiro, S.; Rocha, I.; Nielsen, J.; Villas-Boas, S. G. Metabolite Secretion in Microorganisms: The Theory of Metabolic Overflow Put to the Test. *Metabolomics* **2018**, *14* (4), 43.
(43) Aurich, M. K.; Paglia, G.; Rolfsson, Ó.; Hrafnsdóttir, S.; Magnúsdóttir, M.; Stefaniak, M. M.; Palsson, B. Ø.; Fleming, R. M. T.; Thiele, I. Prediction of Intracellular Metabolic States from Extracellular Metabolomic Data. *Metabolomics* **2015**, *11* (3), 603–619.
(44) Cakir, T.; Efe, C.; Dikicioglu, D.; Hortacsu, A.; Kirdar, B.; Oliver, S. G. Flux Balance Analysis of a Genome-Scale Yeast Model Constrained by Exometabolomic Data Allows Metabolic System Identification of Genetically Different Strains. *Biotechnol Prog* **2007**, *23* (2), 320–326.
(45) Süsstrunk, U.; Pidoux, J.; Taubert, S.; Ullmann, A.; Thompson, C. J. Pleiotropic Effects of CAMP on Germination, Antibiotic Biosynthesis and Morphological Development in Streptomyces Coelicolor. *Mol. Microbiol.* **1998**, *30* (1), 33–46.
(46) Gao, C.; Hindra; Mulder, D.; Yin, C.; Elliot, M. A. Crp Is a Global Regulator of Antibiotic Production in *Streptomyces*. *MBio* **2012**, *3* (6), e00407-12.
(47) Shimada, T.; Fujita, N.; Yamamoto, K.; Ishihama, A. Novel Roles of CAMP Receptor Protein (CRP) in Regulation of Transport and Metabolism of Carbon Sources. *PLoS One* **2011**, *6* (6), e20081.
(48) Butler, M. J.; Bruheim, P.; Jovetic, S.; Marinelli, F.; Postma, P. W.; Bibb, M. J. Engineering of Primary Carbon Metabolism for Improved Antibiotic Production in *Streptomyces Lividans*. *Appl. Environ. Microbiol.* **2002**, *68* (10), 4731 LP – 4739.
(49) Kilstrup, M.; Hammer, K.; Ruhdal Jensen, P.; Martinussen, J. Nucleotide Metabolism and Its Control in Lactic Acid Bacteria. *FEMS Microbiol. Rev.* **2005**, *29* (3), 555–590.
(50) Takano, E. PpGpp and Antibiotic Production in Streptomyces Coelicolor A3(2). **1993**.
(51) Sivapragasam, S.; Grove, A.; Sivapragasam, S.; Grove, A. The Link between Purine Metabolism and Production of Antibiotics in Streptomyces. *Antibiotics* **2019**, *8* (2), 76.
(52) Hesketh, A.; Chen, W.; Ryding, J.; Chang, S.; Bibb, M. The Global Role of PpGpp Synthesis in Morphological Differentiation and Antibiotic Production in Streptomyces Coelicolor A3(2). *Genome Biol.* **2007**, *8* (8), R161.
(53) Saito, N.; Xu, J.; Hosaka, T.; Okamoto, S.; Aoki, H.; Bibb, M. J.; Ochi, K. EshA Accentuates PpGpp Accumulation and Is Conditionally Required for Antibiotic Production in Streptomyces Coelicolor A3(2). *J. Bacteriol.* **2006**, *188* (13), 4952–4961.
(54) Kang, D.-K.; Li, X.-M.; Ochi, K.; Horinouchi, S. Possible Involvement of CAMP in Aerial Mycelium Formation and Secondary Metabolism in Streptomyces Griseus. *Microbiology* **1999**, *145* (5), 1161–1172.
(55) Hockenhull, D. J. D.; Fantes, K. H.; Herbert, M.; Whitehead, B. Glucose Utilization by Streptomyces Griseus. *J. Gen. Microbiol.* **1954**, *10* (3), 353–370.
(56) Borodina, I.; Krabben, P.; Nielsen, J. Genome-Scale Analysis of Streptomyces Coelicolor A3(2) Metabolism. *Genome Res.* **2005**, *15* (6), 820–829.
(57) Gamboa-Suasnavart, R. A.; Valdez-Cruz, N. A.; Gaytan-Ortega, G.; Reynoso-Cereceda, G. I.; Cabrera-Santos, D.; López-Griego, L.; Klöckner, W.; Büchs, J.; Trujillo-Roldán, M. A. The Metabolic Switch Can Be Activated in a Recombinant Strain of Streptomyces Lividans by a Low Oxygen Transfer Rate in Shake Flasks. *Microb. Cell Fact.* **2018**, *17* (1), 189.
(58) van Keulen, G.; Jonkers, H. M.; Claessen, D.; Dijkhuizen, L.; Wösten, H. A. B. Differentiation and Anaerobiosis in Standing Liquid Cultures of Streptomyces Coelicolor. *J. Bacteriol.* **2003**, *185* (4), 1455–1458.
(59) Ryu, Y.-G.; Butler, M. J.; Chater, K. F.; Lee, K. J. Engineering of Primary Carbohydrate

Metabolism for Increased Production of Actinorhodin in *Streptomyces Coelicolor*. *Appl. Environ. Microbiol.* **2006**, *72* (11), 7132–7139.

(60) Viollier, P. H.; Minas, W.; Dale, G. E.; Folcher, M.; Thompson, C. J. Role of Acid Metabolism in *Streptomyces Coelicolor* Morphological Differentiation and Antibiotic Biosynthesis. *J. Bacteriol.* **2001**, *183* (10), 3184 LP – 3192.

(61) Aslan, S.; Noor, E.; Bar-Even, A. Holistic Bioengineering: Rewiring Central Metabolism for Enhanced Bioproduction. *Biochem. J.* **2017**, *474* (23), 3935 LP – 3950.

(62) Borodina, I.; Siebring, J.; Zhang, J.; Smith, C. P.; van Keulen, G.; Dijkhuizen, L.; Nielsen, J. Antibiotic Overproduction in Streptomyces Coelicolor A3 (2) Mediated by Phosphofructokinase Deletion. *J. Biol. Chem.* **2008**, *283* (37), 25186–25199.

(63) Obanye, A. I. C.; Hobbs, G.; Gardner, D. C. J.; Oliver, S. G. Correlation between Carbon Flux through the Pentose Phosphate Pathway and Production of the Antibiotic Methylenomycin in Streptomyces Coelicolor A3(2). *Microbiology* **1996**, *142* (1), 133–137.

(64) Jin, X.-M.; Chang, Y.-K.; Lee, J. H.; Hong, S.-K. Effects of Increased NADPH Concentration by Metabolic Engineering of the Pentose Phosphate Pathway on Antibiotic Production and Sporulation in Streptomyces Lividans TK24. *J Microbiol Biotechnol* **2017**, *27*, 1867–1876.

(65) Coze, F.; Gilard, F.; Tcherkez, G.; Virolle, M.-J.; Guyonvarch, A. Carbon-Flux Distribution within Streptomyces Coelicolor Metabolism: A Comparison between the Actinorhodin-Producing Strain M145 and Its Non-Producing Derivative M1146. *PLoS One* **2013**, *8* (12), e84151.

(66) Romero-Rodríguez, A.; Rocha, D.; Ruiz-Villafan, B.; Tierrafría, V.; Rodríguez-Sanoja, R.; Segura-González, D.; Sánchez, S. Transcriptomic Analysis of a Classical Model of Carbon Catabolite Regulation in Streptomyces Coelicolor. *BMC Microbiol.* **2016**, *16* (1), 77.

(67) Gubbens, J.; Janus, M.; Florea, B. I.; Overkleeft, H. S.; van Wezel, G. P. Identification of Glucose Kinase-Dependent and -Independent Pathways for Carbon Control of Primary Metabolism, Development and Antibiotic Production in *Streptomyces Coelicolor* by Quantitative Proteomics. *Mol. Microbiol.* **2012**, *86* (6), 1490–1507.

(68) Sauer, U.; Eikmanns, B. J. The PEP—Pyruvate—Oxaloacetate Node as the Switch Point for Carbon Flux Distribution in Bacteria. *FEMS Microbiol. Rev.* **2005**, *29* (4), 765–794.

(69) Rodriguez, E.; Navone, L.; Casati, P.; Gramajo, H. Impact of Malic Enzymes on Antibiotic and Triacylglycerol Production in *Streptomyces Coelicolor*. *Appl. Environ. Microbiol.* **2012**, *78* (13), 4571 LP – 4579.

(70) Delbaere, L.; Sudom, A.; Prasad, L.; … Y. L.-C. and M.; 2003, undefined. Structure/Function Studies of the Phosphoryl-Transfer Enzyme: Phosphoenolpyruvate Carboxykinase. *Citeseer*.

(71) Lule, I.; Maldonado, B.; D'Huys, P.-J.; Van Mellaert, L.; Van Impe, J.; Bernaerts, K.; Anné, J. On the Influence of Overexpression of Phosphoenolpyruvate Carboxykinase in Streptomyces Lividans on Growth and Production of Human Tumour Necrosis Factor-Alpha. *Appl. Microbiol. Biotechnol.* **2012**, *96* (2), 367–372.

(72) Zamboni, N.; Maaheimo, H.; … T. S.-M.; 2004, undefined. The Phosphoenolpyruvate Carboxykinase Also Catalyzes C3 Carboxylation at the Interface of Glycolysis and the TCA Cycle of Bacillus Subtilis. *Elsevier*.

(73) Viollier, P. H.; Nguyen, K. T.; Minas, W.; Folcher, M.; Dale, G. E.; Thompson, C. J. Roles of Aconitase in Growth, Metabolism, and Morphological Differentiation of *Streptomyces Coelicolor*. *J. Bacteriol.* **2001**, *183* (10), 3193 LP – 3203.

(74) Ahn, S.; Jung, J.; Jang, I.-A.; Madsen, E. L.; Park, W. Role of Glyoxylate Shunt in Oxidative Stress Response. *J. Biol. Chem.* **2016**, *291* (22), 11928–11938.

(75) Kornberg, H. L. The Role and Control of the Glyoxylate Cycle in Escherichia Coli. *Biochem. J.* **1966**, *99* (1), 1–11.

(76) Esnault, C.; Dulermo, T.; Smirnov, A.; Askora, A.; David, M.; Deniset-Besseau, A.;

Holland, I.-B.; Virolle, M.-J. Strong Antibiotic Production Is Correlated with Highly Active Oxidative Metabolism in *Streptomyces Coelicolor* M145. *Sci. Rep.* **2017**, *7* (1), 200.

(77) Schniete, J. K.; Cruz-Morales, P.; Selem-Mojica, N.; Fernández-Martínez, L. T.; Hunter, I. S.; Barona-Gómez, F.; Hoskisson, P. A. Expanding Primary Metabolism Helps Generate the Metabolic Robustness To Facilitate Antibiotic Biosynthesis in <Em>Streptomyces</Em> *MBio* **2018**, *9* (1), e02283-17.

(78) Hove-Jensen, B.; Rosenkrantz, T. J.; Haldimann, A.; Wanner, B. L. Escherichia Coli PhnN, Encoding Ribose 1,5-Bisphosphokinase Activity (Phosphoribosyl Diphosphate Forming): Dual Role in Phosphonate Degradation and NAD Biosynthesis Pathways. *J. Bacteriol.* **2003**, *185* (9), 2793–2801.

(79) Bramwell, H.; Nimmo, H. G.; Hunter, I. S.; Coggins, J. R. Phosphoenolpyruvate Carboxylase from Streptomyces Coelicolor A3(2): Purification of the Enzyme, Cloning of the Ppc Gene and over-Expression of the Protein in a Streptomycete. *Biochem. J.* **1993**, *293 (Pt 1)* (1), 131–136.

(80) Lewis, R. A.; Laing, E.; Allenby, N.; Bucca, G.; Brenner, V.; Harrison, M.; Kierzek, A. M.; Smith, C. P. Metabolic and Evolutionary Insights into the Closely-Related Species *Streptomyces Coelicolor* and *Streptomyces Lividans* Deduced from High-Resolution Comparative Genomic Hybridization. *BMC Genomics* **2010**, *11* (1), 682.

(81) Park, S.-Y.; Moon, M.-W.; Subhadra, B.; Lee, J.-K. Functional Characterization of the *GlxR* Deletion Mutant of *Corynebacterium Glutamicum* ATCC 13032: Involvement of GlxR in Acetate Metabolism and Carbon Catabolite Repression. *FEMS Microbiol. Lett.* **2010**, *304* (2), 107–115.

(82) Gruchattka, E.; Kayser, O. In Vivo Validation of In Silico Predicted Metabolic Engineering Strategies in Yeast: Disruption of α-Ketoglutarate Dehydrogenase and Expression of ATP-Citrate Lyase for Terpenoid Production. *PLoS One* **2015**, *10* (12), e0144981.

(83) Hudson, G. A.; Zhang, Z.; Tietz, J. I.; Mitchell, D. A.; van der Donk, W. A. In Vitro Biosynthesis of the Core Scaffold of the Thiopeptide Thiomuracin. *J. Am. Chem. Soc.* **2015**, *137* (51), 16012–16015.

(84) Fischer, M.; Schmidt, C.; Falke, D.; Sawers, R. G. Terminal Reduction Reactions of Nitrate and Sulfate Assimilation in Streptomyces Coelicolor A3(2): Identification of Genes Encoding Nitrite and Sulfite Reductases. *Res. Microbiol.* **2012**, *163* (5), 340–348.

(85) Jeong, Y.; Kim, J.-N.; Kim, M. W.; Bucca, G.; Cho, S.; Yoon, Y. J.; Kim, B.-G.; Roe, J.-H.; Kim, S. C.; Smith, C. P.; et al. The Dynamic Transcriptional and Translational Landscape of the Model Antibiotic Producer Streptomyces Coelicolor A3(2). *Nat. Commun.* **2016**, *7* (1), 11605.

(86) Nielsen, J.; Keasling, J. D. Engineering Cellular Metabolism. *Cell* **2016**, *164* (6), 1185–1197.

(87) Stelling, J.; Sauer, U.; Szallasi, Z.; Doyle, F. J.; Doyle, J. Robustness of Cellular Functions. *Cell* **2004**, *118* (6), 675–685.

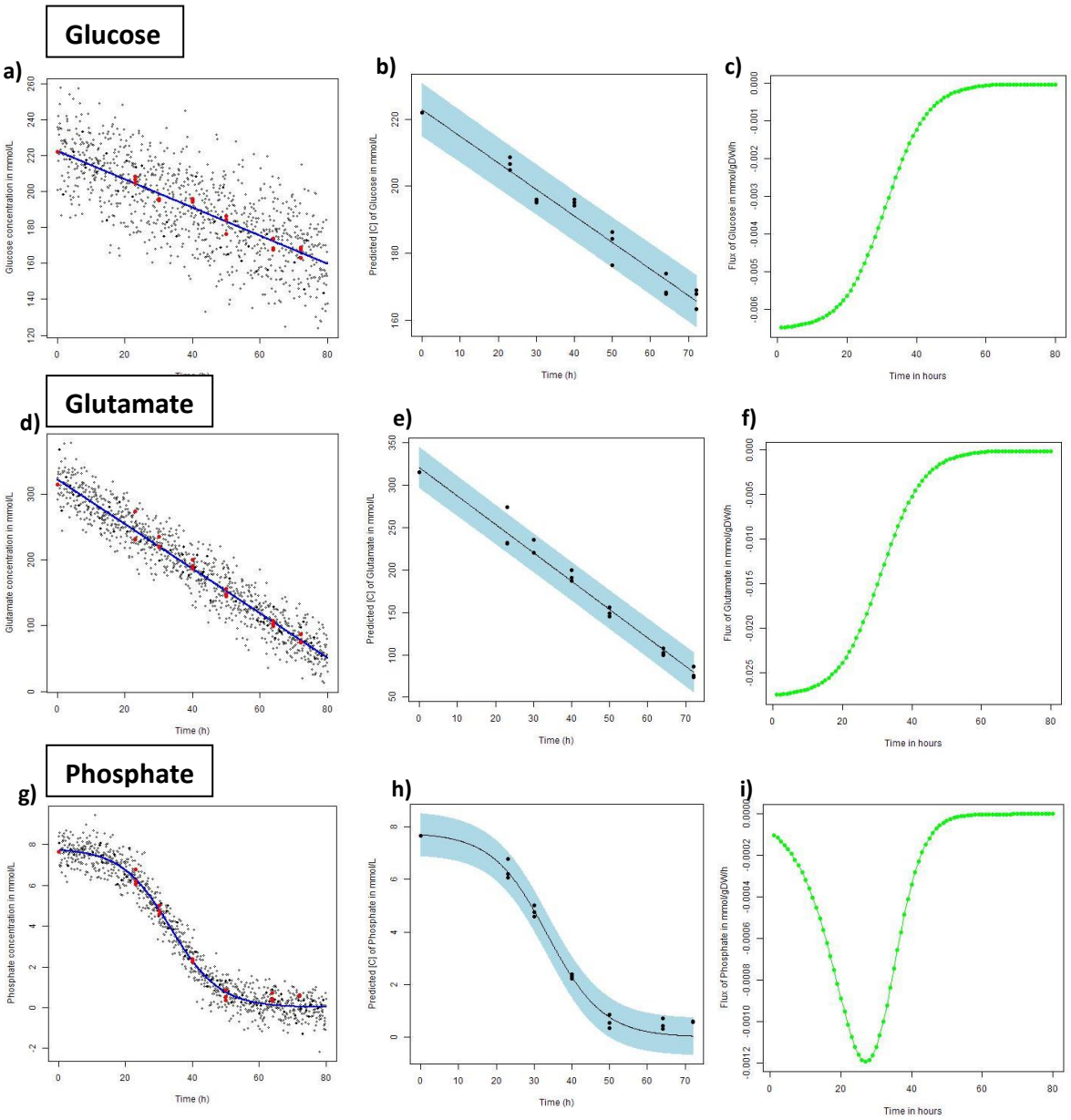
6.8 Supplementary Files

7 Metabolites present in the CMAN media (uptake flux constraint)
 # Glucose 10g/L (55.51 mmol/L) **Carbon source**
Glucose (-55.55)
 # Soluble potato starch 35g/L **Carbon source: mixtures of glucan and malto-polymers**
1,4-alpha-D-glucan (-1000)
Maltohexose (-1000)
Maltopentose (-1000)
Maltotetraose (-1000)
Maltothriose (-1000)
Maltose (-1000)
#Hydrolysed casein 5g/L **Nitrogen source: amino acids from casein**
Alanine (-1000)
Arginine (-1000)
Asparagine (-1000)
Cysteine (-10)
Glutamine (-1000)
Glutamate (-1000)
Glycine (-1000)
Histidine (-1000)
Isoleucine (-1000)
Leucine (-1000)
Lysine (-1000)
Methionine (-1000)
Phenylalanine (-1000)
Proline (-1000)
Serine (-1000)
Threonine (-1000)
Tryptophan (-10)
Tyrosine (-1000)
Valine (-1000)
#Yeast extract 8g/L **Vitamin source and ions**
Panthotenate (-1000)
K^+ (-1000)
Mg^{2+} (-1000)
Na^+ (-1000)
Fe^{3+} (-1000)
Fe^{2+} (-1000)
Cu^{2+} (-1000)
Cl^- (-1000)
Mn^{2+} (-1000)

Choline (-1000)
Inositol (-1000)
Biotine (-1000)
Thiamine (-1000) = Vitamin B1
Pyridoxine (-1000) = Vitamin B6
Phosphate (-1000)
#Calcium carbonate 2g/L **Calcium source**
Ca^{2+} (-1000)

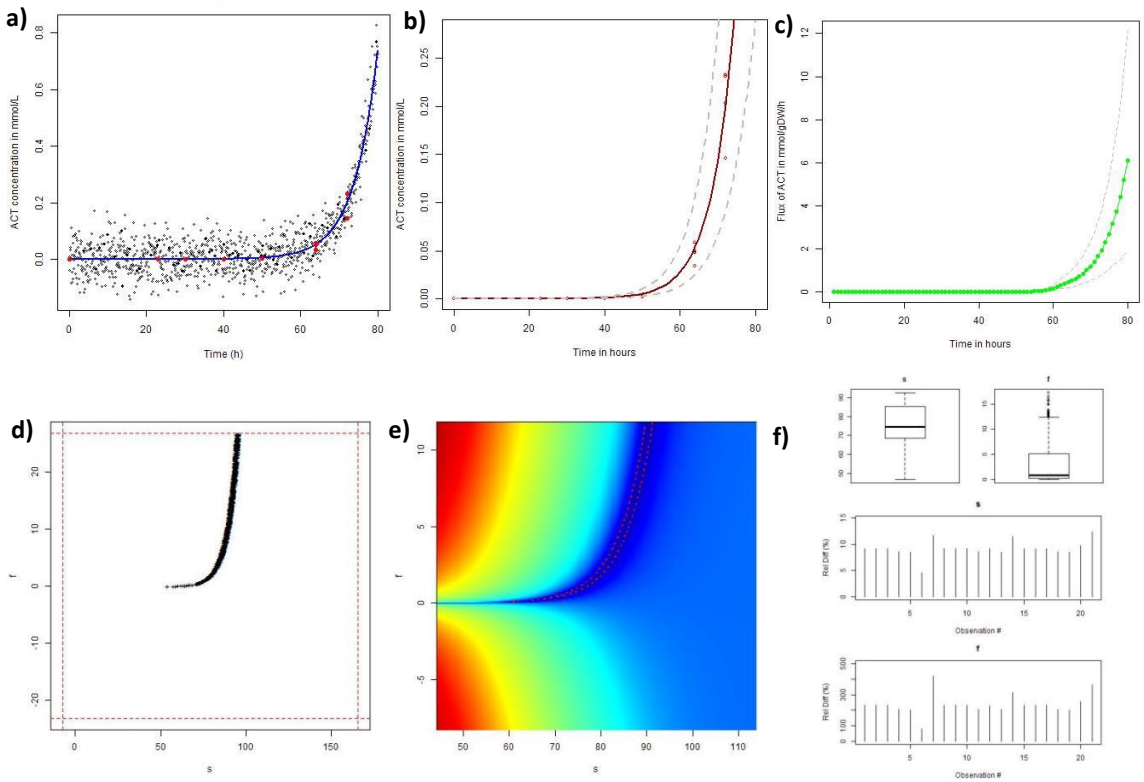
Supplementary Table 6.1: Complex media CMAN composition

Flux tool analysis for M145



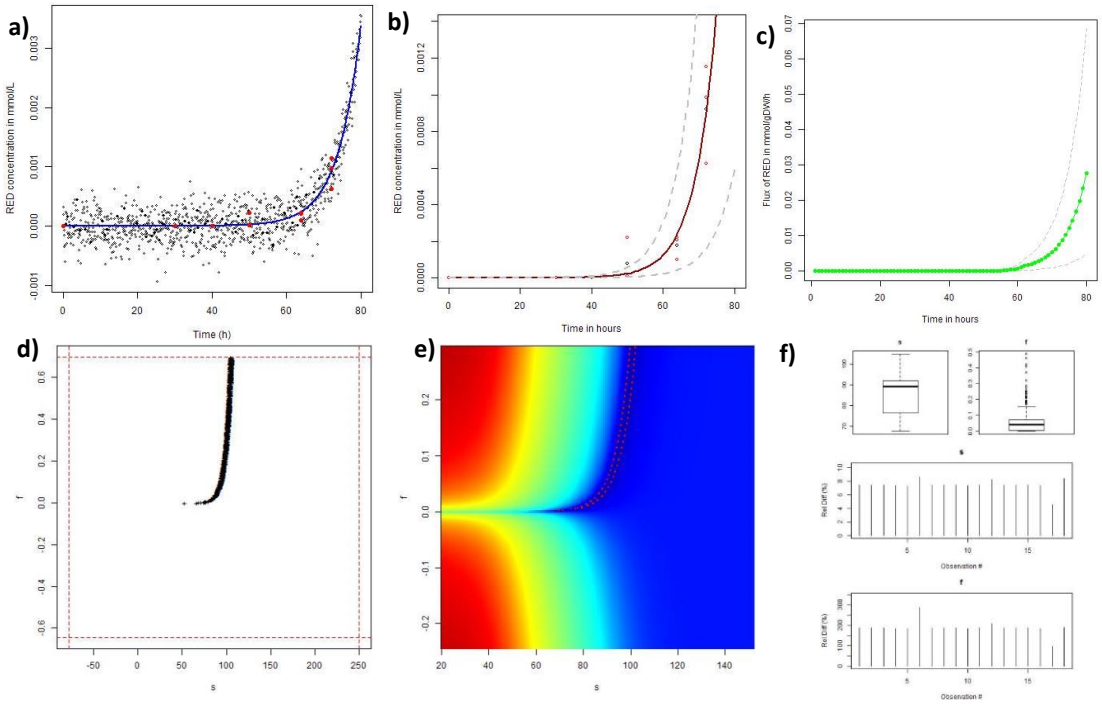
Supplementary Figure 6.1 Flux tool analysis for glucose, glutamate, and phosphate uptake in M145

a) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glucose)
b) Fitting of the equation to the data with the 95% CI band for the glucose data
c) Uptake flux values across times for the glucose
d) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glutamate)
e) Fitting of the equation to the data with the 95% CI band for the glutamate data
f) Uptake flux values across times for the glutamate
g) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (phosphate)
h) Fitting of the equation to the data with the 95% CI band for the phosphate data
i) Uptake flux values across times for the phosphate



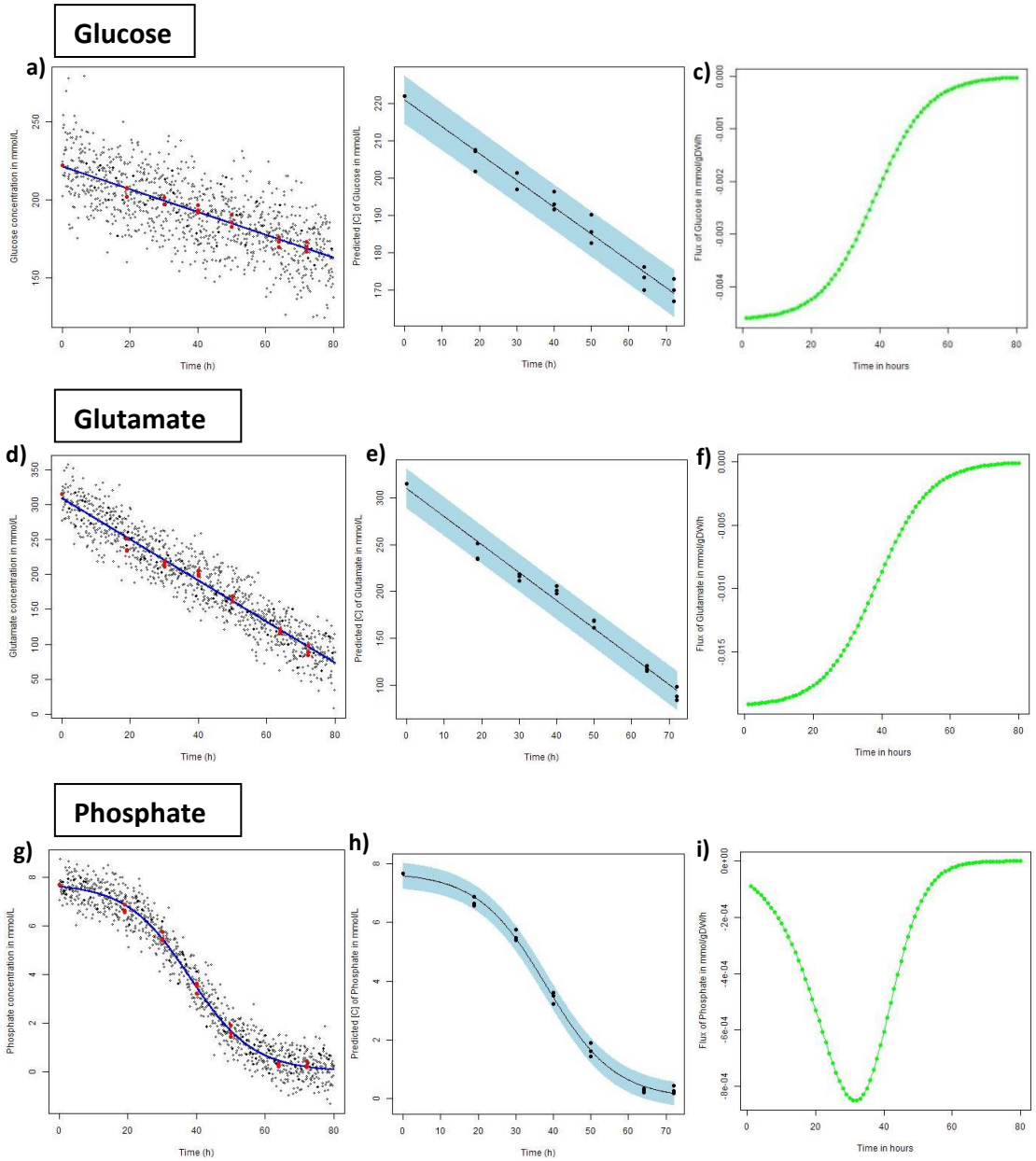
Supplementary Figure 6.2: Flux tool analysis for Actinorhodin secretion in M145

a) Fitting of the equation to the data with noisy simulated data around a sigma=0.1
b) Fitted equation with upper and lower values of the 95% CI for the predicted concentration.
c) Absolute actinorhodin export flux over time.
d) Couple parameters Beale's 95% unlinearized confidence region.
e) Couple parameters RSS contour with 95% CI. Heatmap colours corresponds to the couple of parameters from the least fitting in red to the most fitting in dark blue.
f) Residuals analysis for the parameters estimation during equation fitting.



Supplementary Figure 6.3: Flux tool analysis for Undecylprodigiosin secretion in M145

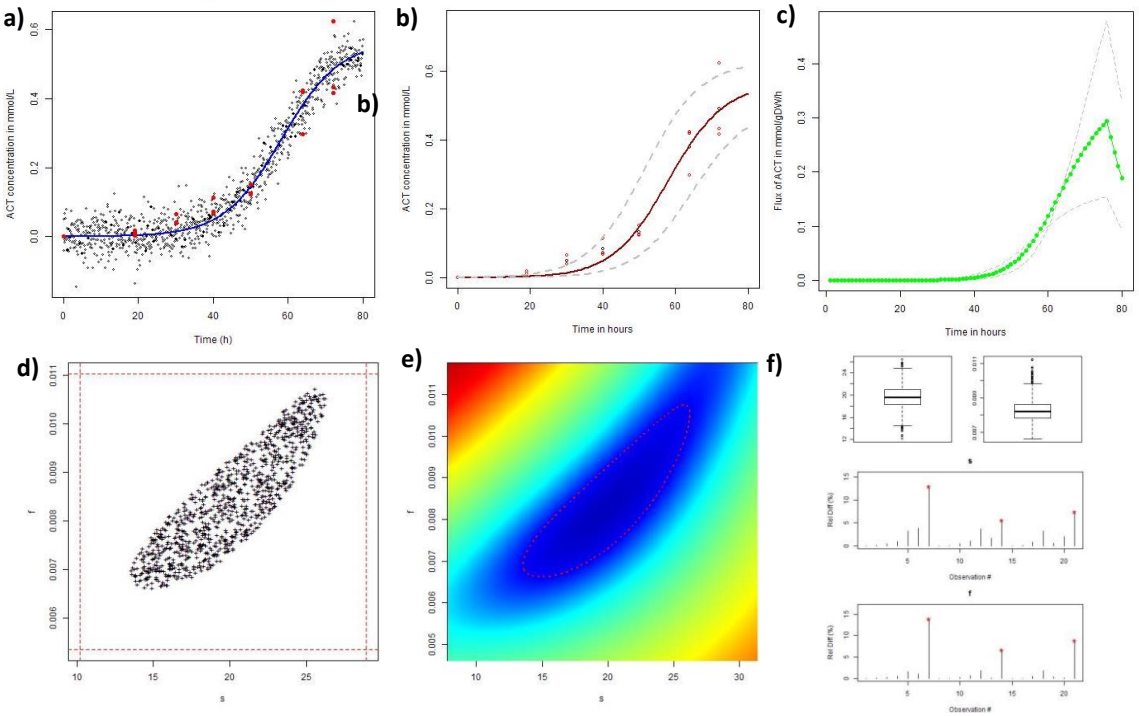
a) Fitting of the equation to the data with noisy simulated data around a sigma=0.1
b) Fitted equation with upper and lower values of the 95% CI for the predicted concentration.
c) Absolute undecylprodigiosin export flux over time.
d) Couple parameters Beale's 95% unlinearized confidence region.
e) Couple parameters RSS contour with 95% CI. Heatmap colours corresponds to the couple of parameters from the least fitting in red to the most fitting in dark blue.
f) Residuals analysis for the parameters estimation during equation fitting.



Supplementary Figure 6.4: Flux tool analysis for glucose, glutamate, and phosphate uptake in M145+ACT

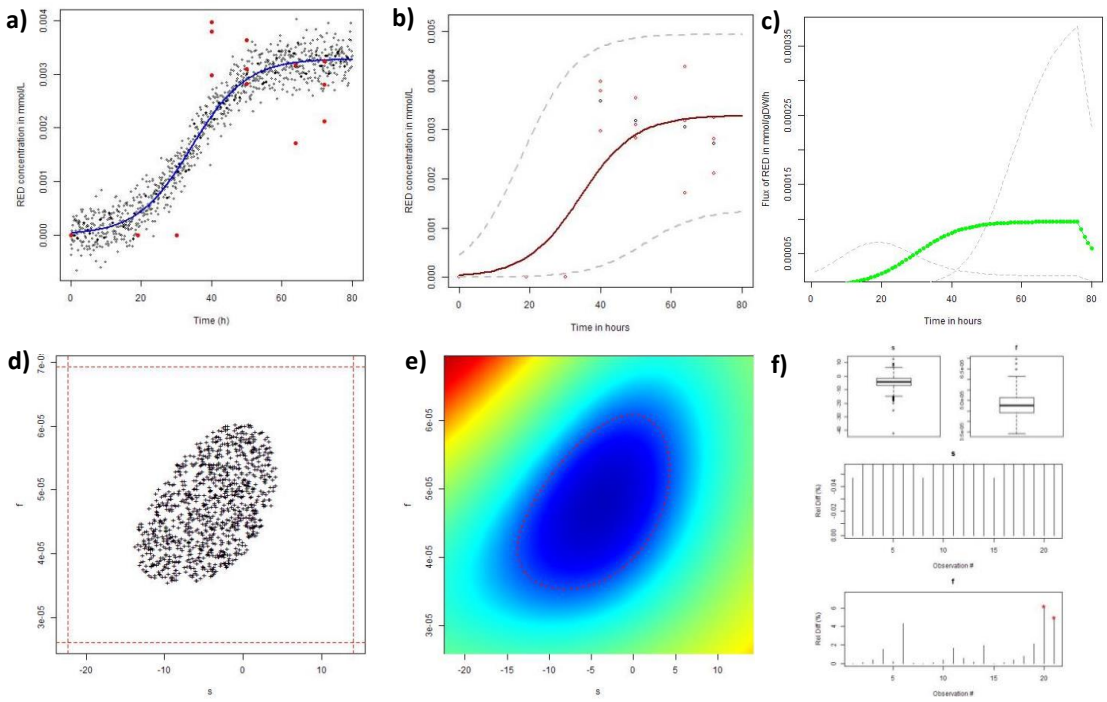
a) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glucose)
b) Fitting of the equation to the data with the 95% CI band for the glucose data
c) Uptake flux values across times for the glucose
d) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glutamate)
e) Fitting of the equation to the data with the 95% CI band for the glutamate data
f) Uptake flux values across times for the glutamate
g) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (phosphate)
h) Fitting of the equation to the data with the 95% CI band for the phosphate data
i) Uptake flux values across times for the phosphate

Flux tool analysis for M145 + ACT



Supplementary Figure 6.5: Flux tool analysis for Actinorhodin secretion in M145 + ACT

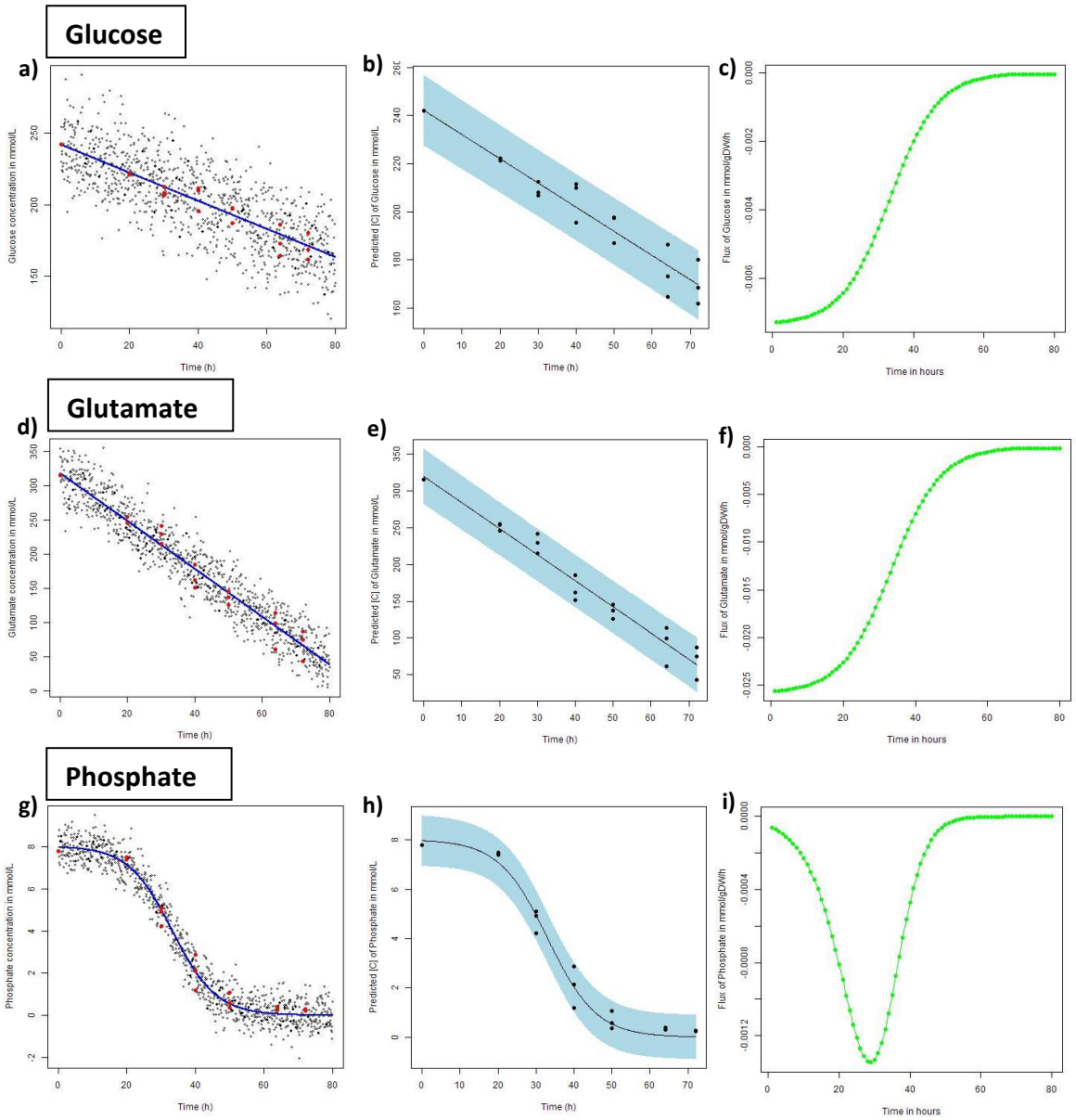
a) Fitting of the equation to the data with noisy simulated data around a sigma=0.1
b) Fitted equation with upper and lower values of the 95% CI for the predicted concentration
c) Absolute actinorhodin export flux over time.
d) Couple parameters Beale's 95% unlinearized confidence region.
g) Couple parameters RSS contour with 95% CI. Heatmap colours corresponds to the couple of parameters from the least fitting in red to the most fitting in dark blue.
e) Residuals analysis for the parameters estimation during equation fitting.



Supplementary Figure 6.6: Flux tool analysis for Undecylprodigiosin secretion in M145 + ACT

a) Fitting of the equation to the data with noisy simulated data around a sigma=0.1
b) Fitted equation with upper and lower values of the 95% CI for the predicted concentration.
c) Absolute undecylprodigiosin export flux over time.
d) Couple parameters Beale's 95% unlinearized confidence region.
e) Couple parameters RSS contour with 95% CI. Heatmap colours corresponds to the couple of parameters from the least fitting in red to the most fitting in dark blue.
f) Residuals analysis for the parameters estimation during equation fitting.

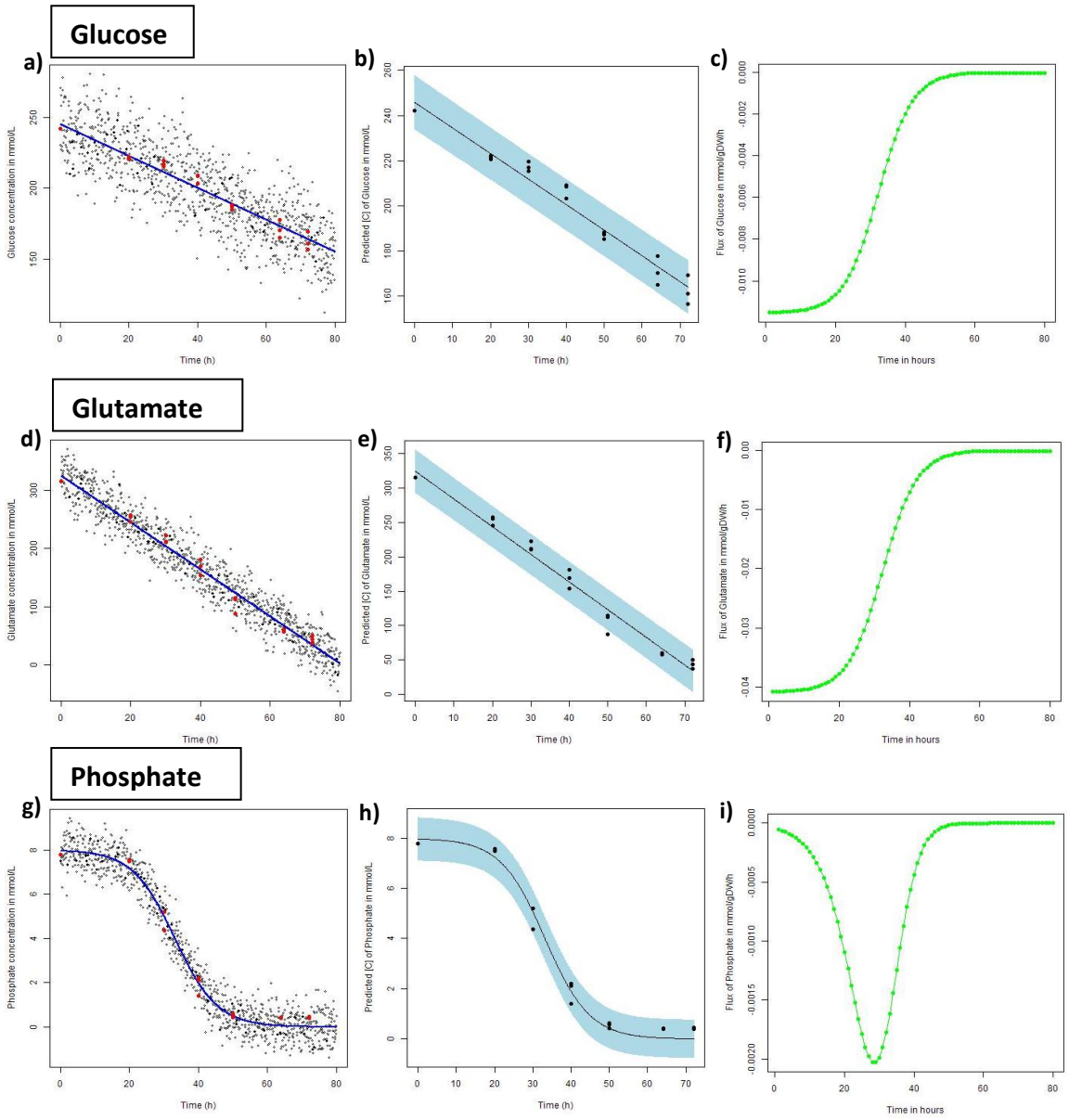
Flux tool analysis for M1146



Supplementary Figure 6.7: Flux tool analysis for glucose, glutamate, and phosphate uptake in M1146

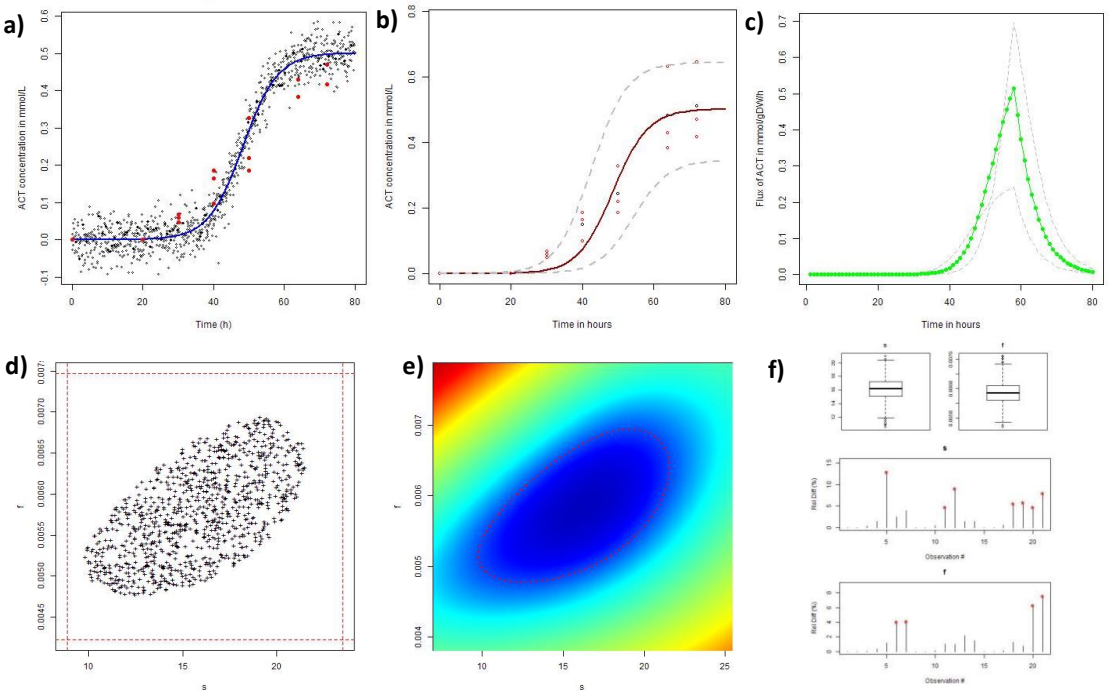
a) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glucose)
b) Fitting of the equation to the data with the 95% CI band for the glucose data
c) Uptake flux values across times for the glucose
d) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glutamate)
e) Fitting of the equation to the data with the 95% CI band for the glutamate data
f) Uptake flux values across times for the glutamate
g) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (phosphate)
h) Fitting of the equation to the data with the 95% CI band for the phosphate data
i) Uptake flux values across times for the phosphate

Flux tool analysis for M1146 + ACT



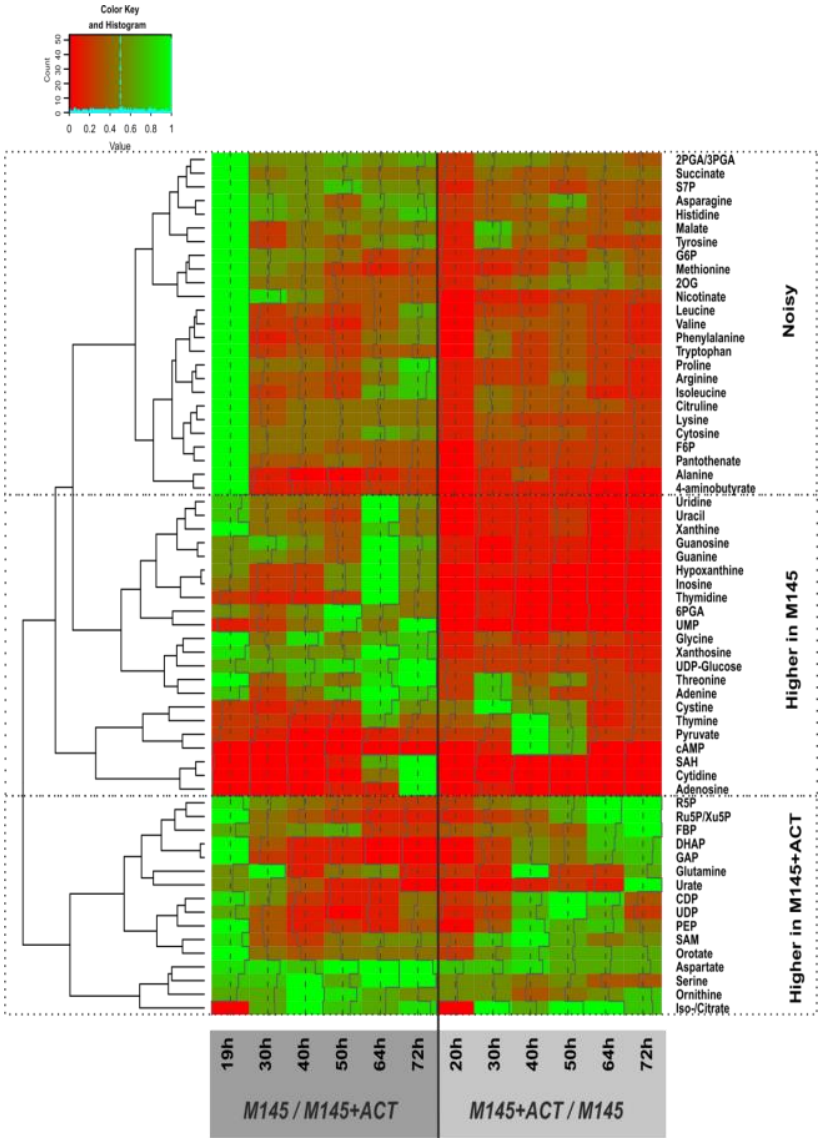
Supplementary Figure 6.8: Flux tool analysis for glucose, glutamate, and phosphate uptake in M1146+ACT

a) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glucose)
b) Fitting of the equation to the data with the 95% CI band for the glucose data
c) Uptake flux values across times for the glucose
d) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (glutamate)
e) Fitting of the equation to the data with the 95% CI band for the glutamate data
f) Uptake flux values across times for the glutamate
g) Fitting of the equation to the data with noisy simulated data with a sigma=0.1 (phosphate)
h) Fitting of the equation to the data with the 95% CI band for the phosphate data
i) Uptake flux values across times for the phosphate



Supplementary Figure 6.9: Flux tool analysis for Actinorhodin secretion in M1146 + ACT

a) Fitting of the equation to the data with noisy simulated data around a sigma=0.1
b) Fitted equation with upper and lower values of the 95% CI for the predicted concentration.
c) Absolute actinorhodin export flux over time.
d) Couple parameters Beale's 95% unlinearized confidence region.
e) Couple parameters RSS contour with 95% CI. Heatmap colours corresponds to the couple of parameters from the least fitting in red to the most fitting in dark blue.
f) Residuals analysis for the parameters estimation during equation fitting.



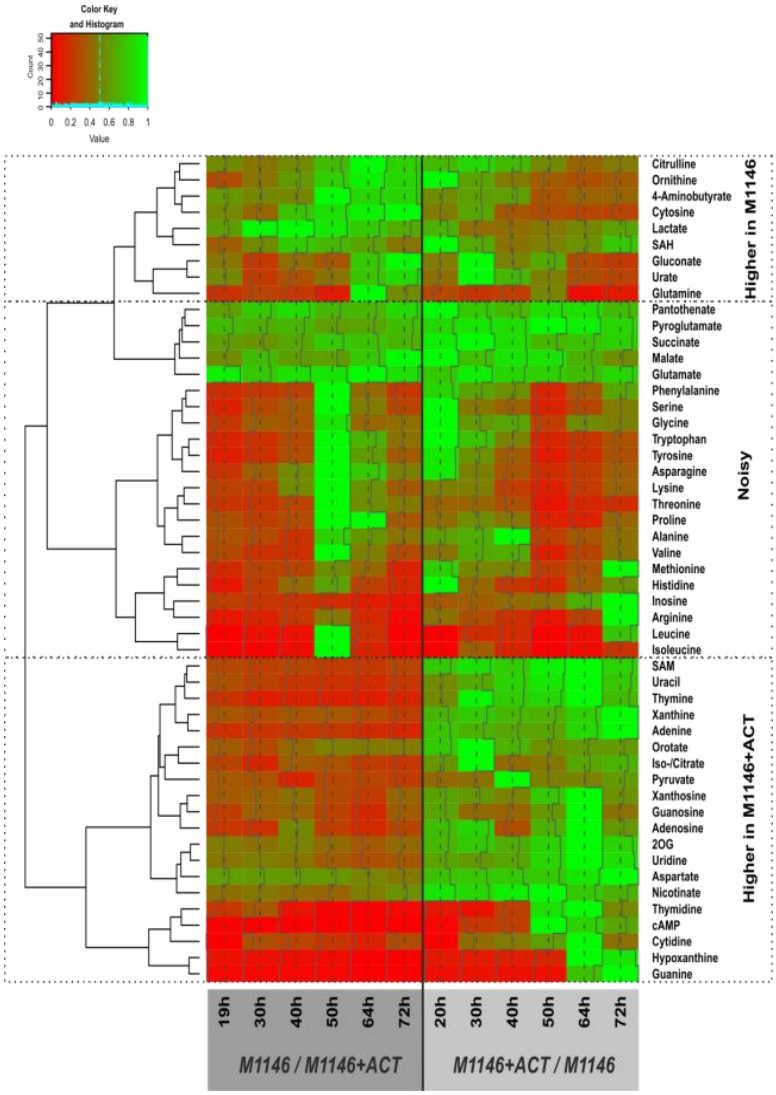
Supplementary Figure 6.10: Exometabolome data analysis M145 and M145+ACT

Metabolome measured in the culture media at 6 different time points (19/20, 30, 40, 50, 64, and 72h) for the four strains. The ratios of metabolites relative quantities are calculated and plotted as heatmaps where the relative high values are in green and low values are in red. The heatmap colour code corresponds to lower relative ion count in red and in green for the higher relative ion count.

Exometabolome of M145 and M145+ACT

The ratios of normalised ion counts for the metabolites measured are calculated as M145/M145+ACT and M145+ACT/M145. The metabolites are clustered in 3 groups, the ones with higher levels in the strains with the ACT cosmid and the ones with higher levels with the strain without the cosmid.

The cAMP, pyruvate, thymine, and cystine are not clustered in the right group as these are clearly higher in the M145+ACT. This clustering issue seems to be caused by the different trend of these metabolites.

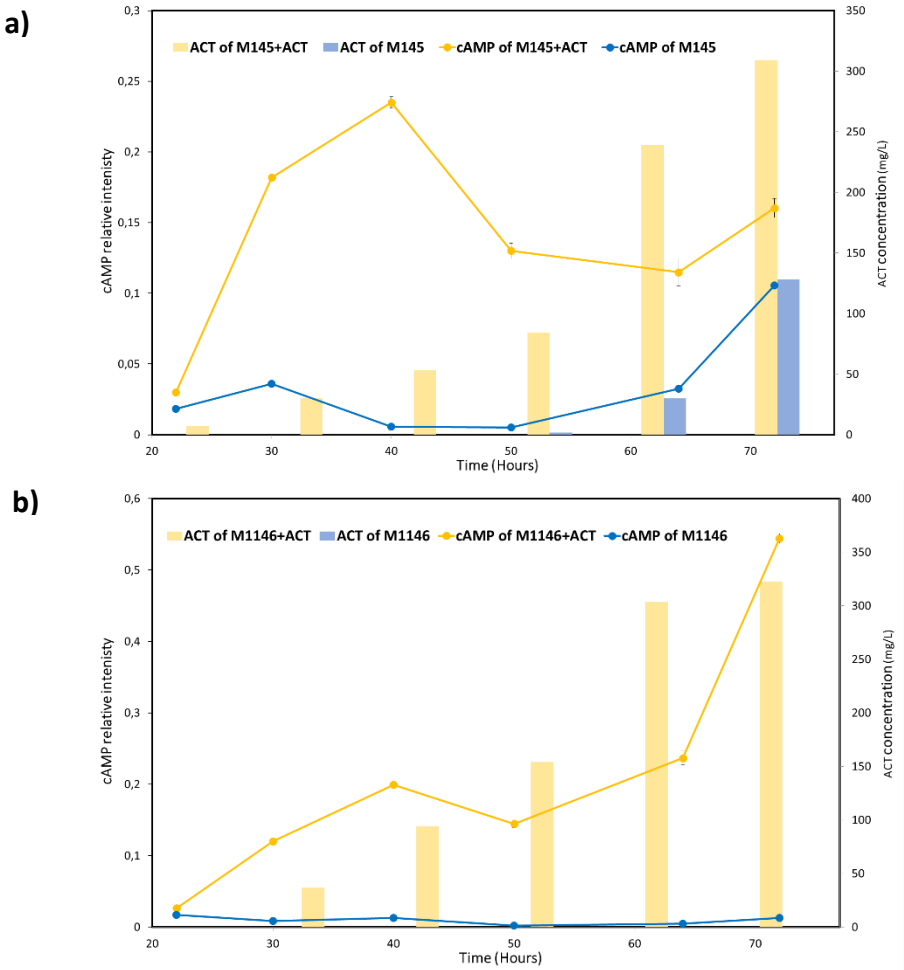


Supplementary Figure 6.11: Exometabolome data analysis for M1146 and M1146+ACT

Metabolome measured in the culture media at 6 different time points (19/20, 30, 40, 50, 64, and 72h) for the four strains. The ratios of metabolites relative quantities are calculated and plotted as heatmaps where the relative high values are in green and low values are in red.

Exometabolome of M1146 and M1146+ACT

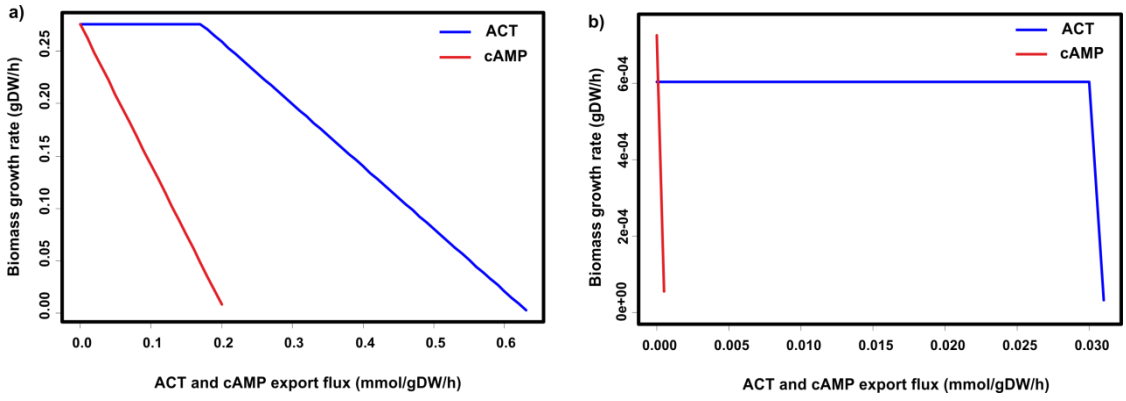
The ratios of normalised ion counts for the metabolites measured are calculated as M1146/M1146+ACT and M1146+ACT/M1146. The metabolites are clustered in 3 groups, the ones with higher levels in the strains with the ACT cosmid and the ones with higher levels with the strain without the cosmid.



Supplementary Figure 6.12: Extracellular cAMP relative intensity in all the strains

a) Extracellular levels of cAMP in M145 and M145+ACT

b) Extracellular levels of cAMP in M1146 and M1146+ACT



Supplementary Figure 6.13: Metabolic trade-off of cAMP compared to ACT

a) Exponential phase: ACT and cAMP secretion fluxes competition with biomass growth at 32 hours

b) Stationary phase: ACT and cAMP secretion fluxes competition with biomass growth at 64 hours

a)

Reaction name	Subsystem	M145	M145+Act	M1146	M1146+Act	Correlation ACT	
S. coelicolor biomass - objective function						M145	M1146
Actinorhodin Pathway Summary	Actinorhodin Biosynthesis					1.00	1.00
pyruvate kinase	Glycolysis and Gluconeogenesis					0.95	-0.21
phosphoglycerate mutase	Glycolysis and Gluconeogenesis					0.88	-0.34
enolase	Glycolysis and Gluconeogenesis					0.88	-0.34
ribulose 5-phosphate 3-epimerase	Pentose Phosphate Pathway					0.67	-0.36

b)

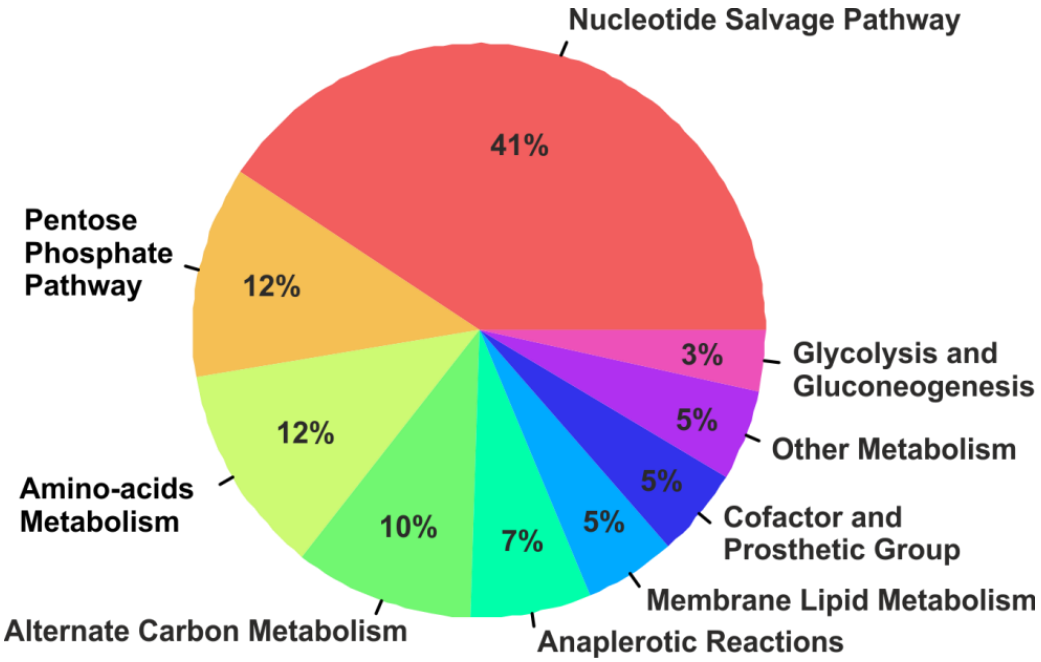
Reaction name	Subsystem	M145	M145+Act	M1146	M1146+Act	Correlation ACT	
S. coelicolor biomass - objective function						M145	M1146
Actinorhodin Pathway Summary	Actinorhodin Biosynthesis					1.00	1.00
2',3'-Cyclic AMP exchange	Exchange					0.00	0.88
NAD transhydrogenase	Oxidative Phosphorylation					-0.12	0.67

Supplementary Figure 6.14: Reactions correlated to actinorhodin production after exometabolomics constraints and media constraints for each strain

a) Reactions correlated to ACT production in the M145 and M145+ACT

b) Reactions correlated to ACT production in M1146+ACT

Only the reactions with a Pearson correlation superior to 0.5 were taken into account as correlated to ACT production. In the heatmap the red colour corresponds to lower predicted flux and green higher predicted flux.



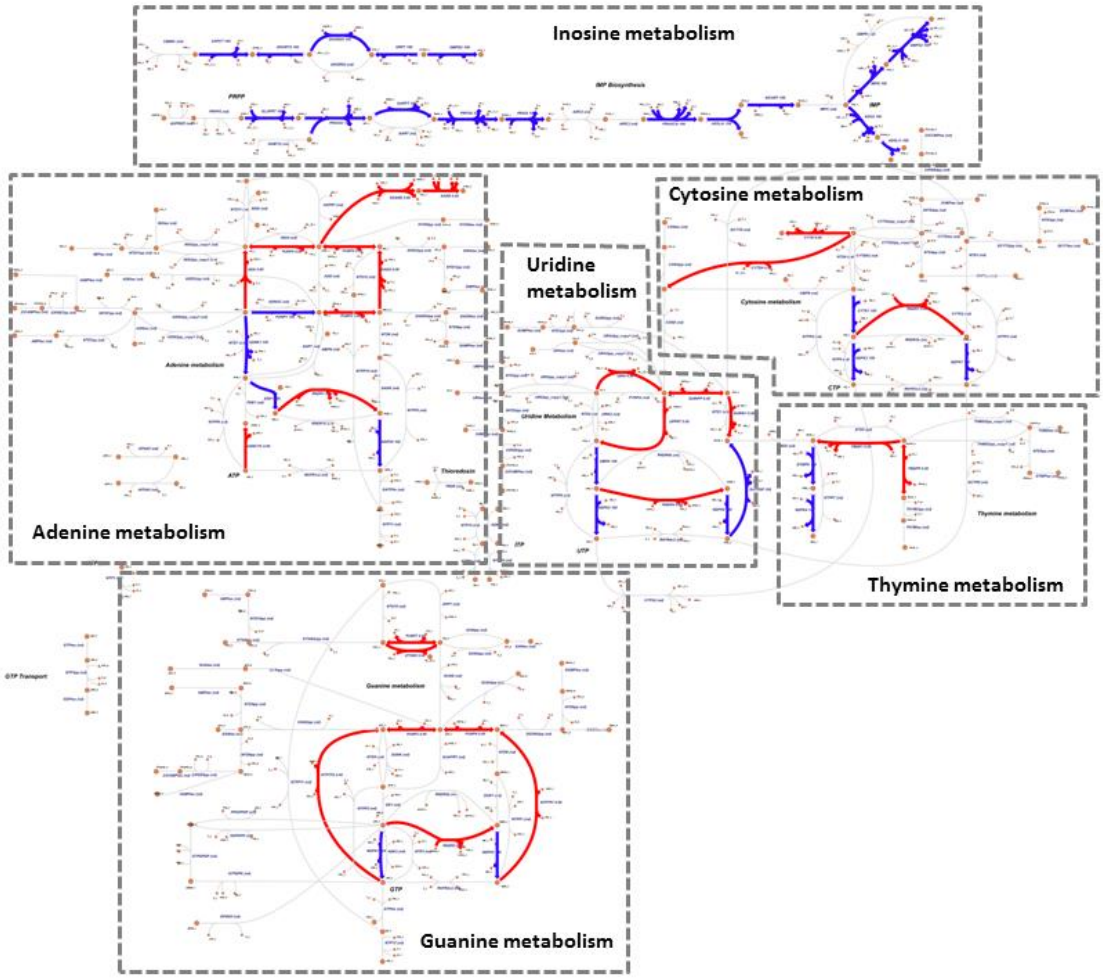
Supplementary Figure 6.15: Metabolic pathways activated by the metabolites secretions

Metabolic pathways associated to the metabolic reactions predicted as activated by the metabolic exports constrained in the model.

RXN_ID	Rxn_name	Metabolic pathway	ACT 0 to 100%	Correlation
ACTt	actinorhodin transport via facilit	Transport and Membrane		NA
Biomass	S. coelicolor biomass - objective	Exchange		-0.93
ACTS	Actinorhodin Biosynthesis pathw	Actinorhodin Biosynthesis		1.00
ACCOAC	acetyl-CoA carboxylase	Membrane Lipid Metabolism		1.00
HCO3E	HCO3 equilibration reaction	Unassigned		1.00
PDH	pyruvate dehydrogenase	Glycolysis and Gluconeogenesis		1.00
ENO	enolase	Glycolysis and Gluconeogenesis		0.98
PGM	phosphoglycerate mutase	Glycolysis and Gluconeogenesis		0.98
GAPD	glyceraldehyde-3-phosphate del	Glycolysis and Gluconeogenesis		0.98
PGK	phosphoglycerate kinase	Glycolysis and Gluconeogenesis		0.98
PYK	pyruvate kinase	Glycolysis and Gluconeogenesis		0.97
TPI	triose-phosphate isomerase	Glycolysis and Gluconeogenesis		0.97
ME2	malic enzyme (NADP)	Anaplerotic Reactions		0.97
EX_nh4(e)	Ammonia exchange	Exchange		0.97
NH4t	ammonia reversible transport	Inorganic Ion Transport and Metal		0.97
GLUDxi	glutamate dehydrogenase (NAD,	Glutamate Metabolism		0.96
GLCt2	D-glucose transport in via proton	Transport and Membrane		0.95
EX_glc(e)	D-Glucose exchange	Exchange		0.95
PFK	phosphofructokinase	Glycolysis and Gluconeogenesis		0.94
FBA	fructose-bisphosphate aldolase	Glycolysis and Gluconeogenesis		0.94
EX_h2o(e)	H2O exchange	Exchange		0.93
H2Ot	H2O transport via diffusion	Transport and Membrane		0.93
CYO2a	cytochrome bc1 oxidoreductase	Oxidative Phosphorylation		0.92
CYO2b	cytochrome aa3 oxidase (menaq	Oxidative Phosphorylation		0.92
CITL	Citrate lyase	Citric Acid Cycle		0.91
ACKr	acetate kinase	Pyruvate Metabolism		0.91
PTAr	phosphotransacetylase	Pyruvate Metabolism		0.91
EX_h(e)	H+ exchange	Exchange		0.90
XYLI2	xylose isomerase	Alternate Carbon Metabolism		0.89
HEX7	hexokinase (D-fructose:ATP)	Alternate Carbon Metabolism		0.89
HCYSMT	homocysteine S-methyltransfera	Methionine Metabolism		0.88
AHC	adenosylhomocysteinase	Methionine Metabolism		0.83
GLUt2r	L-glutamate transport via proton	Transport and Membrane		0.82
EX_glu-L(e)	L-Glutamate exchange	Exchange		0.82
TALA	transaldolase	Pentose Phosphate Pathway		0.77

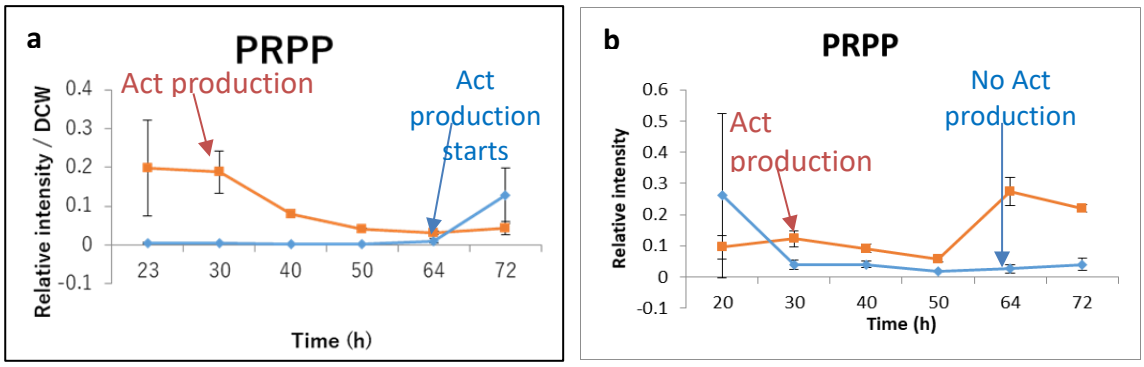
Supplementary Figure 6.16: Intracellular predicted fluxes highly correlated to increasing ACT production

The predicted metabolic fluxes are represented as a heatmap with low relative to the row predicted fluxes in red and higher relative to the row predicted fluxes in green.



Supplementary Figure 6.17: Metabolic map of nucleotides metabolism with fluxes associated to ACT production and with metabolic exports

Reactions in red corresponds to the reaction fluxes associated to metabolic exports while the reactions in blue corresponds to the reactions associated with ACT production.

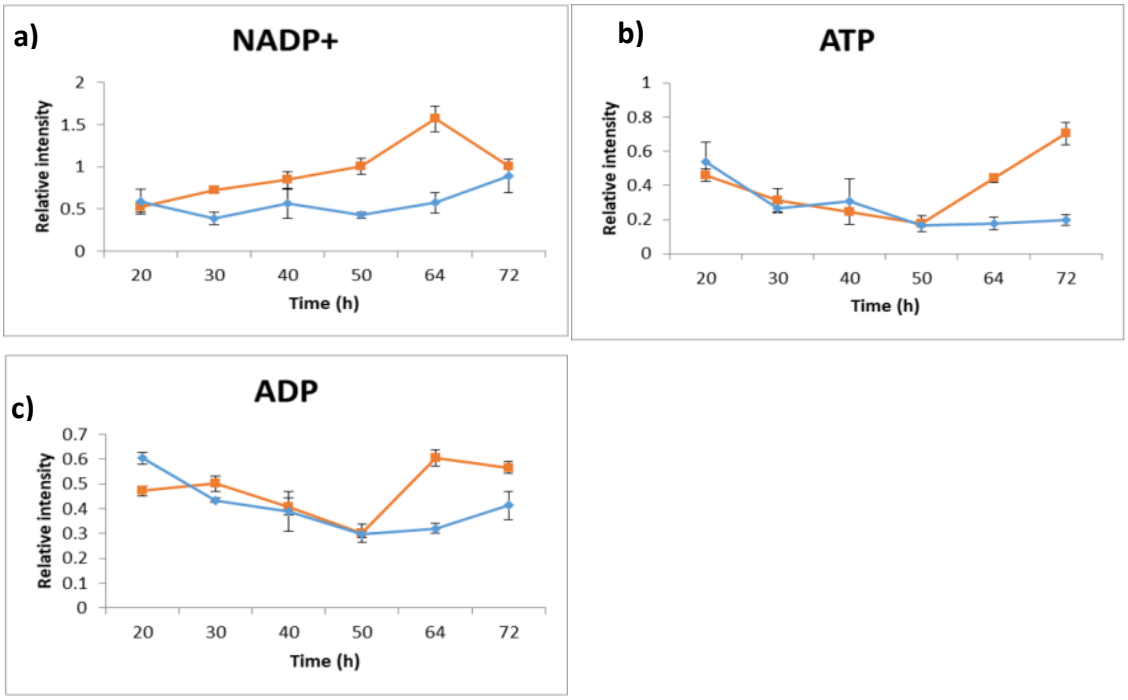


Supplementary Figure 6.18: Intracellular PRPP relative intensity in all the strains

a) Intracellular levels of PRPP in M145 and M145+ACT. The orange curve corresponds to M145+ACT and the blue curve corresponds to M145.

b) Intracellular levels of PRPP in M1146 and M1146+ACT. The orange curve corresponds to M1146+ACT and the blue curve corresponds to M1146.

Data visualisation by Katsuaki Nitta



Supplementary Figure 6.19: Intracellular NADP+, ATP, and ADP relative intensity in all the strains

a) Intracellular levels of NADP+ in M1146 and M1146+ACT

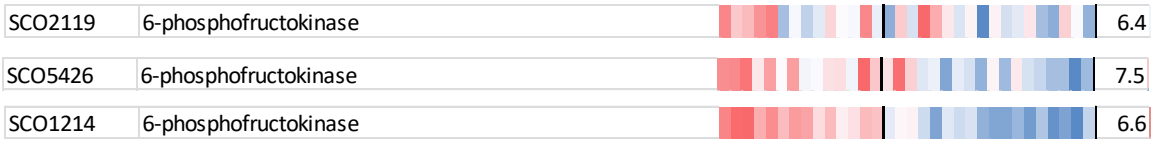
b) Intracellular levels of ATP in M1146 and M1146+ACT

c) Intracellular levels of ADP in M1146 and M1146+ACT

The orange curve corresponds to M1146+ACT and the blue curve corresponds to M1146.

Data visualisation by Katsuaki Nitta

a) Phosphofructokinase (PFK)



b) PRPP synthase

c) Phosphoenolpyruvate carboxykinase (PEPCK)

d) Phosphoenolpyruvate carboxylase (PEPC)

e) Malate

Supplementary Figure 6.20 Gene expression data in the minimal media for M145 for deletion targets

a) Gene expression for Phosphofructokinase (PFK)

b) Gene expression for PRPP synthase

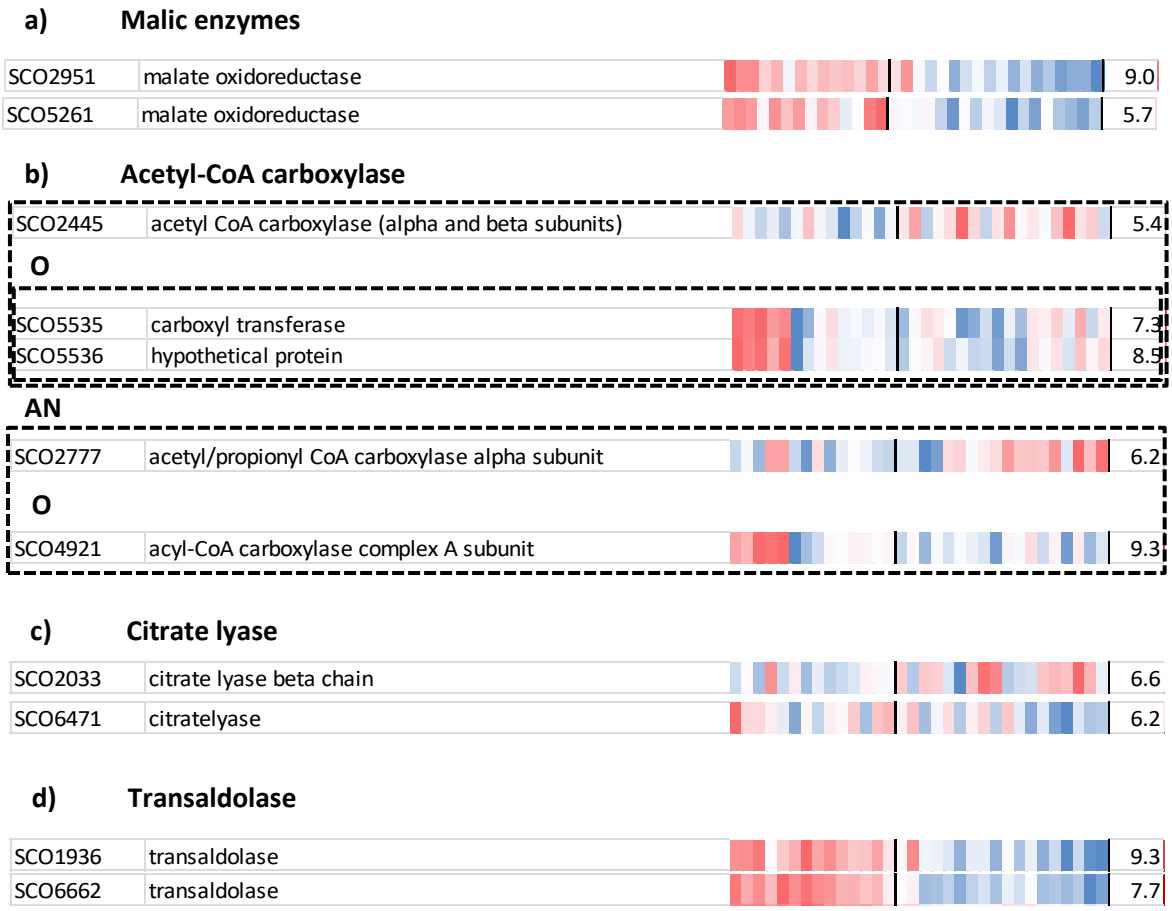
c) Gene expression for Phosphoenolpyruvate carboxykinase (PEPCK)

d) Gene expression for Phosphoenolepyruvate carboxylase (PEPC)

e) Gene expression for Malate synthase

The heatmap colour code corresponds to low gene expression level in blue and high gene expression level in red.

Data by Nieselt et al 2010



Supplementary Figure 6.21: Gene expression data in the minimal media for M145 for the overexpression targets

a) Gene expression for Malic enzymes

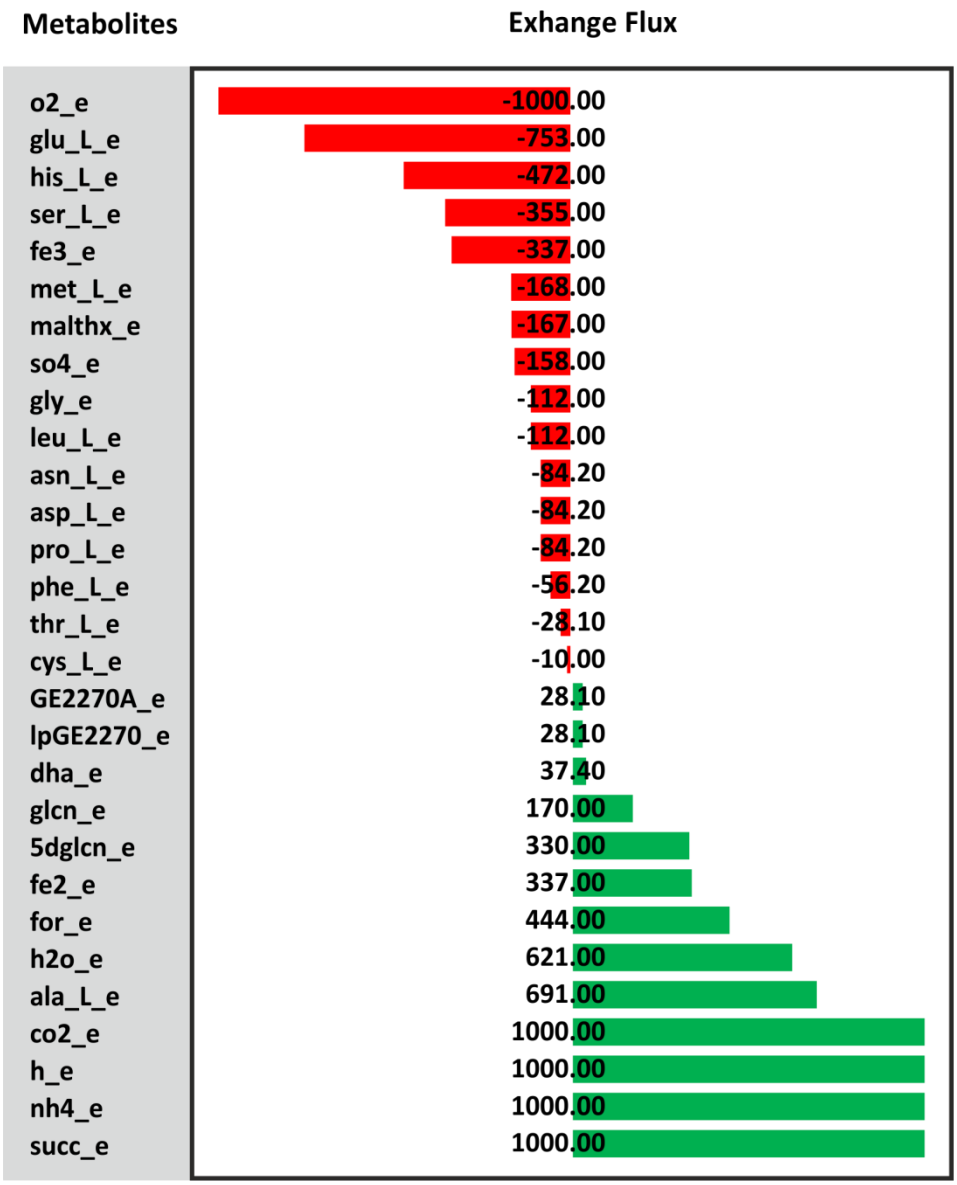
b) Gene expression for Acetyl-CoA carboxylase (ACC)

c) Gene expression for citrate lyase

d) Gene expression for transaldolase

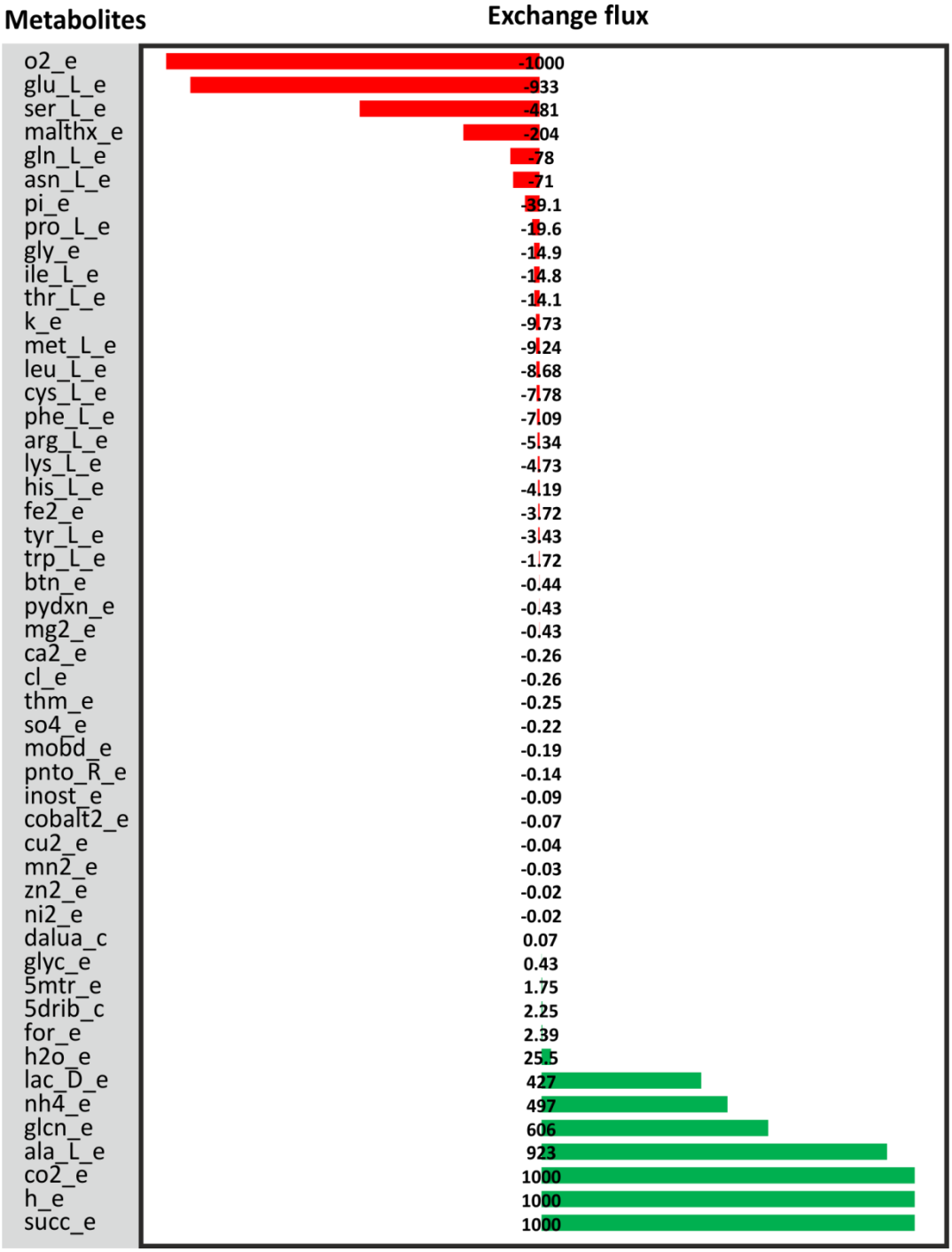
The heatmap colour code corresponds to low gene expression level in blue and high gene expression level in red.

Data by Nieselt et al 2010



Supplementary Figure 6.22: Predicted metabolites imports and exports from media for GE2270A production only in CMAN media

Red histogram corresponds to the predicted uptake fluxes, while the green histogram corresponds to the predicted export fluxes.



Supplementary Figure 6.23: Predicted metabolites imports and exports from media for *S. coelicolor* growth only in CMAN media

Red histogram corresponds to the predicted uptake fluxes, while the green histogram corresponds to the predicted export fluxes.

Strains	Number of essential genes for growth	Genes reducing growth by >10%	Genes reducing growth by <10%	Non-essential genes (not reducing growth)
***S. coelicolor* M1146**	152	12	52	986

Supplementary Table 6.2: Single gene deletion results for biomass optimisation.

Strains	Number of essential genes for GE2270A production	Genes reducing GE2270A production by >10%	Genes reducing GE2270A production by <10%	Non-essential genes (not reducing production)
***S. coelicolor* M1146**	14	27	38	1123

Supplementary Table 6.3: Single gene deletion results for GE2270A production optimisation.

Strains	Number of essential gene pairs for growth	Gene pairs reducing growth by >10%	Gene pairs reducing growth by <10%	Non-essential gene pairs (not reducing growth)
***S. coelicolor* M1146**	172,502	12,829	218,374	475,798

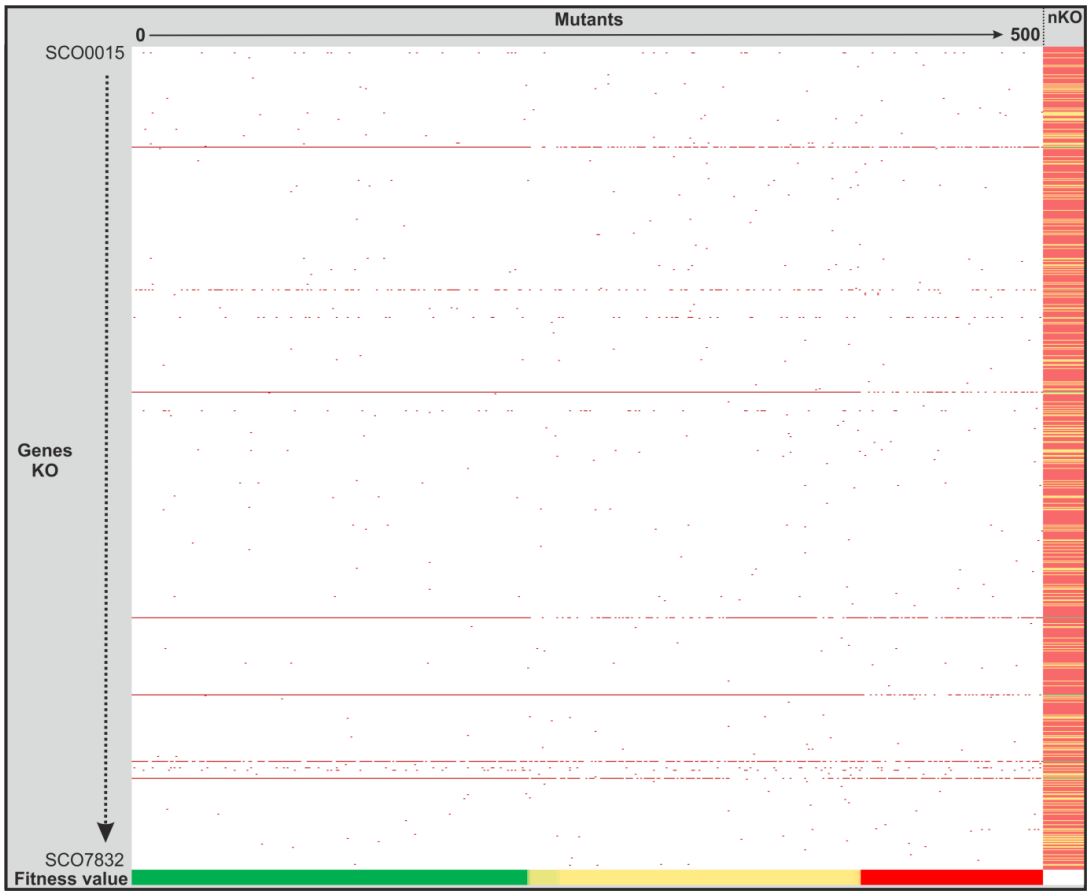
Supplementary Table 6.4: Double gene deletion results for biomass optimisation.

Strains	Number of essential genes pair for GE2270A production	Genes pair reducing GE2270A production by >10%	Genes pair reducing GE2270A production by <10%	Non-essential genes pair (not reducing production)
***S. coelicolor* M1146**	17,970	29,572	41,036	635,629

Supplementary Table 6.5: Double gene deletion results for GE2270A production optimisation.

Gene	nKO (%)	Reactions	Function	Priority
SCO6102	88.6	FADRx2, FLVR, SULRi, FMNRx2	Cofactor nitrite or sulphite reductase: Reduce NADH or NADPH to generate FMNH2, FADH2, Riboflavin, or H2S	**Group I** (>25%)
SCO3295	88.4	FMNRx, FMNRx2		
SCO1223	80	ORNTA	Ornithine transaminase: Use alpha-ketoglutarate to transform ornithine in Glutamate	
SCO6754	79.2	GLYCDx	Glycerol dehydrogenase	
SCO5515	78.8	PGCD	Phosphoglycerate dehydrogenase	
SCO6658	74.4	GND*, GND2*	Phosphogluconate dehydrogenase	
SCO2390	28.2	40 Reactions:* 3OAS*, KAS14*	3-oxoacyl-ACP synthase, and beta-ketoacyl-ACP synthase	
SCO2640	12.2	ASAD*	Aspartate-semialdehyde dehydrogenase	**Group II** (>5%)
SCO6697	10.8	4CMLCL*, OXOAEL*	4-carboxymuconolactone decarboxylase, and 3-oxoadipate enol-lactone hydrolase	
SCO6700	9.4	PCADYOX	Protocatechuate 3,4-dioxygenase	
SCO3473	8.6	DDPGA*, EDA*	2-dehydro-3-deoxy-phosphogluconate aldolase, 2-dehydro-3-deoxy-phosphogluconate aldolase	
SCO0213	7.2	NO2t2r*, NO3t7*	Nitrite transport in via proton symport, and Nitrate transport via nitrite antiport	
SCO2148	1.6	CYO2a*	Cytochrome *bc1* oxidoreductase complex	**Group III** (>0.75%)
SCO1170	1.4	XYLK*, DXYLK*	Xylulokinase, 1-Deoxy-D-xylulose kinase	
SCO5366	1.4	ATPS4r*	ATP synthase	
SCO7266	1.2	38 Reactions: 3OAR*	3-oxoacyl-ACP reductase	
SCO1488	1	UPPRT	Uracil phosphoribosyltransferase	
SCO3899	1	MI1PS*, MI1PS2*	Myo-Inositol-1-phosphate synthases	
SCO4830	1	PROabc*	L-proline transport via ABC system	
SCO5848	1	TGBPA*	Tagatose-bisphosphate aldolase	
SCO6731	1	3OXCOAT*, OXDHCOAT*	3-oxoadipyl-CoA thiolase, and 3-oxo-5,6-dehydrosuberyl-CoA thiolase	
SCO0814	0.8	RMK	Rhamnulokinase	
SCO2093	0.8	HCO3E*	HCO3 equilibration	
SCO2546	0.8	ADA*, DADA*	Adenosine deaminase, and Deoxyadenosine deaminase	
SCO4388	0.8	CS	Cytosine transport in via proton symport	

Supplementary Table 6.6: List of 25 genes knockout candidates to increase production of GE2270A in S. coelicolor M1146.



Supplementary Figure 6.24: Matrix of all the metabolic genes knocked out in the last generation of mutants by OptGene

Most frequent genes knockouts on the column in green and least frequent in red. In the row the fitness value is ordered in green for the highest and red for the lowest.

Strain	Strain1	Strain2	Strain3	Strain4	Strain5	Strain6	Strain7	Strain8	Strain9	Strain10
KO n°1	SCO6102	SCO3295	SCO1223	SCO6754	SCO5515	SCO6102	SCO6102	SCO6102	SCO6102	SCO3295
KO n°2	NA	NA	NA	NA	NA	SCO3295	SCO1223	SCO6754	SCO5515	SCO1223
KO n°3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KO n°4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KO n°5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Strain	Strain11	Strain12	Strain13	Strain14	Strain15	Strain16	Strain17	Strain18	Strain19	Strain20
KO n°1	SCO3295	SCO3295	SCO1223	SCO1223	SCO6754	SCO6102	SCO6102	SCO6102	SCO6102	SCO6102
KO n°2	SCO6754	SCO5515	SCO6754	SCO5515	SCO5515	SCO3295	SCO3295	SCO3295	SCO1223	SCO1223
KO n°3	NA	NA	NA	NA	NA	SCO1223	SCO6754	SCO5515	SCO6754	SCO5515
KO n°4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KO n°5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Strain	Strain21	Strain22	Strain23	Strain24	Strain25	Strain26	Strain27	Strain28	Strain29	Strain30	Strain31
KO n°1	SCO6102	SCO3295	SCO3295	SCO3295	SCO1223	SCO6102	SCO6102	SCO6102	SCO6102	SCO3295	SCO6102
KO n°2	SCO6754	SCO1223	SCO1223	SCO6754	SCO6754	SCO3295	SCO3295	SCO3295	SCO1223	SCO1223	SCO3295
KO n°3	SCO5515	SCO6754	SCO5515	SCO5515	SCO5515	SCO1223	SCO1223	SCO6754	SCO6754	SCO6754	SCO1223
KO n°4	NA	NA	NA	NA	NA	SCO6754	SCO5515	SCO5515	SCO5515	SCO5515	SCO6754
KO n°5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	SCO5515

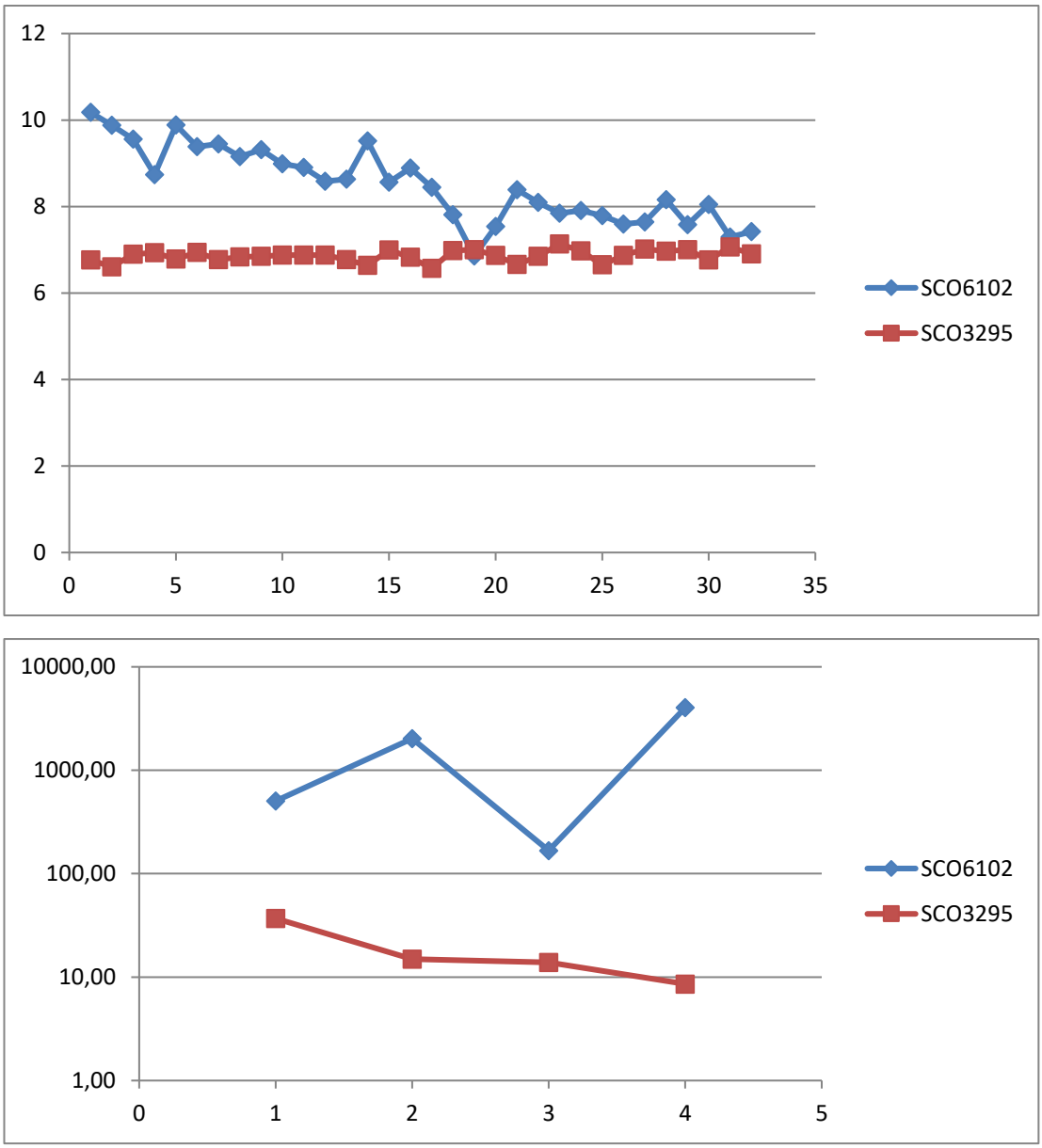
Supplementary Figure 6.25: Combination of all possible viable mutant strains tested *in silico*.

Each gene has a different colour code, brown for SCO6102, pink for SCO3295, blue for SCO1223, green for SCO6754, and purple for SCO5515.

		Strains	WT	S30	S24	S5	S10	S25	S11	S14	S22	S12	S23	S2	S4	S3	S13	S15	S29	S9	S21	S20	S7	S1	S8	S19	S6	S27	S26	S31	S17	S28	S18	S16	DIFF
		Sum Total Flux (mmol/gDW/h)										20034.74006							20034.83492				20047.12361				20106.25049								
		Higher in non-producing mutants										WT phenotype group							Group 2				Group 3				GE2270A producers								
	Reaction	Reaction Name	Metabolic Subsystem																																
	H2Ot	H2O transporter	Transport	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	326	326	326	326	326	326	326	326	26.1324
	D-LAC2	D-lactate transport	Transport	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	393	393	393	393	393	393	393	393	2.81681
	LDH_D	D-lactate dehydrogenase	Pyruvate metabolism	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	396	393	393	393	393	393	393	393	393	2.81681
	CYTBD1	cytochrome oxidase bd I	Oxidative Phosphorylation	471	471	471	471	471	471	471	471	471	471	471	471	471	471	471	471	471	471	471	471	471	471	469	469	469	469	469	469	469	469	1.78572
	NH4t	NH4 transporter	Transport	404	404	404	404	404	404	404	404	404	404	404	404	404	404	404	404	404	404	404	404	404	404	403	403	403	403	403	403	403	403	1.39396
	G1DH	glucose 1-dehydrogenase	Pentose Phosphate Pathway	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	605	605	605	605	605	605	605	605	1.09769
	GLCNt2r	D-gluconate transport	Alternate Carbon Metabolism	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	605	605	605	605	605	605	605	605	1.09769
	GLNLASE	gluconolactonase	Alternate Carbon Metabolism	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	606	605	605	605	605	605	605	605	605	1.09769
	Biomass_SCI	Biomass	Biomass	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	52.4	51.9	51.9	51.9	51.9	51.9	51.9	51.9	51.9	0.54292
		Higher in GE2270A producing mutants										WT phenotype group							Group 2				Group 3				GE2270A producers								
	Reaction	Reaction Name	Metabolic Subsystem																																
	PROI2r	L-Proline transport	Amino Acids Transport	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	-0.9282
	ICDHyr	isocitrate dehydrogenase	Citric acid cycle	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-1	-1	-1	-1	-1	-1	-1	-1	-1.005
	ICL	Isocitrate lyase	Citric acid cycle	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	-1.005
	ASN2r	L-asparagine transport	Amino Acids Transport	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	9.61	10.6	10.6	10.6	10.6	10.6	10.6	10.6	10.6	-1.0313
	ASNTRS	Asparaginyl-tRNA synthetase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	-1.1309
	ASPTRS	Asparaginyl-tRNA synthetase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	-1.1309
	PROTRS	Prolyl-tRNA synthetase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	-1.1309
	GLUDyi	glutamate dehydrogenase	Glutamate Metabolism	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	32.7	32.7	32.7	32.7	32.7	32.7	32.7	32.7	-1.1811
	GLY2r	glycine transport	Amino Acids Transport	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	16.2	16.2	16.2	16.2	16.2	16.2	16.2	16.2	-1.3536
	ACLS	acetolactate synthase	Valine Leucine and Isoleucine Metabolism	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	14.8	14.8	14.8	14.8	14.8	14.8	14.8	14.8	-1.3685
	DHAD1	dihydroxy-acid dehydratase	Valine Leucine and Isoleucine Metabolism	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	14.8	14.8	14.8	14.8	14.8	14.8	14.8	14.8	-1.3685
	KARA1i	acetohydroxy acid isomeroreductase	Valine Leucine and Isoleucine Metabolism	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5	14.8	14.8	14.8	14.8	14.8	14.8	14.8	14.8	-1.3685
	VALTA	valine transaminase	Valine Leucine and Isoleucine Metabolism	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-12	-13	-13	-13	-13	-13	-13	-13	-13	-1.3876
	LEUabc		Amino Acids Transport	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	8.68	10.1	10.1	10.1	10.1	10.1	10.1	10.1	10.1	-1.418
	ALATRS	Alanyl-tRNA synthetase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.51	1.51	1.51	1.51	1.51	1.51	1.51	1.51	-1.5079
	LEUTRS	Leucyl-tRNA synthetase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.51	1.51	1.51	1.51	1.51	1.51	1.51	1.51	-1.5079
	VALTRS	Valyl-tRNA synthetase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.51	1.51	1.51	1.51	1.51	1.51	1.51	1.51	-1.5079
	ATPS4r	ATP synthase	Oxidative Phosphorylation	774	774	774	774	774	774	774	774	774	774	774	774	774	774	774	774	774	774	774	774	774	774	775	775	775	775	775	775	775	775	-1.6121
	CYO2a	cytochrome bc1 oxidoreductase comple	Oxidative Phosphorylation	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	529	529	529	529	529	529	529	529	-1.8076
	CYO2b	cytochrome aa3 oxidase	Oxidative Phosphorylation	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	527	529	529	529	529	529	529	529	529	-1.8076
	GLYTRS	Glycyl-tRNA synthetase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.88	1.88	1.88	1.88	1.88	1.88	1.88	1.88	-1.8849
	SER2r	Serine transport	Amino Acids Transport	481	481	481	481	481	481	481	481	481	481	481	481	481	481	481	481	481	481	481	481	481	481	483	483	483	483	483	483	483	483	-1.9587
	METabc	Methionine transport	Amino Acids Transport	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	9.24	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	-2.1661
	CYSabc	Cystein transport	Amino Acids Transport	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	7.78	9.96	9.96	9.96	9.96	9.96	9.96	9.96	9.96	-2.1813
	CYSTRS	Cysteinyl-tRNA synthetase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.26	2.26	2.26	2.26	2.26	2.26	2.26	2.26	-2.2618
	METTRS	Methionyl-tRNA formyltransferase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.26	2.26	2.26	2.26	2.26	2.26	2.26	2.26	-2.2618
	GLUTRS	Glutamyl-tRNA synthetase	tRNA Charging	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	7.47	7.47	7.47	7.47	7.47	7.47	7.47	7.47	-2.5882
	SERTRS	Seryl-tRNA synthetase	tRNA Charging	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	-2.6388
	GLUt2r	Glutamate transport	Amino Acids Transport	933	933	933	933	933	933	933	933	933	933	933	933	933	933	933	933	933	933	933	933	933	933	936	936	936	936	936	936	936	936	-3.3345
	P5CR	pyrroline-5-carboxylate reductase	Arginine and Proline Metabolism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6.19	6.19	6.19	6.19	6.19	6.19	6.13	6.13	6.13	6.13	6.13	6.13	6.13	6.13	-6.1276
	PROD2	Proline dehydrogenase	Arginine and Proline Metabolism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6.19	6.19	6.19	6.19	6.19	6.19	6.13	6.13	6.13	6.13	6.13	6.13	6.13	6.13	-6.1276
	PPA_1	inorganic diphosphatase	Anapleiotropic reactions	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	180	180	180	180	180	180	180	180	-19.824
	ADK1	adenylate kinase	Nucleotides Salvage Pathway	613	613	613	613	613	613	613	613	613	613	613	613	613	613	613	613	613	613	613	613	613	613	82.2	82.2	82.2	82.2	82.2	82.2	82.2	82.2	-20.852

Supplementary Figure 6.26: Metabolic differences between the producing and non-producing mutants (and WT)

In the row the predicted growth rates in red are the highest and blue ones are the lowest. While the row with the fitness value is ordered in green for the highest and red for the lowest.



Supplementary Figure 6.27: Transcription levels of SCO6102 and SCO3295 under minimal media conditions

a) Gene expression data in minimal media (Nieselt et al 2010)

b) Gene expression data in complex media (Jeong et al 2016)

Gene	Enzyme	Reaction function	Reversible reaction
SCO5711	FMN adenylyltransferase	FMN conversion to FAD	No
SCO5711	riboflavin kinase	Riboflavin conversion into FMN	No
SCO2781	ferricoelichelin reductase	Production of coelichelin	No
SCO2781	ferrioxamine B reductase	Production of Desferrioxamine B	No
SCO2781	ferrioxamine E reductase	Production of Desferrioxamine E	No
SCO5743	thymidylate synthase (Flavin-dependent)	Synthesis of thymidylate from Methylenetetrahydrofolate	No

Supplementary Table 6.7: Genes with FMN binding sites and their associated reactions

Chapter VII

7. Conclusions and future perspectives

In this thesis, I have presented the reconstruction and analysis of constraint-based genome-scale metabolic models to study and engineer primary and secondary metabolism of *Streptomyces* species. The aim of this work is to better understand how it can be used to increase antibiotic production in *Streptomyces* spp., which would ultimately help in the discovery and production of new antibiotics to face the rise of antimicrobial resistance. This involved the update and validation of a high-quality genome-scale metabolic model of *Streptomyces coelicolor* to study its primary and secondary metabolism (*Chapter II*) [1]. This new model was then used to reconstruct a model of *Streptomyces lividans*, a strain that is phylogenetically very closely related, which is then compared with the *S. coelicolor* metabolism to identify differences potentially responsible for antibiotics production differences (*Chapter III*). The comparative reconstruction and comparison method were automated to compare about 50 different Actinobacteria strains and identify metabolic differences between these organisms (*Chapter IV*). The step necessary to obtain condition-specific predictions by constraining metabolic exchanges in the model was automated by building an *R* tool, which was then tested and validated with experimental data (*Chapter V*). Finally, the models and tools developed and validated in this thesis were used to help design theoretically better *S. coelicolor* antibiotic producing strains; such as the native production of actinorhodin and the heterologous production of the antibiotic GE2270A (*Chapter VI*). Each chapter addresses a specific challenge in the

design-build-test-learn (DBTL) cycle of secondary metabolites synthetic biology, with a strong focus on learn and design phases of the cycle (*Figure 7.1*).

7.1 *Streptomyces coelicolor* metabolic modelling and future improvements

The updated genome-scale metabolic model *i*AA1259 of the antibiotic producer model strain *S. coelicolor* presented in this thesis (*Chapter I*) is one of the most advanced metabolic model of this species currently available [1,2]. A new updated metabolic model *i*KS1317 of *S. coelicolor* was published recently with reactions and transporters recently identified [3]. The new model predictions were validated using gene knockout predictions, but the quantitative growth predictions and the fluxes predictions were not validated or compared to previous models [3]. The *i*AA1259 model is validated with multi-omics data across the growth curve, and benchmarked to previous metabolic models to show its superiority in predicting metabolic processes of this organism [1]. In future, the *i*AA1259 model can easily be updated by introducing the secondary metabolites pathways and the transports added in the *i*KS1317 model. This metabolic model can now be used in the synthetic biology DBTL cycle by integrating multi-omics data from the learn phase to the design phase (*Figure 7.1*) to better design *S. coelicolor* strains for the production of antibiotics and other secondary metabolites.

Furthermore, with the recent development of hybrid genome-scale metabolic models, this model is a good starting point to build the next generation of models. For example, by constraining the intracellular metabolic fluxes based on the kinetic data available on the organism, as applied to *Saccharomyces cerevisiae* [4]. Additionally, next generation models will include more information about the cell processes, such as protein structural information, or gene regulation networks, as applied for *Escherichia*

coli [5]. Of course, *S. coelicolor* does not benefit the same profusion of large-scale datasets as *Escherichia coli* and *Saccharomyces cerevisiae.* Nevertheless, once sufficient data becomes available in the future for *S. coelicolor,* building an hybrid model of this strain will help to further improve the metabolic predictions as it did for other organisms [4,5].

7.2 The importance of primary metabolism for secondary metabolites production

Despite the very high similarity between *S. coelicolor* and *S. lividans*, the *S. coelicolor* strain is a superior actinorhodin and undecylprodigiosin natural producer (*Chapter III*). The comparative metabolic modelling of *S. coelicolor* and *S. lividans* strains helped identifying metabolic differences in the central metabolism (e.g., flux distribution, or enzymes absent), potentially responsible for the lower antibiotics production in *S. lividans*. This work also highlighted once more the importance of better understanding the central metabolism of *Streptomyces* spp. and its influence on antibiotics production. Central metabolism is the "metabolic platform" distributing resources for carbon-based molecules, the rewiring of this metabolism has been a major objective for metabolic engineering and synthetic biology [6,7]. The production of polyketide antibiotics, such as actinorhodin, is directly dependent on metabolic precursors produced by this metabolism (i.e., acetyl-CoA and malonyl-CoA) [8,9]. Furthermore, studying the connection between primary and secondary metabolism in the light of the evolution theories of secondary metabolisms [10,11] will help better understand and better engineer secondary metabolic pathways; as described for actinorhodin in *Chapter III*. A more holistic understanding of the metabolic regulations and flux distributions from primary to secondary metabolism will help to rationally

engineer organisms. In the meantime, as our knowledge is still insufficient for fully rational designs, applying genome-scale metabolic modelling and the integrative analysis of multi-omics data helps to identify engineering targets to design better antibiotics producers (*Figure 7.1*) [12,13].

7.3 Large-scale comparison and analysis of microorganisms metabolisms

The rapid development of high throughput sequencing technologies has recently allowed the full-genome sequencing of thousands of microbial genomes, including those of Actinobacteria. At the same time, new tools were developed to rapidly reconstruct genome-scale metabolic models from the genome sequences [14,15], such as the pipeline developed and presented in *Chapter IV*. The combination of the available full-genome data with the fast metabolic model reconstruction tools allows building of tens or hundreds of metabolic models at a time for different organisms [16–19]. The comparative analysis of these metabolic models can help to identify major metabolic differences between organisms (from individuals to entire phyla), informing on their evolutionary and environmental differences [19]. In this thesis, the development of a comparative reconstruction pipeline allowed the reconstruction and analysis of 50 different Actinobacteria metabolic models, to compare metabolic characteristics such as their core and accessory metabolism, or their different predicted environmental conditions. The comparative metabolic analysis of relevant strains can help to identify alternative chassis strains for the bioproduction of a compound, for example by identifying the strains able to metabolically produce a target compound in the learn phase (*Figure 7.1*).

Also, future metabolic models developed here can be used as a base to build models of similar quality to that of *S. coelicolor* for other Actinobacteria strains, to engineer secondary metabolites production or integrate omics data. Then, the high-quality metabolic models of the different chassis strains can be compared *in-silico* for their production capabilities and help create a list of candidate producer strains.

7.4 Metabolic modelling driven strain design for synthetic biology

The development of predictive models, such as genome-scale metabolic models, aims at studying biological systems. As the quality of these models increase in each generation, as shown in *Chapter II* with the updated computational model for *S. coelicolor* [1], their predictive power increases as well. The integration of omics data is a crucial step to learn from the data to represent the organism metabolic state more accurately and better predict the biological system behaviours across the growth phases, as it was presented *in Chapter II*, *III*, and *VI* (*Figure 7.1*).

Automation of the DBTL cycle is a key step to accelerate synthetic biology. Hence, the introduction of new tools to automate the data integration and the reconstruction of metabolic models (*Chapter III* and *V*) can help to speed-up the DBTL cycle by accelerating the learn and design phases (*Figure 7.1*). In the future, the development of more automated user-friendly tools working at different steps of the DBTL cycle would help to move toward true computer-aided design (CAD) of biological systems. The success of CAD biological engineering will also depend on the capacity of computational models to take into account the inherent uncertainty of biological systems. So, the integration of methods to estimate uncertainty in predictive

modelling, such as ensemble modelling [22,23] (as introduced in *Chapter V*), will help to have a greater predictive power for the next generation of CAD tools.

7.5 Other applications and final remarks

The work presented in this thesis can be used beyond secondary metabolite production in *Streptomyces* and other Actinobacteria. The metabolic models developed in Chapter II, III, and IV can be used to design strains producing other valuable biomolecules, such as proteins [24], or biofuels [25]. The tools developed in *Chapter IV* and *V* can be used for any other microorganisms to study and/or engineer their biological systems.

With an ever-increasing amount of biological data, particularly with multi-omics approaches, scientists need new tools to rapidly and reliably make sense of these data. In parallel, systems and synthetic biology applications have the potential to help us better understand and engineer life. This ultimately aims at helping us address major challenges that our societies are facing, from the rise of antimicrobial resistances to the climate catastrophe. The work carried for this thesis, hopefully, makes a useful contribution to the scientific and technologic endeavour of engineering biology to solve current and future problems.

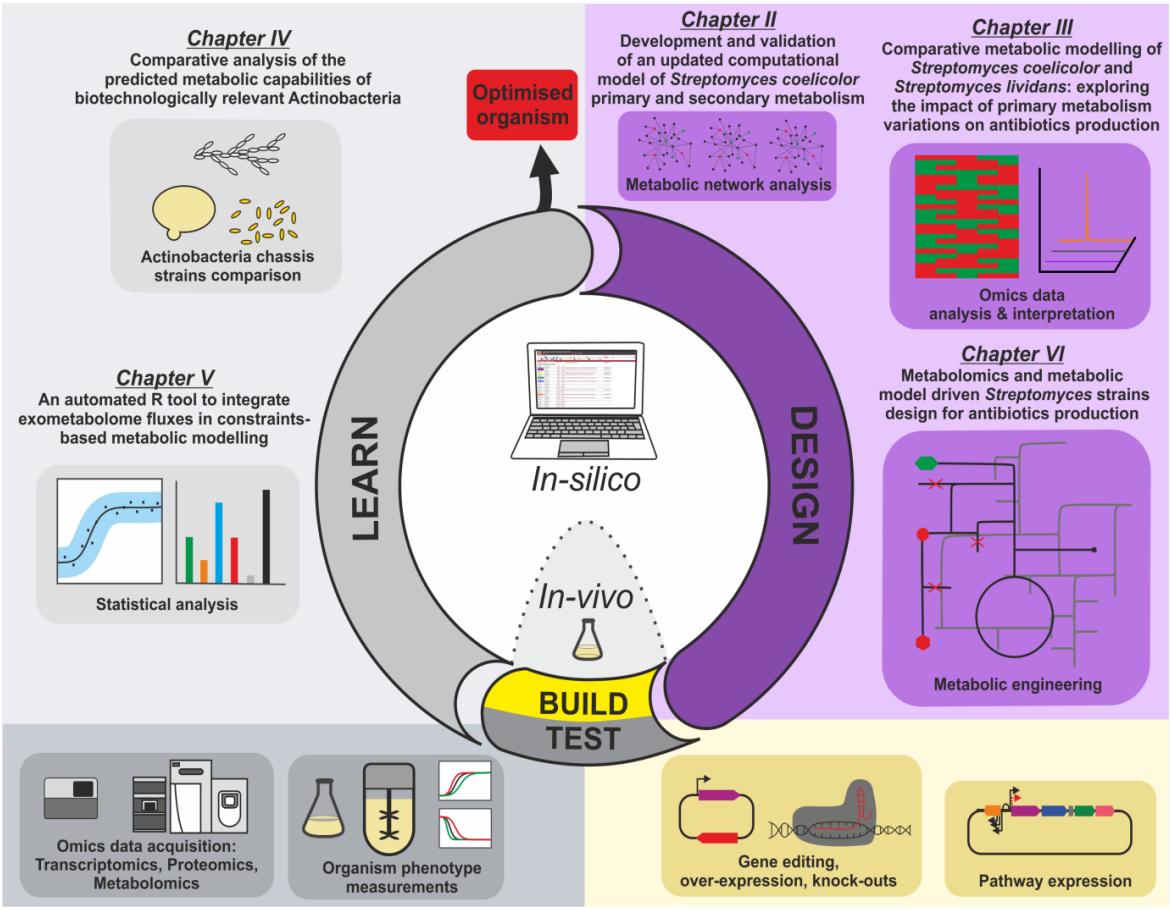


Figure 7.1: Thesis overview

Overview of the thesis chapters and their place in the design-build-test-learn cycle for the synthetic biology of antibiotics.

Chapter II: Development and validation of an updated computational model of *Streptomyces coelicolor* primary and secondary metabolism. Design phase.
Chapter III: Comparative metabolic modelling of *Streptomyces coelicolor* and *Streptomyces lividans*: exploring the impact of primary metabolism variations on antibiotics production. Design phase.
Chapter IV: Comparative analysis of the predicted metabolic capabilities of biotechnologically relevant Actinobacteria. Learn phase.
Chapter V: An automated *R* tool to integrate exometabolome fluxes in constraints-based metabolic modelling. Learn phase.
Chapter VI: Metabolomics and metabolic model driven *Streptomyces* strains design for antibiotics production. Design phase.

7.6 References

(1) Amara, A.; Takano, E.; Breitling, R. Development and Validation of an Updated Computational Model of *Streptomyces Coelicolor* Primary and Secondary Metabolism. *BMC Genomics* **2018**, *19* (1), 519.

(2) Mohite, O. S.; Weber, T.; Kim, H. U.; Lee, S. Y. Genome-Scale Metabolic Reconstruction of Actinomycetes for Antibiotics Production. *Biotechnol. J.* **2019**, *14* (1), 1800377.

(3) Kumelj, T.; Sulheim, S.; Wentzel, A.; Almaas, E. Predicting Strain Engineering Strategies Using IKS1317: A Genome-Scale Metabolic Model of *Streptomyces Coelicolor*. *Biotechnol. J.* **2019**, *14* (4), 1800180.

(4) Sánchez, B. J.; Zhang, C.; Nilsson, A.; Lahtvee, P.-J.; Kerkhoven, E. J.; Nielsen, J. Improving the Phenotype Predictions of a Yeast Genome-Scale Metabolic Model by Incorporating Enzymatic Constraints. *Mol. Syst. Biol.* **2017**, *13* (8), 935.

(5) Monk, J. M.; Lloyd, C. J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; et al. *i*ML1515, a Knowledgebase That Computes *Escherichia Coli* Traits. *Nat. Biotechnol. 2017 3510* **2017**.

(6) Nielsen, J.; Keasling, J. D. Engineering Cellular Metabolism. *Cell* **2016**, *164* (6), 1185–1197.

(7) Pandit, A. V.; Srinivasan, S.; Mahadevan, R. Redesigning Metabolism Based on Orthogonality Principles. *Nat. Commun.* **2017**, *8*, 15188.

(8) Ryu, Y.-G.; Butler, M. J.; Chater, K. F.; Lee, K. J. Engineering of Primary Carbohydrate Metabolism for Increased Production of Actinorhodin in *Streptomyces Coelicolor*. *Appl. Environ. Microbiol.* **2006**, *72* (11), 7132–7139.

(9) Butler, M. J.; Bruheim, P.; Jovetic, S.; Marinelli, F.; Postma, P. W.; Bibb, M. J. Engineering of Primary Carbon Metabolism for Improved Antibiotic Production in *Streptomyces Lividans*. *Appl. Environ. Microbiol.* **2002**, *68* (10), 4731 LP – 4739.

(10) Hodgson, D. A. Primary Metabolism and Its Control in Streptomycetes: A Most Unusual Group of Bacteria. *Adv. Microb. Physiol.* **2000**, *42*, 47–238.

(11) Bu'Lock, J. D. Intermediary Metabolism and Antibiotic Synthesis. *Adv. Appl. Microbiol.* **1961**, *3*, 293–342.

(12) Rokem, J. S.; Lantz, A. E.; Nielsen, J. Systems Biology of Antibiotic Production by Microorganisms. *Nat. Prod. Rep.* **2007**, *24* (6), 1262–1287.

(13) Breitling, R.; Achcar, F.; Takano, E. Modeling Challenges in the Synthetic Biology of Secondary Metabolism. *ACS Synth. Biol.* **2013**, *2* (7), 373–378.

(14) Overbeek, R.; Olson, R.; Pusch, G. D.; Olsen, G. J.; Davis, J. J.; Disz, T.; Edwards, R. A.; Gerdes, S.; Parrello, B.; Shukla, M.; et al. The SEED and the Rapid Annotation of Microbial Genomes Using Subsystems Technology (RAST). *Nucleic Acids Res.* **2014**, *42* (D1), D206–D214.

(15) Thiele, I.; Palsson, B. Ø. A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction. *Nat. Protoc.* **2010**, *5* (1), 93–121.

(16) Alam, M. T.; Medema, M. H.; Takano, E.; Breitling, R. Comparative Genome-scale Metabolic Modeling of Actinomycetes: The Topology of Essential Core Metabolism. *FEBS Lett.* **2011**, *585* (14), 2389–2394.

(17) Karlsen, E.; Schulz, C.; Almaas, E. Automated Generation of Genome-Scale Metabolic Draft Reconstructions Based on KEGG. *BMC Bioinformatics* **2018**, *19* (1), 467.

(18) Pitkänen, E.; Jouhten, P.; Hou, J.; Syed, M. F.; Blomberg, P.; Kludas, J.; Oja, M.; Holm, L.; Penttilä, M.; Rousu, J.; et al. Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. *PLoS Comput. Biol.* **2014**, *10* (2), e1003465.
(19) Magnúsdóttir, S.; Heinken, A.; Kutt, L.; Ravcheev, D. A.; Bauer, E.; Noronha, A.; Greenhalgh, K.; Jäger, C.; Baginska, J.; Wilmes, P.; et al. Generation of Genome-Scale Metabolic Reconstructions for 773 Members of the Human Gut Microbiota. *Nat. Biotechnol.* **2017**, *35* (1), 81–89.
(20) Palsson, B. *Systems Biology : Constraint-Based Reconstruction and Analysis*.
(21) Schellenberger, J.; Que, R.; Fleming, R. M. T.; Thiele, I.; Orth, J. D.; Feist, A. M.; Zielinski, D. C.; Bordbar, A.; Lewis, N. E.; Rahmanian, S.; et al. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6* (9), 1290–1307.
(22) Tsigkinopoulou, A.; Baker, S. M.; Breitling, R. Respectful Modeling: Addressing Uncertainty in Dynamic System Models for Molecular Biology. *Trends Biotechnol.* **2017**, *35* (6), 518–529.
(23) Tsigkinopoulou, A.; Hawari, A.; Uttley, M.; Breitling, R. Defining Informative Priors for Ensemble Modeling in Systems Biology. *Nat. Protoc.* **2018**, *13* (11), 2643–2663.
(24) Anné, J.; Mellaert, L. *Streptomyces Lividans* as Host for Heterologous Protein Production. *FEMS Microbiol. Lett.* **1993**, *114* (2), 121–128.
(25) Anteneh, Y. S.; Franco, C. M. M. Whole Cell Actinobacteria as Biocatalysts. *Front. Microbiol.* **2019**, *10*, 77.