

Computational studies of industrial hosts for improved production of the 2nd generation of biofuels

Thèse N° 9068

Présentée le 1^{er} février 2019

à la Faculté des sciences de base

Laboratoire de biotechnologie computationnelle des systèmes

Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

MILENKO TOKIC

Acceptée sur proposition du jury

Prof. A.-C. Corminboeuf, présidente du jury

Prof. V. Hatzimanikatis, Dr L. Miskovic, directeurs de thèse

Prof. S. Waldherr, rapporteur

Prof. L. Blank, rapporteur

Prof. F. Maréchal, rapporteur

2019

To Edit and Klara
My day to day inspiration

Acknowledgments

First, I would like to express my sincere gratitude to my supervisor Vassily Hatzimanikatis for giving me the opportunity to work and learn from him. Vassily, you have been an endless source of ideas but also challenges at the same time. Thank you very much for your patience and scientific enthusiasm you shared with me.

A big part of my gratitude goes to my co-supervisor, Ljubisa Miskovic. Misko, you have been a constant support to me since the day I came for the interview. Thank you very much for the fruitful scientific discussions, for abstracts proofreading, for day-to-day small talks about sport, films, recent updates on the criterion collection... It is a never-ending list. Beside it, I enjoyed sharing the office with you.

Beside my supervisors, I would like to thank people from my PhD committee: Prof. Lars Blank, Prof. Steffen Waldherr, Prof. Francois Marechal and Prof. Clemence Corminboeuf for giving me a proper sweating during the exam. Also, here I would like to acknowledge people from the SynPath project with whom I collaborated during my time at EPFL, namely Meric (for answering on every single stupid question I asked) and Noushin (the pathway master) from the LCSB, and Dario, Birgitta, and Lars from RWTH Aachen.

I would also to acknowledge Christine Kupper for the administrative support and the help in finding my first Lausanne apartment.

Special thanks go to the past and present LCSB members. You made every day of my PhD life enjoyable: Tuure and Yves as members of La Résistance, Georgios F. for being the calmest man in the universe (you know that half of the COBRA'15 prize is yours), Anush for always willing to help, Stefano for the endless laugh and introduction to the Lebanese, Joana for the Portugal trip invitation and the manuscript proofreading, Vikash for delicious Indian food and mint sauce, Sofia for being past but also future concert buddy, Homa for generating 1TB of fingerprints, Daniel H, Tiziano, Stepan, Georgios S, Alex, Julen, Jasmin, Daniel W, Pierre, Maria, Thomas, Liliana, Robin and Zhaleh for keeping the lab's atmosphere bright.

Some people deserve a special mentioning, people from Trebavci group: Baki, Mara, Pedja M, Ana, Losmi, Mica, Andrej, Tamara, Nevena, Laza, Zlatko, Natasa & co, Obren, the rest of EPFLers from Serbia: Spa, Mima, Drazen, Nikola Kln, Andrici and people from Room 343: kum Todor, scientific and day-of-birth brother Pedja, one and only roommate Moki, the best football manager Mich, Bjoris, Ogi, Boza, Ozzmen. Thank you, guys, you have been the light at the end of the tunnel.

It is hard to find suitable words to express how grateful I am to my family for all the love and support they gave to me. In the absence of words, I can say only: mama, tata, hvala vam na svemu i puno vas volim. And my brother and his wife, Nikola and Boka, for being always there and for the endless discussions about the origin of swiss cheese. I also thank Karvak family for the love, acceptance and for taking care of Klara while I was writing this thesis.

Last but not least, Edit and Klara. You are my endless source of inspiration, my origin of happiness and my lifetime loves. Words can't describe how much you mean to me. You are my whole world. I love you both to the moon and back!

Abstract

In an effort of overcoming the limited availability of fossil energy resources and moving toward a sustainable economy, the focus of the research and development in the area of biofuels has shifted towards developing the 2nd generation of fuels that should be produced via microbial fermentation. The 2nd generation biofuels should satisfy several criteria such as lower emission, higher energy density and should be less corrosive to engines. Although for many of these molecules, natural producers are known, they are not produced in the appreciated quantities. Heterologous expression of biosynthetic pathways taken from natural producers or expression of *de novo* synthetic pathways into microbial workhorses such as *Escherichia coli* allows for production of a wide spectra of biofuels. Recently, *Pseudomonas putida* has emerged as an amenable production host with a number of advantages over natural producers. *P. putida* is a non-pathogenic soil bacterium known for its versatile metabolism. This highly adaptive bacterium has been found to survive and grow on a wide range of substrates from pure caffeine to toxic industrial waste. Moreover, *P. putida* is tolerant to high toxicity compounds such as 2nd generation biofuel butanol. Counterintuitively, *P. putida* was seldom used as a host for the production of biofuels.

In this thesis, we performed a computational analysis of this organism to evaluate its metabolic capacities to serve as a potential 2nd generation biofuels production host. Its capacity was compared against heavily used host *E. coli* on the test example of production of one of the most prominent fuel candidate Methyl Ethyl Ketone (MEK).

To this end, we first performed a thermodynamic curation of the genome-scale iJN1411 model of *P. putida*, and we then used redGEM and lumpGEM algorithms to derive a consistently reduced large-scale stoichiometric model of *P. putida*. We integrated different omics data into resulting models and we proposed a novel way of constraining concentrations of the same species across different compartments while maintaining the consistency with the experimental measurements. To assess its capability to serve as a host, we evaluated and analyzed more than 3.6 millions biosynthetic pathways for production of 5 MEK precursors, in both heavily used industrial workhorse *E. coli* and

rising *P. putida*. We compared their capability and performance with respect to thermodynamic feasibility and yield and we identified the most promising pathways for MEK production. Beside the discovered and evaluated pathways, we present a new way of clustering of feasible pathways and pathway precursors that allows us to classify and evaluate alternative ways for production and to better understand chemistry that leads towards the target molecule.

Identification of metabolic engineering targets for the improved biofuel production requires kinetic models. We used the ORACLE framework to generate a population of large-scale kinetic models of *P. putida*, and we employed these models in two studies. In the first study, for a wild-type strain of *P. putida* grown under aerobic conditions using glucose as a carbon source, we evaluated and validated the predictions of the generated kinetic models against a collection of experimental single-gene knockouts. In the second study, we analyzed the capacity of *P. putida* to adapt to increased energy demand, and we identified potential metabolic engineering targets for improved resistance of this organism to stress conditions.

Keywords: *Pseudomonas putida*, *Escherichia coli*, thermodynamic, model reduction, pathway feasibility, pathway similarity, pathway clustering, Methyl Ethyl Ketone, large-scale and genome-scale kinetic models, kinetic parameters uncertainty

Résumé

Afin de surmonter la disponibilité limitée des ressources en énergie fossile et de progresser vers une économie durable, la recherche et le développement dans le domaine des biocarburants ont mis l'accent sur le développement de la deuxième génération de carburants issus de la fermentation microbienne. Les biocarburants de 2ème génération devraient satisfaire à plusieurs critères, tels que la réduction des émissions et la densité énergétique, et devraient être moins corrosifs pour les moteurs. Bien que pour beaucoup de ces molécules, les producteurs naturels soient connus, ils ne sont pas produits en quantités appréciables. L'expression hétérologue de voies de biosynthèse provenant de producteurs naturels ou l'expression de voies de synthèse *de novo* dans des « bêtes de somme » microbiennes tels que *Escherichia coli* permet la production d'un large spectre de biocarburants. Récemment, *Pseudomonas putida* s'est avéré hôte favorable à la production avec de nombreux avantages par rapport aux producteurs naturels. *P. putida* est une bactérie du sol non pathogène connue pour son métabolisme polyvalent. On a constaté que cette bactérie hautement adaptative survit et se développe sur une large gamme de substrats, de la caféine pure aux déchets industriels toxiques. De plus, *P. putida* est tolérant à des composés hautement toxiques tels que le biocarburant butanol de 2ème génération. Contre intuitivement, *P. putida* était rarement utilisé comme hôte pour la production de biocarburants.

Dans cette thèse, nous avons effectué une analyse computationnelle de cet organisme pour évaluer ses capacités métaboliques en tant qu'hôte potentiel de production de biocarburants de 2e génération. Sa capacité a été comparée à celle de l'hôte fortement utilisé *E. coli*, en se basant sur un test de production de l'un des principaux candidats au carburant, la méthyléthylcétone (MEK).

À cette fin, nous avons d'abord procédé à une curation thermodynamique du modèle iJN1411 à l'échelle du génome de *P. putida*, puis nous avons utilisé des algorithmes redGEM et lumpGEM pour obtenir un modèle stœchiométrique à grande échelle de *P. putida* réduit de manière consistante. Nous avons intégré différentes données omiques dans les modèles résultants et nous avons proposé une nouvelle méthode pour

limiter les concentrations des mêmes espèces dans différents compartiments tout en maintenant la cohérence avec les mesures expérimentales. Pour évaluer sa capacité à servir d'hôte, nous avons évalué et analysé plus de 3,6 millions de voies de biosynthèse pour la production de 5 précurseurs de MEK, sur les deux bêtes de somme industrielles que sont *E. coli* et la nouvelle venue *P. putida*. Nous avons comparé leurs capacités et leurs performances en termes de faisabilité thermodynamique et de rendement, et nous avons identifié les voies les plus prometteuses pour la production de MEK. Outre les voies découvertes et évaluées, nous présentons une nouvelle façon de regrouper les voies et les précurseurs de voies possibles, ce qui nous permet de classer et d'évaluer des méthodes alternatives de production et de mieux comprendre la chimie menant à la molécule cible.

L'identification de cibles d'ingénierie métabolique pour l'amélioration de la production de biocarburant nécessite des modèles cinétiques. Nous avons utilisé l'outil ORACLE pour générer une population de modèles cinétiques à grande échelle de *P. putida*, et nous avons utilisé ces modèles dans deux études. Dans la première étude, pour une souche sauvage de *P. putida* cultivée dans des conditions aérobies en utilisant le glucose comme source de carbone, nous avons évalué et validé les prédictions des modèles cinétiques générés par rapport à un ensemble expérimental de KO à gène unique. Dans la seconde étude, nous avons analysé la capacité de *P. putida* à s'adapter à une demande énergétique accrue et nous avons identifié des cibles potentielles d'ingénierie métabolique pour améliorer la résistance de cet organisme aux conditions de stress.

Mots-clés: *Pseudomonas putida*, *Escherichia coli*, thermodynamique, réduction de modèle, faisabilité de voies métaboliques, similarité des voies, regroupement des voies, méthyléthylcétone, modèles cinétiques à grande échelle et à l'échelle du génome, incertitude des paramètres cinétiques

Table of Contents

Acknowledgments.....	ii
Abstract.....	iii
Résumé.....	v
List of figures	x
List of tables.....	xi
Introduction	1
Motivation	1
Aim and scope	4
Thesis outline	4
Articles included in this thesis	5
Chapter 1. Background.....	7
1.1 Genome scale models	7
1.2 Constrained-based modeling.....	8
1.2.1 Flux balance analysis	9
1.2.2 Thermodynamics-based flux analysis	9
1.2.3 Other constrained-based methods.....	10
1.3 Systems-level analysis and data integration.....	10
1.4 Systematic reduction of Genome Scale Models.....	11
1.5 Optimization of synthetic pathways for efficient biofuel production, building a kinetic model.....	11
Chapter 2. Curation and systematic reduction of <i>Pseudomonas putida</i> GEM iJN1411	15
2.1 Introduction	15
2.2 Methods.....	17
2.2.1 Thermodynamic-based flux analysis (TFA)	17
2.2.2 Integration of metabolomics data while considering cellular compartments.	17
2.2.3 Gap-filling of thermodynamically curated iJN1411.....	19
2.2.4 Systematic reduction.....	19
2.3 Results and Discussion	20
2.3.1 Thermodynamic curation of the <i>P. putida</i> GEM iJN1411	20
2.3.2 Core reduced stoichiometric models of <i>P. putida</i>	24

2.3.3 Generation of lumped reaction for production of biomass building blocks	27
2.3.4 Consistency checks of core reduced models	28
2.3.5 Essentiality of genes encoding for EDA and EDD	31
2.4 Conclusions	33
Chapter 3. Discovery and evaluation of biosynthetic pathways for the production of five methyl ethyl ketone precursors	35
3.1 Introduction	35
3.2 Methods.....	38
3.2.1 Metabolic network generation	39
3.2.2 Pathway reconstruction	41
3.2.3 Pathway evaluation	41
3.2.4 Subnetwork reconstruction	43
3.2.5 Ranking and visualization of <i>in silico</i> pathways	45
3.2.6 Experimental implementation and pathway optimization.....	46
3.3 Results and Discussion	46
3.3.1 Generated metabolic network around Methyl Ethyl Ketone.....	46
3.3.2 Pathway reconstruction towards five target compounds.....	48
3.3.3 Evaluation of reconstructed pathways.....	49
3.3.4 Identification and analysis of anabolic subnetworks capable of synthesizing target molecules	54
3.3.5 Clustering of feasible pathways.....	61
3.3.6 Ranking of biosynthetic pathways and recommendations	66
3.3.7 Further experimental implementation and pathway optimization	68
3.4 Conclusions	69
Chapter 4. Comparative study on <i>Escherichia coli</i> and <i>Pseudomonas putida</i> metabolic capabilities.....	73
4.1 Introduction	73
4.2 Methods.....	76
4.2.1 Reduced <i>P. putida</i> and <i>E. coli</i> models.....	76
4.2.2 Displacements from the thermodynamic equilibrium	77
4.2.3 Pathway evaluation	77
4.3 Results and Discussion	78
4.3.1 Stoichiometric comparison of reduced models	78
4.3.2 Cofactor production costs.....	80
4.3.3 Displacements from the thermodynamic equilibrium.....	82

4.3.3 <i>Pseudomonas putida</i> as a potential host for production of 5 methyl ethyl ketone precursors	86
4.4 Conclusions	90
Chapter 5. Large-scale kinetic metabolic models of <i>Pseudomonas putida</i> for consistent design of metabolic engineering strategies	93
5.1 Introduction	93
5.1.1 Issues in building kinetic models	94
5.1.2. Standard Requirements in Kinetic Modeling	97
5.1.3 Current scope, level of details and consistency in kinetic models	101
5.1.4 Review of published kinetic models of <i>P. putida</i>	102
5.2 Methods	103
5.2.1 Configuring stoichiometric model for kinetic studies of wild-type physiology	103
5.2.2 Configuring stoichiometric model for kinetic studies of stress conditions	107
5.2.3 Construction of the large-scale kinetic models	110
5.3 Results	111
5.3.1 Kinetic study of wild type <i>P. putida</i> physiology	111
5.3.2 Kinetic study of increased ATP demand in <i>P. putida</i>	117
5.4 Conclusions	122
Conclusions and Perspectives	123
Bibliography	129

List of figures

Figure 1.1 The conceptual basis of constraint-based modeling.	8
Figure 2.1 The core networks generated by the redGEM algorithm from iJN1411 genome-scale model.....	26
Figure 2.2 Flux variability of reactions in 4 starting subsystems in D2 and GEM.....	30
Figure 2.3 Starch and sucrose metabolism and cell envelope biosynthesis cellulose metabolism from GEM	31
Figure 2.4 The directionality of transketolase 2 (TKT2) impacts the in silico essentiality of two genes encoding enzymes EDD and EDA from the Entner-Doudoroff pathway	32
Figure 3.1 Computational pipeline for discovery, evaluation and analysis of biosynthetic pathways.....	40
Figure 3.2 Growth of the generated metabolic network over 5 generations	47
Figure 3.3 Three alternative ways to produce 3oxPNT from acetate through 2 intermediate metabolites: 2- ethylmalate and 3-hydroxypentanoate.....	53
Figure 3.4 Alternative ways of producing 3oxPNT from glucose.	56
Figure 3.5 Three different pathways from acetate to 3oxPNT sharing the same lumped reaction. Ac (acetate), r5p (D-ribose 5-phosphate), sucCoA (succinyl CoA), g3p (glyceraldehyde 3-phosphate).....	58
Figure 3.6 Clustering of the 115 reconstructed pathways from acetate to 3oxPNT.	62
Figure 3.7 Clustering of 115 active subnetworks corresponding to 115 reconstructed pathways from acetate to 3oxPNT.	65
Figure 3.8 The highest ranked candidate pathways for the production of (a) 3oxPNT, (b) MVK, (c) BuNH ₂ , (d) iB2OT and (e) MEKCNH.....	68
Figure 4.1 Upper glycolysis in <i>P. putida</i> (a) and <i>E. coli</i> (b).	75
Figure 4.2 Displacements from the thermodynamic equilibrium for <i>E. coli</i>	84
Figure 4.3 Displacements from the thermodynamic equilibrium for <i>P. putida</i>	85
Figure 5.1 8 bidirectional reactions (BDRs) from D2 used to compute 48 feasible flux directionality profiles	106
Figure 5.2 Biomass yield for minimal sum of fluxes per FDP	107
Figure 5.3 Minimal glucose uptake rate as a function of ATP hydrolysis flux for different combinations of allowed maximal rates of the oxygen uptake and ATP synthesis.....	109
Figure 5.4 Distribution of the control coefficients of glucose uptake and specific cell growth in wild-type <i>P.</i> <i>putida</i> cells	113
Figure 5.5 Ranges of K_m 's predicted by ORACLE.....	116
Figure 5.6 Ranges of V_{max} 's predicted by ORACLE	117
Figure 5.7 Fermentation profile of <i>P. putida</i> metabolism under increased ATP demand.....	118
Figure 5.8 Control coefficients of the growth in the stress conditions	120
Figure 6.1 Design-based strain engineering workflow	127

List of tables

Table 2.1 Reactions that are determined as bi-directional in FBA, but unidirectional in TFA.....	22
Table 2.2 Additionally constrained reactions upon implementation of experimental data.....	23
Table 2.3 An overview of the delivered reduced models	28
Table 3.1 Reconstructed pathways towards five target compounds.	49
Table 3.2 Number of known reaction steps versus all reaction steps in the predicted pathways.	51
Table 3.3 Alternative anabolic subnetworks for 5 target compounds together with their lumped reactions and precursors.....	55
Table 3.4 Top ten connecting precursor combinations for the production of 3OXPNP.....	60
Table 4.1 Comparison of reduced models of <i>P. putida</i> and <i>E. coli</i>	79
Table 4.2 Comparison between <i>P. putida</i> and <i>E. coli</i> alternative subnetworks for production of different cofactors.	81
Table 4.3 Reconstructed pathways towards 5 compounds used as a basis for comparison of <i>P. putida</i> and <i>E. coli</i> capabilities.....	86
Table 4.4 Comparison of TFA feasible host specific pathways.....	88
Table 4.5 Yield dependency with the respect to the chosen host	89
Table 5.1 Ranges of important parameters predicted by the iSCHRUNK algorithm	114

Introduction

Motivation

In 2009, Shafiee and Topal, in one of the most impactful publications in the field of energy production & sources (1289 citations based on the Google scholar in November 2018), estimated fossil fuel reserve depletion times for oil, coal and gas to approximately 35, 107 and 37 years, respectively [1]. More recently, in June 2017, British Petroleum (BP) reported that the proved oil reserves rose to the total of 1707 billion barrels – just enough oil to sustain the world consumption for 50.6 years at 2016 production rates [2]. Despite the fact that in reality oil reserves may be higher, mankind has to find a solution to its fossil fuel dependency. As a part of that solution, biofuels will play a big role and they will represent an essential contribution to our future energy supply [3].

The idea of producing biofuels and biochemicals from biomass-derived sugars can be traced back to the 1970's [4]. A series of global fossil fuel crises between 1973 and 1979 sparked an intensive development of technologies for production of biofuels. Since then, the first generation of biofuels (mainly ethanol and biodiesel) has been produced at industrial scale. According to the Renewable Fuel Association (RFA) [5], the world production of ethanol in 2017 was around 27 million gallons (102 million liters). However, ethanol is not an ideal replacement for conventional fuels mainly because both food and biofuel industry use corn as one of the principal production resources. Limited arable land for food crops and the expected growth of the world population (according to UN the world population will reach 9.6 billion by 2050) will further toughen up the competition for resources and motivate the development of technologies for sustainable production of food and fuels.

Motivated by ethical (fuel vs. food), and practical (relatively low energy density of ethanol compared to the fossil fuels and the highly hygroscopic nature of ethanol) reasons, recent research efforts focus on a sustainable production of the 2nd generation of biofuels. Compared to the currently used fossil fuels and bioethanol, these 2nd generation biofuels should provide lower carbon emissions, higher energy density, and should be less corrosive to engines and distribution infrastructures. Recently, a large number of potential candidates for the 2nd generation biofuels has been proposed such as n-butanol [6], isobutanol [6], 2-methyl-1-butanol [6], 3-methyl-1-butanol [6], C₁₃ to C₁₇ mixtures of alkanes and alkenes [7], fatty esters, fatty alcohols [8], and Methyl Ethyl Ketone (MEK) [9].

While certain candidates for 2nd generation biofuels are native to some living cells, their production does not occur in appreciable quantities [10]. In this case, metabolic engineering strategies applied in native organisms can target rate limiting steps and allow an enhanced production. For chemicals whose natural microbial producers are not known, the feasibility of their bioproduction has to be assessed and potential novel biosynthetic pathways for production of these chemicals are yet to be discovered [11, 12]. Even when production pathways for target chemicals are known, it is important to find alternatives in order to further reduce cost and greenhouse gas emissions, and as well to avoid possible patent issues.

Nowadays, knowledge assembled in the fields of metabolic engineering and systems biology provide the tools to design hybrid organisms with novel biosynthetic capabilities and suitable production properties [13]. The ideal microorganism for biofuel production should possess high substrate utilization and processing capacities, fast and deregulated pathways for sugar transport, good tolerance to inhibitors and product, high metabolic fluxes, and should lead to the production of a single fermentation product [14]. The most frequently suggested candidates for the host organisms are prokaryotic *Escherichia coli* and eukaryotic yeast *Saccharomyces cerevisiae* [15-17]. These organisms have fast growth rates and are facultative anaerobes, allowing a flexible and economical process design for large-scale production[16]. However, during the expression of non-native pathways, several important factors have to be taken into account: the affinity of an enzyme for substrate(s) and its catalytic efficiency[18], the NADH/NAD⁺ imbalance[15] that may

result from the introduction of non-native pathways and their expression, additional stresses caused by the toxicity of the non-native product [14] etc.

Recently, *Pseudomonas putida* emerged as a good alternative due to its fast growth with low nutrient demand [19], considerable metabolic versatility [20], low cellular energy demand [21], ability to grow in wide range of chemicals [22, 23] and its robustness and high flexibility to adapt and counteract different stresses [24]. In addition, *P. putida* is suitable for genetic manipulation [25]. For these reasons, *P. putida* rose as one of the most promising production hosts for a wide range of chemicals including biofuels.

A systematic strain design for the production of biochemicals usually requires several gene insertions/deletions, measurements of product titers and different omics data to assess the performance of engineered strains. Combinatorial identification of target genes and rate-limiting steps require inclusion of this information in a mathematical model. To this end, mathematical models of metabolism serve as one of the essential tools for designing recombinant organisms for the production of biofuels. In recent years, the prevalent frameworks for modeling metabolic pathways were constraint-based approaches that make use of network stoichiometry to characterize the intracellular fluxes at steady state [26-29]. While proving their utility in studies of cellular physiology and metabolic engineering [30-32], the stoichiometric models lack information about metabolic regulation and enzyme kinetics. Therefore, these static descriptions are unable to capture the dynamic features of metabolic pathways and they cannot be used for predicting the complex dynamic responses to environmental and genetic perturbations, or, e.g., for studying dynamic transitions of the metabolism [33-35] or oscillatory phenomena [36, 37]. Such studies require a kinetic model. The overarching ambition of kinetic metabolic modeling is to capture the dynamic behavior of metabolism to such an extent that systems and synthetic biology strategies can reliably be tested *in silico*.

Aim and scope

This PhD thesis is a part of SynPath, a transnational research project between four universities¹ under the ERA-SynBio initiative. The aim of this project was to develop a workflow for the design and integration of **novel** synthetic biochemical pathways into two **chassis**: *Escherichia coli* and *Pseudomonas putida* for the production of **novel** molecules. This workflow was tested for the case of production of novel biofuels.

In spite of *Pseudomonas putida* versatility and superior stress tolerance compared to other industrially-relevant organisms such as *E. coli* [38], this organism is rarely used as a host for the production of biofuels. The principal objective of this thesis is to investigate metabolic capabilities of this organism as a prospective host for the production of second-generation biofuels. To this end, we performed series of computational analysis of *Pseudomonas putida* metabolism, assessed its capacity for producing biofuels and compared it to the one of *Escherichia coli*. This way, we were able to underscore benefits and drawbacks of using *Pseudomonas putida* as an alternative to *Escherichia coli* for biofuel production.

Thesis outline

This thesis is organized in 5 chapters. Each chapter contains the state-of-the-art related to the chapter topic.

In **chapter 1**, we introduced some basic concepts and methods such as constrained based modeling and Genome-Scale Models which appear throughout the whole thesis. In **chapter 2**, we present the first thermodynamically curated genome-scale model of *P. putida* and its three reduced counterparts of different complexity. We studied how different constraints imposed in the model can affect our conclusions about gene essentiality. In **chapter 3**, we used Biochemical Network Integrated Computational Explorer (BNICE.ch) to explore the space of potential biotransformations around promising 2nd generation biofuel candidate methyl ethyl ketone. The computational

¹ RWTH Aachen University (RWTH), Technical University of Denmark (DTU), École Polytechnique Fédérale de Lausanne (EPFL) Laboratory of Computational Systems Biotechnology (LCSB), and University of California, Berkeley & Lawrence Berkeley National Laboratory (JBEI)

techniques presented in this chapter for generating and analyzing biochemical pathways can be useful for the development and evaluation of similar novel biochemical pathways for many other compounds of biotechnological importance. Since the selection of a host organism is critical step in metabolic engineering, in **Chapter 4** we compared the metabolic capacities of *P. putida* and *E. coli* from different aspects like thermodynamics, capacity to produce different cofactors and we evaluated them as potential second-generation biofuels production hosts. We have shown that choice of the host organism has a huge impact on the biosynthetic pathway performance, e.g., pathway is thermodynamically feasible in one organism but not in the other. Finally, to identify metabolic targets for the optimization of production of desired molecules, in **Chapter 5** we built a population of large-scale kinetic models of *P. putida*. This way, we are able to identify enzymes that are potential candidates for metabolic engineering strategies towards improving metabolic capabilities of a *P. putida* strain.

Supplementary material for this thesis is available in the electronic form, and it is accessible at the following link:

https://zenodo.org/record/1545523#.W_vZzZNXhBw

Articles included in this thesis

Milenko Tokic, Noushin Hadadi, Meric Ataman, Dário Neves, Birgitta E. Ebert, Lars M. Blank, Ljubisa Miskovic and Vassily Hatzimanikatis (2018): Discovery and Evaluation of Biosynthetic Pathways for the Production of Five Methyl Ethyl Ketone Precursors, **ACS Synthetic Biology** 7 (8), pp 1858–1873

Ljubisa Miskovic, **Milenko Tokic**, Georgios Fengos and Vassily Hatzimanikatis (2015): Rites of Passage: Requirements and Standards for Building Kinetic Models of Metabolic Phenotypes, **Current Opinion in Biotechnology**, Vol 36, P 146-153

Milenko Tokic, Ljubisa Miskovic and Vassily Hatzimanikatis (2018): Large-scale kinetic metabolic models of *Pseudomonas putida* for consistent design of metabolic engineering strategies (in preparation)

Milenko Tokic, Ljubisa Miskovic and Vassily Hatzimanikatis (2018): Comparison of metabolic capabilities of *Pseudomonas putida* and *Escherichia coli* (in preparation)

Chapter 1. Background

This chapter summarizes some of the basic concepts and definitions, which propagate through whole thesis. Also, all computational tools used for production and generation of obtained results, are reviewed and conceptually described.

1.1 Genome scale models

A mathematical model is a system of equations that represents the features of a certain process. Typically, a metabolic network representing cell metabolism consists of a list of biochemical reactions and the association between these reactions and relevant proteins and genes. A reconstruction can be converted into a model by including the assumptions necessary for computational simulation, for example, maximum reaction rates and nutrient uptake rates [39]. Integration of experimental studies with metabolic modeling allows comprehension of systems properties, prediction of perturbation effects and generation of hypotheses for further research [40]. Models can be applied to generate novel, testable and often quantitative predictions of cellular behavior [39]. Analysis of the metabolic network allows us to simulate experimental conditions that are difficult or very expensive and time consuming to perform. Also, model simulations can be repeated relatively easily for different conditions.

Metabolic models that capture information about all known metabolic reactions and the genes with their respective enzymes are known as Genome Scale Models (GEMs). GEMs have been proved as very useful in the field of System Biology and they have been used as tools for drug design and personalized medicine [41], for prediction of gene manipulation targets for metabolic engineering [42-45], for understanding diverse phenotypes [46, 47] and for studying of the effect of gene knockouts [48].

1.2 Constrained-based modeling

The most commonly used computational tool for research and the understanding of cell metabolism is constrained-based modeling. Basic concepts of constraint-based modeling are shown in Figure 1.1.

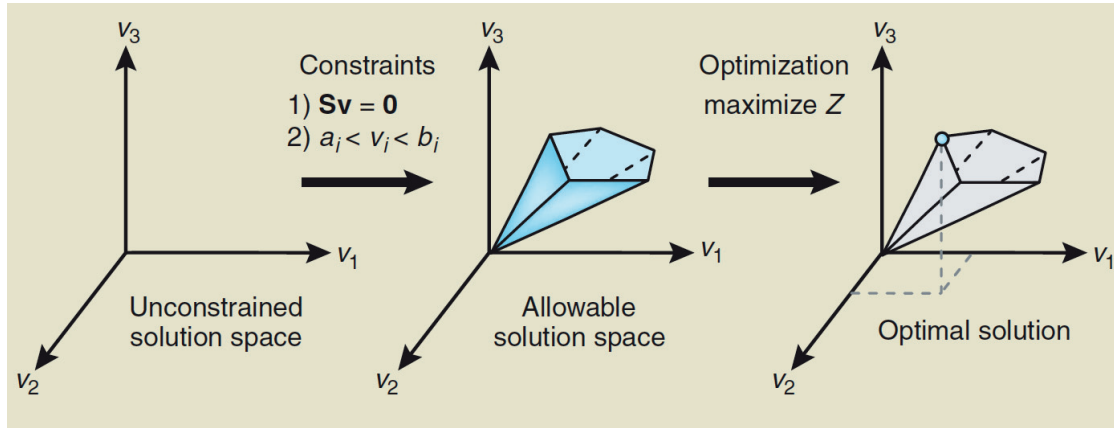


Figure 1.1 The conceptual basis of constraint-based modeling.

(Taken from Orth *et al* [49])

A metabolic network is represented as a stoichiometric matrix S . Every row of this matrix corresponds to a compound (for a system with m compounds) and every column corresponds to a reaction (n reactions). The flux of each metabolic reaction in a network is represented by the vector v . If no constraints are applied to the network the flux distribution may lie at any point of the n -dimensional solution space. An allowable solution space is defined by constraints: mass balance equations, experimental flux measurements and physiological upper and lower flux bounds (a_i and b_i). The network may acquire any flux distribution within this space, but points outside this space are denied by the constraints. For a given objective function (Z) an optimal set of fluxes can be obtained subject to the mass balance ($Sv=0$) and linear inequality ($a_i < v_i < b_i$) constraints [49, 50].

1.2.1 Flux balance analysis

The aim of Flux balance Analysis (FBA), as a constraint-based modeling approach, is to find a flux distribution within n-dimensional constrained space which will maximize or minimize an objective function Z (e.g. maximum growth or production of certain metabolite). Such flux distribution has to satisfy mass balance equations. Solutions obtained with FBA can be unreliable (although they are mathematically correct) because they can violate thermodynamic constraints (e.g. a flux is forced through a reaction into a direction that is not feasible due to the second law of thermodynamics).

1.2.2 Thermodynamics-based flux analysis

In order to address the thermodynamic feasibility, Henry *et al.* [51, 52] proposed the introduction of additional constraints into the FBA workflow in order to generate thermodynamically feasible flux distributions for a network and metabolite activity profiles. This workflow is known as thermodynamics-based flux analysis (TFA) [53-57]. Additional thermodynamics constraints reduce further the allowable solution space. This gives us the confidence that our metabolite concentrations and flux distribution (directionality of reactions) are consistent with values of the Gibbs free energy of reaction. TFA computes feasible flux profiles while taking into account thermodynamic information about Gibbs free energy of compounds and reactions. To extract thermodynamic information we use Group Contribution method (GCM) [58, 59], in which each individual reaction participant is broken down into a smaller substructures and linked with matching constituent molecular substructures from GCM with known Gibbs free energy.

Thermodynamic feasibility can be determined based on the Gibbs free energy change of reaction, which is defined as:

$$\Delta_R G'^{\circ} = \sum_{i=1}^m n_i \Delta_f G_i'^{\circ} + RT \ln \left(\prod_{i=1}^m x_i^{n_i} \right) \quad (1.1)$$

where $\Delta_f G_i'^{\circ}$ is the standard Gibbs free energy of formation of compound i , R is the universal gas constant, T is the temperature (here assumed to be 298 K), m is the

number of metabolites involved in the reaction, x_i is the activity of metabolites i , and n_i is the stoichiometric coefficient of metabolites i in the reaction (n_i is negative for reactants and positive for products). If $\Delta_R G'^{\circ}$ is smaller than 0 the reaction proceeds spontaneously and if $\Delta_R G'^{\circ}$ is bigger than 0 the reaction is not energetically favorable and will not occur spontaneously.

1.2.3 Other constrained-based methods

Beside FBA and TFA, whole gamut of constrained-based methods have been developed for studying different phenomena. Minimization of Metabolic Adjustment (MOMA) [47] was used for studying a flux redistribution in a mutants. The main assumption in MOMA is that flux redistribution in a mutant is minimal compared with the flux distribution in the wild-type. Regulatory Flux Balance Analysis (rFBA) [60], steady-state rFBA (SR-FBA) [61] and probabilistic regulation of metabolism (PROM) [62] are methods based on the Boolean rules for the genome-scale integration of transcription and metabolism. Integrated FBA (iFBA) [63] simultaneously model the metabolic, regulatory and signal transduction networks where signaling network is described by the set of ordinary differential equations. Dynamic FBA (dFBA) [64] extended FBA with the time component. Constrained Allocation Flux Balance Analysis (CAFBA) combines mass balance and proteomic constraints [65]. Dynamic enzyme-cost FBA (deFBA) [66] couples metabolism and gene expression by combining a quasi-steady state for the intracellular metabolism, and a dynamic part for the evolution of biomass and substrate concentrations.

1.3 Systems-level analysis and data integration

Strain design and production optimization traditionally rely on *in vitro* experiments and measurements. Today, modern measurement techniques are capable of measuring a wide range of intracellular (fluxes, concentrations, enzyme saturations, etc.) and fermentation data. To capture the full benefit of these data, they have to be integrated into the metabolic model. The model can use these data for further characterization and optimization of cell performance. Due to the underdetermined nature of a model, it is not possible to identify a single, unique flux profile but rather the large number of

different flux profiles. Understanding the properties of all possible flux profiles is imperative for understanding and optimizing the pathways. As mentioned above, TFA integrates the information about the intracellular metabolite concentrations allowing identifying the thermodynamic displacements from equilibrium of the network enzymes and characterizes the bioenergetic state of the cell. The integrated metabolome and fluxomic data from TFA are further used for the identification of the reactions constituting rate-limiting steps in the network [67-69].

1.4 Systematic reduction of Genome Scale Models

Since its induction, GEMs are constantly growing in their size and complexity. Along with the explosion in the number of reaction and metabolites, computational efforts for their analysis gradually increased. Although reduced models are frequently used as an alternative, they are often built around certain pathway [70-72] or subsystems of interest [37, 73-75] in *ad hoc* manner. As a consequence, their utility is hampered by the inconsistency in stoichiometry. To overcome this, Ataman *et al.* [76] proposed an algorithm called redGEM, for systematic reduction of GEMs to core models. Erdreich *et al.* [77] proposed the alternative called NetworkReducer. Both methods share the same goal: an automated and unbiased way of delivering smaller models in a systematic way. There are many advantages of reduced models compared with their GEM counterparts: reduced complexity while keeping in mind that all the properties of GEM have to be preserved, they are modular, easy expandable with a pathway of special interest and reproducible.

1.5 Optimization of synthetic pathways for efficient biofuel production, building a kinetic model

Traditional analysis of metabolic network is based on stoichiometric models and FBA/TFA. Such a formulation does not allow the prediction of metabolic responses to changes in cellular and process parameters. Such studies require a kinetic model of the metabolic network.

Reliable kinetic models require information about kinetic properties of enzymes, and quite often, this information is not available. To overcome this, Wang and Hatzimanikatis have proposed an approach based on Metabolic Control Analysis (MCA) and uses Monte Carlo techniques to simulate the uncertainty in the values of the system parameters [78, 79]. The approach laid a foundation for a computational method that integrates available information into a mathematical structure. The new method is named ORACLE (**O**ptimization and **R**isk **A**nalysis of **C**omplex **L**iving **E**ntities) [69, 80]. ORACLE methodology allows for the integration of the flux profile analysis from TFA along with metabolomics and fluxomics data and kinetic expressions for the development of large-scale kinetic models under great deal of uncertainty.

Within MCA two important variables are defined: flux control coefficients (FCC) and concentration control coefficients (CCC). Control coefficients (CC) are defined as the fractional change of metabolic fluxes and metabolite concentrations in response to fractional changes of system parameters [78, 79]. Control coefficients are determined by the elasticities, which quantify the strength of interactions of the enzymes with substrates, products, inhibitors and activators. Control coefficients capture the sensitivity of a metabolic network to a small perturbation of the metabolic parameters and they are used for understanding the properties of the metabolic pathways and to predict cellular responses [69].

The control coefficients are defined as:

- Flux control coefficient (FCC)

$$C_p^v = \frac{dv/v}{dp/p} = \frac{d \ln v}{d \ln p} = \Pi - E(NVE)^{-1}NV\Pi \quad (1.2)$$

- Concentration control coefficients (CCC)

$$C_p^x = \frac{dx/x}{dp/p} = \frac{d \ln x}{d \ln p} = -(NVE)^{-1}NV\Pi \quad (1.3)$$

where N is the stoichiometric matrix, V the diagonal matrix of fluxes, E the elasticity matrix and Π the parameter matrix; p is the parameter, x concentration and v is the flux.

MCA is a local sensitivity analysis approach. It is based on small perturbations of the parameter values around the steady state solution. We can use a kinetic model to

simulate any kind of perturbation of its components [81]. A properly built kinetic model can be used to:

- Assess possible responses of engineered strains
- Guide the design of industrial organisms
- Uncover couplings in metabolic networks

Chapter 2. Curation and systematic reduction of *Pseudomonas putida* GEM iJN1411

2.1 Introduction

Pseudomonas putida is a highly adaptive, non-pathogenic, soil bacterium that can grow on a wide range of substrates, and it is tolerant to high toxicity compounds[82]. For these reasons, it emerged recently as one of the most promising production hosts for a wide range of chemicals. Recent efforts toward understanding and improving the behavior and systemic properties of *P. putida* metabolism resulted in several genome-scale reconstructions. The first reconstructed Genome-Scale Model (GEM) of *P. putida*, iJN746, was published in 2008 and it comprised 911 metabolites, 950 reactions, and 746 genes [83]. It was rapidly followed by the publication of iJP815 [84] and several other reconstructions [85, 86]. These models have shown their value in studying metabolic features of *P. putida* such as the enhanced production of poly-hydroxyalkanoates [87], reconciliation of key biological parameters for growth on glucose under carbon-limited conditions [88], or identification of essential genes for growth on minimal medium [89]. The inconsistencies among these models motivated Yuan *et al.* to build so-called pathway-consensus model PpuQY1140 [90]. The so far most complete GEM of *P. Putida*, iJN1411, was published in 2017 by Nogales *et al.* [91], and it contains 2057 metabolites, 2581 reactions, and 1411 genes.

Thermodynamics is an important property of the metabolic network with a broad impact on the directionality of reactions [52, 55]. Methods that use thermodynamics data such as the thermodynamics-based flux analysis TFA [51-53, 55-57] allow us to integrate the metabolomics data together with the fluxomics data, to eliminate in silico designed biosynthetic pathways not obeying the second-law of thermodynamics [92-94], to eliminate infeasible thermodynamic cycles [95-97], to identify how far or how close are reactions from thermodynamic equilibrium [54, 98] or to constrain available flux and concentration spaces [53, 55]. Despite the fact that usefulness of thermodynamics has been demonstrated in many applications, only a few reconstructed GEMs consider this important network property [54, 99-102].

Analyzing GEMs can be a challenging task. Some methods like enumeration of elementary modes [103] can be even unmanageable in GEMs composed of several thousand reactions and the complexity of this method is widely recognized in the literature [104]. While this method works well for small scale models, it might fail even for the medium scale network [105]. For example, Flynn *et al.* [106] used elementary modes method to analyze metabolic model of *Shevanella oneidensis* MR1. Although their analysis was limited to a model composed of 64 reactions and 61 metabolites, they constructed 368'545 elementary modes. It's not hard to imagine an explosion in number of modes in a bigger model. On the other side, availability of intracellular metabolomics and fluxomic data is typically limited to the central carbon pathways, leaving most of the GEM uncovered. Reduced model came naturally as an alternative. While many reduced models are delivered in biased way, with *ad hoc* stoichiometry around pathway/s or subsystem/s of interest and tailored made for the "in the lab" study, Ataman *et al.* [76] proposed a systematic semi-automatic method for reduction of GEMs to core models in unbiased way

In this chapter, we used the thermodynamics-based flux analysis (TFA) [51-53, 55-57] to introduce thermodynamics into iJN1411. With TFA, we imposed thermodynamic information for 62.3% metabolites and 59.3% reactions from the model. We analyzed thermodynamic behavior of the model and we corrected the topology of the network. Finally, we applied redGEM algorithm and we delivered 3 reduced modes of different

complexity. All reduced models are validated against their GEM counterpart in terms of gene essentiality, flux variability and the growth predictions.

2.2 Methods

2.2.1 Thermodynamic-based flux analysis (TFA)

We integrated thermodynamic properties into existing stoichiometric model iJN1411 [91] using Thermodynamic-based flux analysis (TFA) framework [51-53, 55-57]. TFA computes feasible flux profiles while considering thermodynamic information about Gibbs free energy of compounds and reactions. To extract the thermodynamic information we used Group Contribution method (GCM) [58, 59], in which each individual reaction participant is broken down into a smaller substructures and linked with matching constituent molecular substructures from GCM with known Gibbs free energy.

2.2.2 Integration of metabolomics data while considering cellular compartments.

Here we propose a novel set of constraints that allow for concentrations of the same species across different compartments to be different while maintaining the consistency with the experimental measurements.

For the concentration C_M of a metabolite M measured in the range $C_M \in (\underline{C}_M, \overline{C}_M)$ we have:

$$C_M = \frac{N_t}{V_t} = \frac{\sum_i N_{Ci}}{\sum_i V_{Ci}} \quad (2.1)$$

where N_t is the number of moles of M and V_t is the total volume of the cell. N_{Ci} and V_{Ci} are the corresponding quantities in compartments i . Considering that $\sum_i V_{Ci} = V_t$, i.e., $\sum_i \frac{V_{Ci}}{V_t} = \sum_i \alpha_i = 1$, by dividing (2.1) with V_t we obtain

$$C_M = \frac{\sum_i \frac{N_{Ci} V_{Ci}}{V_t V_{Ci}}}{\sum_i \frac{V_{Ci}}{V_t}} = \frac{\sum_i \alpha_i C_{Mi}}{\sum_i \alpha_i} \quad (2.2)$$

where C_{Mi} is the concentration of metabolite M in the compartment i and α_i is the volume fraction of the compartment i with respect to the entire cell. Observe that α_i and C_{Mi} are positive quantities.

If we apply logarithm to (2.2), we have:

$$\log C_M = \log \frac{\sum_i \alpha_i C_{Mi}}{\sum_i \alpha_i}. \quad (2.3)$$

Considering that log is a concave function, we can use Jensen's inequality [107] where for a concave function φ and positive weights α_i it holds that:

$$\varphi \left(\frac{\sum_i \alpha_i x_i}{\sum_i \alpha_i} \right) \geq \frac{\sum_i \alpha_i \varphi(x_i)}{\sum_i \alpha_i}. \quad (2.4)$$

Therefore, by combining (2.3) and (2.4) we get:

$$\log C_M = \log \frac{\sum_i \alpha_i C_{Mi}}{\sum_i \alpha_i} \geq \sum_i \alpha_i \log C_{Mi}. \quad (2.5)$$

Moreover, if we denote the physiological lower and upper bound on intracellular metabolite concentrations as $LB = 1 \mu\text{M}$ and $UB = 50 \text{ mM}$, respectively, then the upper bound on C_{Mi} , $\overline{C_{Mi}}$, can be derived from the following expression:

$$\overline{C_M} = \alpha_i \overline{C_{Mi}} + (1 - \alpha_i) * LB, \quad (2.6)$$

hence

$$\overline{C_{Mi}} = \frac{\overline{C_M} + (1 - \alpha_i) * LB}{\alpha_i}. \quad (2.7)$$

To prevent the case $\overline{C_{Mi}} > UB$ for some values of α_i , we put the upper bound on $\overline{C_{Mi}}$ as follows:

$$\overline{C_{Mi}} = \min \left(\frac{\overline{C_M} + (1 - \alpha_i) * LB}{\alpha_i}, UB \right). \quad (2.8)$$

Analogously for the lower bound on the concentration of the metabolite M in the compartment i, $\underline{C_{Mi}}$, we have:

$$\underline{C_{Mi}} = \max \left(\frac{\overline{C_M} + (1 - \alpha_i) * UB}{\alpha_i}, LB \right). \quad (2.9)$$

Therefore, instead of using i constraints on the compartment species of metabolite M in the form of $\log \underline{C}_M \leq \log C_{Mi} \leq \log \overline{C}_M$, we propose to use $i+2$ constraints providing more flexibility and relaxing the assumption on equal concentrations of metabolite M in all compartments:

$$\log \underline{C}_{Mi} \leq \log C_{Mi} \leq \log \overline{C}_{Mi} \quad (2.10)$$

together with (2.5) and

$$\log \underline{C}_M \leq \log C_M \leq \log \overline{C}_M \quad (2.11)$$

where \underline{C}_{Mi} and \overline{C}_{Mi} are computed as in (2.8) and (2.9).

The volume fractions of cytosol, α_1 , and periplasm, α_2 , were taken respectively as 0.88 and 0.12 [108].

2.2.3 Gap-filling of thermodynamically curated iJN1411

We merged two genome-scale models, iJN1411 of *P. putida* [91] and iJO1366 of *E. coli* [29] into one compound model that was used for gap-filling of iJN1411. We removed duplicate reactions from the compound model along with phosphofructokinase (PFK) that is absent from *P. putida* metabolism [109]. Compared to iJN1411 the compound model had additional 1201 reactions originating from iJO1366. We imposed experimentally measured ranges of ATP concentrations, glucose uptake and the specific growth rate, and performed TFA while minimizing the number of reactions from the set of the added 1201 that can carry flux. The optimization revealed that it is sufficient to add a single reaction, sulfate adenylyltransferase (SADT₂), from iJO1366 to iJN1411 and obtain consistency of iJN1411 TFA solutions with the experimental data.

2.2.4 Systematic reduction

We used the redGEM [76] and lumpGEM [110] algorithms to deliver reduced models of three different sizes (referred in the results section as D₁, D₂ and D₃). The first step in the redGEM algorithm is to select the metabolic subsystems of interest around which the reduced models are built. We selected the following six metabolic subsystems from

iJN1411: glycolysis and gluconeogenesis, pentose phosphate pathway, pyruvate metabolism, TCA cycle and oxidative phosphorylation. From the reactions belonging to these six subsystems we removed all cofactor pairs and small metabolites such as protons, phosphate groups and inorganics. We then used a graph search algorithm to identify all one-reaction, two-reaction, and three-reaction steps pairwise connections between six subsystems and formed the core metabolic networks of D₁, D₂ and D₃ model, respectively. We next performed another graph search to find the connections of D₁-D₃ core networks with the extracellular space. With this step the core networks of D₁, D₂ and D₃ models were finalized.

We then used the lumpGEM[110] algorithm to connect the core networks of D₁, D₂ and D₃ with the building blocks of the iJN1411 biomass reaction. For each of 102 iJN1411 biomass building blocks (BBBs), lumpGEM identified a set of alternative minimal subnetworks that were able to connect precursors belonging to the core network and the BBB. The size of minimal networks is denoted S_{\min} [110]. For some studies it is of interest to identify subnetwork of higher sizes. Herein we identified subnetworks of the size $S_{\min}+2$. Finally, lumpGEM collapses the identified subnetworks into lumped reactions that together with the core networks constitute the core reduced model.

All delivered models passed a series of tests, to ensure consistency between them and GEM. These tests include thermodynamic feasibility, theoretical maximal biomass yield, gene essentiality of common genes and thermodynamic flux ranges of common reactions.

2.3 Results and Discussion

2.3.1 Thermodynamic curation of the *P. putida* GEM iJN1411

We used Group Contribution method (GCM) [58, 59] to assign the standard Gibbs free energy of formation to 62.3% metabolites and the standard Gibbs free energy of reaction to 59.3% reactions from the iJN1411 model. We could calculate the standard Gibbs free energies for all metabolites and reactions participating in the pathways of central carbon metabolism (glycolysis, gluconeogenesis, pentose phosphate pathway, tricarboxylic acid (TCA) cycle). In contrast, we could estimate the standard Gibbs free energy of reaction

for only 3.3% reactions in the poly-hydroxyalkanoates (PHA) metabolism because the majority of involved metabolites from these pathways have the structures with unknown residuals which precluded computation of the thermodynamic properties. With TFA we identified 103 bidirectional reactions, while this number was 115 when thermodynamics were not considered (Table 2.1). In this case, only constraint imposed in the model was upper glucose uptake of 10 mmol/gDCW/hr.

We integrated experimental measurements of glucose uptake and biomass yield on glucose [111] and metabolite concentrations [112] into the thermodynamically curated model iJN141. The performed TFA indicated that the model predicted ranges of ATP concentrations could not match the values reported in the literature [109, 112]. A reason behind this mismatch could lie in the fact that the H^+ /ATP stoichiometry in the electron transport chain (ETC) of *P. putida* might be inaccurately determined in iJN141 which would lead to large discrepancies in ATP yield on glucose [20, 113]. Here, we investigated another venue and hypothesized that iJN141 is missing a critical reaction in the ATP-related metabolism. Therefore, to make model predictions consistent with the experimental observations, we performed gap-filling with the iJO1366 GEM of *E. coli* [29] (Methods 2.2.3). Our analysis indicated that one reaction, sulfate adenylyltransferase (SADT2), is missing in the iJN141. SADT2 plays a role in cysteine formation, and similarly to sulfate adenylyltransferase (SADT), which already exists in the iJN141, it catalyzes the production of cysteine precursor adenosine 5'-phosphosulfate from ATP and SO_4 . The production of adenosine 5'-phosphosulfate catalyzed by SADT2 is coupled with GTP consumption, whereas this coupling is absent in SADT. Since the experimental evidence supports that GTP hydrolysis enhances the rate of adenosine 5'-phosphosulfate formation [114], we included this reaction into iJN141. The thermodynamically curated, gap-filled, model iJN141 was consistent with the experimental values of both fluxomics and metabolomics data.

Table 2.1 Reactions that are determined as bi-directional in FBA, but unidirectional in TFA

Reaction abbreviation	Reaction Name
3HAACOAT ₁₂₀	3 Hydroxyacyl ACPCoA Transacylase
3HAACOAT ₆₀	3 Hydroxyacyl ACPCoA Transacylase
3HAACOAT ₈₀	3 Hydroxyacyl ACPCoA Transacylase
FALDM	formaldehyde dismutase
RECOAH ₂	3 hydroxyacyl Coa dehydratase 3R 3 hydroxyhexanoyl CoA
RECOAH ₃	3 hydroxyacyl Coa dehydratase 3R 3 hydroxyoctanoyl CoA
RECOAH ₅	3 hydroxyacyl Coa dehydratase 3R 3 hydroxydodecanoyl CoA
RHACOAR ₁₀₀	3R 3 Hydroxyacyl CoANADP oxidoreductase
RHACOAR ₁₂₀	R Hydroxyacyl CoANADP oxidoreductase
RHACOAR ₁₄₀	3R 3 Hydroxyacyl CoANADP oxidoreductase
RHACOAR ₆₀	3R 3 Hydroxyacyl CoANADP oxidoreductase
RHACOAR ₈₀	3R 3 Hydroxyacyl CoANADP oxidoreductase

Table 2.2 Additionally constrained reactions upon implementation of experimental data

Reaction abbreviation	Reaction Name
ALCD19	alcohol dehydrogenase glycerol
ATPS4rpp	ATP synthase four protons for one ATP periplasm
CO2tex	CO ₂ transport via diffusion extracellular to periplasm
CO2tpp	CO ₂ transporter via diffusion periplasm
DM_co2_e	CO ₂ exchange
DM_h2o_e	H ₂ O exchange
FALDtex	formaldehyde transport via diffusion extracellular to periplasm
FALDtpp	formaldehyde transport via diffusion periplasm
FMNR2r	FMN reductase
G3PD2	glycerol 3 phosphate dehydrogenase NADP
GLYALDDr	D Glyceraldehyde dehydrogenase
H2Otex	H ₂ O transport via diffusion extracellular to periplasm
H2Otpp	H ₂ O transport via diffusion periplasm
MTHFC	methenyltetrahydrofolate cyclohydrolase
MTHFD	methylenetetrahydrofolate dehydrogenase NADP
TPI	triose phosphate isomerase

The integration of thermodynamics data into models restricts the available flux and concentration spaces [53, 55] because thermodynamics determines the directionality in which reactions can operate [52, 55]. With the Flux Balance Analysis (FBA) we found that 108 reactions could operate in both forward and reverse direction (bi-directional reactions) while still being consistent with the integrated fluxomics data [111]. We used TFA to integrate additional metabolomics data, and we found that there were 87 bi-

directional reactions. This means that 21 reactions that were bi-directional with the stoichiometric constraints could not operate in both directions due to thermodynamic constraints. Compared with unconstrained TFA case, 16 additional reactions were determined as unidirectional (Table 2.2).

2.3.2 Core reduced stoichiometric models of *P. putida*

We next performed redGEM algorithm [76] on the thermodynamically curated GEM of *P. putida*, iJN1411 [91], to generate three different reduced models of different complexity. Firstly, we selected a set of starting central carbon subsystems (glycolysis, gluconeogenesis, pyruvate metabolism, pentose phosphate pathway, TCA cycle and oxidative phosphorylation), as an initial set of subsystems around which we did expansion (Fig 2.1). These subsystems include 90 reactions and 99 metabolites. Next, we performed the intra-expansion between subsystems in a pairwise manner with respect to a different degree of connectivity D (Methods 2.2.4).

The simplest out of three core models (subsequently referred to as D₁) contained the D₁ core network formed by the reactions and metabolites from the six subsystems and the reactions that belonged to one-reaction-step pairwise connections between these six subsystems (Fig 2.1). Most of them were around L-aspartate node like L aspartate 3 hydroxylase (ASP₃H) and L aspartate oxidase (ASPO₅), which connect L-aspartate, and succinate or aspartate transaminase (ASPTA), which connects L-aspartate with oxaloacetate. The algorithm also added two glutamate dehydrogenases (one uses NADH (GLUD_x) and second NADPH (GLUD_y)) as a direct link between 2-oxoglutarate and L-glutamate, which is a precursor of one of the *P. putida* biomass building blocks L-glutamine. It also added a direct link between 2-oxoglutarate and succinate via Kdo 2 hexa acyl lipid A hydroxylase (LIPAH). Finally redGEM added reactions whose reactants are only cofactors like adenylate kinase (ADK₁), NAD kinase (NADK), nucleoside triphosphatase ATP (NTP₁), Nucleoside triphosphate pyrophosphorylase ATP (NTPP₆) or NADH peroxidase (NADHPO).

In a medium complexity core model, D₂, algorithm added many two reaction step connections between initial subsystems such as link between acetyl-CoA and acetate via acetaldehyde dehydrogenase (acetylating) (ACALD) and aldehyde dehydrogenase

(acetaldehyde, NAD) (ALDD2x) through intermediate metabolite acetaldehyde or link between phosphoenolpyruvate and 3-dehydroquinate via 3 deoxy D arabino heptulosonate 7 phosphate synthetase (DDPA) and 3 dehydroquinate synthase (DHQS) through intermediate metabolite 2-dehydro-3-deoxy-D-arabino-heptonate 7-phosphate (Fig 2.1). The biggest number of added reactions were connections between L aspartate and fumarate like phosphoribosylaminoimidazolesuccinocarboxamide synthase (PRASCSi) and adenylosuccinate lyase (ADSL2r). Finally, algorithm added reactions around GTP and GTP nodes like adenylylate kinase GTP (ADK3), nucleoside triphosphatase GTP (NTP3) and nucleoside diphosphate kinase ATP GDP (NDPK1).

The core network of the highest complexity model, D3, included reactions around chorismate node from phenylalanine tyrosine tryptophan biosynthesis pathway such as 3 phosphoshikimate 1 carboxyvinyltransferase (PSCVT), chorismate synthase (CHORS), anthranilate synthase (ANS) (Figure 2.1). It also added three-reaction step connection between 3-phospho-D-glycerate and glyoxylate (glycerate kinase (GLYCK), tartronate semialdehyde reductase (TRSARr) and glyoxalate carboligase (GLXCL)). The biggest number of reactions was added around glycerol-3-phosphate and glycerol nodes (not shown in the figure). Starting from the initial set of 7 subsystems (including extracellular metabolites as the extracellular subsystem), by network expansion, reactions from D3 covered in total 38 different subsystems. In GEM there are 89 unique subsystems which mean that only with three-step connections we were able to cover more than 40 % of all systems defined in GEM.

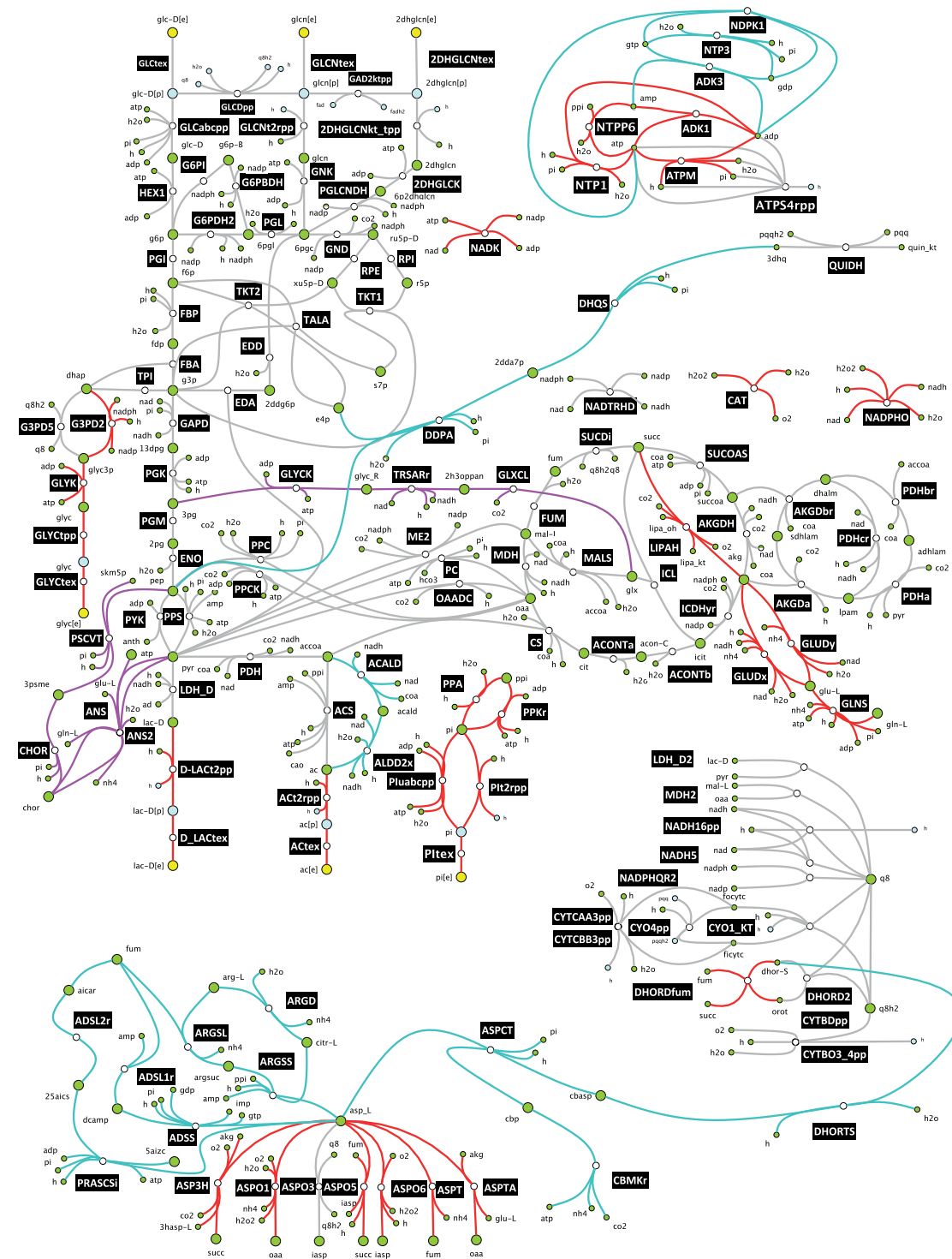


Figure 2.1 The core networks generated by the redGEM algorithm from iJN1411 genome-scale model

The core network was built around reactions (grey) that belong to the six subsystems of central carbon metabolism (glycolysis and gluconeogenesis, pentose phosphate pathway, pyruvate metabolism, TCA cycle and oxidative phosphorylation). Reactions belonging to one-reaction-step, two-reaction-step, and three-reaction-step connections between the six subsystems are marked in red, cyan and magenta, respectively.

2.3.3 Generation of lumped reaction for production of biomass building blocks

Next step in building systematically reduced models is to connect systematically reduced core networks with the biomass building blocks (BBBs). The biomass reaction in iJN1411 contains 102 biomass building blocks. To connect BBBs with systematically reduced core networks we used redGEM [76] accompanying algorithm called lumpGEM [110], which identifies a subnetwork and its elementally balanced lumped reactions that can produce a BBB from the precursors that belong to the systematically reduced core network. By its definition, lumpGEM is identifying a subnetwork through minimization of the number of reactions that don't belong to the core network (non-core reactions), which is able to connect a precursor belonging to the core network and a BBB. This minimal number of reactions is called S . It is also able to identify all the possible alternatives of the size S , but also alternatives with the higher S . For example, for production of 10-formyltetrahydrofolate from the D1 core network, the minimal number of non-core reactions is 41 (S_{min}). There are 8 different subnetworks of size 41, but only 4 unique lumped reactions. If we go to the higher S , there are 20 subnetworks of size 42 (S_{min} plus 1) with 10 unique lumped reactions and 10 subnetworks of size 43 (S_{min} plus 2) with 7 unique lumped reactions. On contrary, for production of L-alanine, only one non-core reaction is enough (L alanine transaminase). For this case, $S_{min} = 1$. In this work, we generated all possible alternative subnetworks up to the size of S_{min} plus 2, for all systematically reduced core network (D1 to D3). In total, we added 550 unique reactions generated by lumpGEM in the D1, 397 in D2 and 407 in D3 (Table 2.3).

Table 2.3 An overview of the delivered reduced models

	D1	D2	D3
Reactions	828	704	750
<i>Core</i>	278	307	343
<i>Lumped</i>	550	397	407
% of reactions with estimated standard Gibbs free energy	70.4	70.8	62.3
Metabolites	286	306	336
<i>Cytosolic</i>	156	174	200
<i>Periplasmic</i>	70	71	74
<i>Extracellular</i>	60	61	62
% of metabolites with estimated standard Gibbs free energy	81.8	82.7	82.1

2.3.4 Consistency checks of core reduced models

An important step in the generating systematically reduced model is checking their properties against their GEM counterpart. In other words, we have to be assured that our reduced model behaves exactly in the same way as GEM. We performed a battery of tests to validate the consistency of the systemic properties of the core reduced models D1, D2 and D3 with their GEM counterpart, iJN1411. Here we present and discuss results for D2.

We first performed FBA and TFA for the glucose uptake of 10 mmol/gDCW/hr, and we found the identical maximum specific growth rate of $\mu=0.939 \text{ h}^{-1}$ for both D2 and iJN1411, meaning that D2 was able to capture well the physiology of the growth on glucose.

We then carried out the comparison of essential genes between D2 and GEM. *In silico* gene deletion represents one of the most common analysis of metabolic networks, and it is used to assess the predictive potential of the model [83] or to identify main genetic targets for strain engineering [87, 115]. Out of 314 genes that D2 shared with GEM, we identified 47 as *in silico* essential. Out of these 47, 36 were essential in both D2 and GEM and 11 were essential in D2 only. These 11 genes were essential in D2 because this model was missing some of the alternative pathways from GEM. For example, aceF PP_0338 (encoding for acetyltransferase component of pyruvate dehydrogenase complex) and aceE PP_0339 (encoding for pyruvate dehydrogenase, E1 component) are essential in D2 because they encode for enzymes necessary for synthesizing acetyl-CoA from pyruvate, whereas GEM contains additional alternative pathways for this synthesis. Interestingly, among the 11 genes is the gene tpiA PP_4715 that encodes for triose-phosphate isomerase although it is reported as essential in the literature [116].

A third criterion to assess compatibility between GEM and D2 is thermodynamic-based flux variability analysis (TVA). We performed TVA on all common reactions between GEM and D2 to compare allowable flux ranges. Although for the majority of the reactions we obtained consistent flux ranges, there are some reactions that had reduced flexibility in D2 compared with its GEM counterpart (Figure 2.2). The majority of these reactions were in the upper glycolysis (GAD2ktp (gluconate 2 dehydrogenase periplasm), GLCDpp (glucose dehydrogenase), HEX_1 (hexokinase), GNK (gluconokinase) and gluconeogenesis (PGK (phosphoglycerate kinase), PGM (phosphoglycerate mutase), ENO (enolase)). Reason for this lies in the absence of some subsystems in the D2. In this case additional flexibility in these reactions in the GEM comes from the starch and sucrose metabolism and cell envelope biosynthesis cellulose metabolism, which are absent in the D2 (Figure 2.3).

Final step in the validation is concentration variability analysis with TFA on shared metabolites between D2 and GEM. In this validation step, we didn't see any discrepancy between the two. Similar result was reported for the case of *E. coli* where the discrepancy for concentration ranges was reported for only few metabolites [76]. Detailed consistency checks for D1, D2 and D3 can be consulted in Appendix Chapter 2.

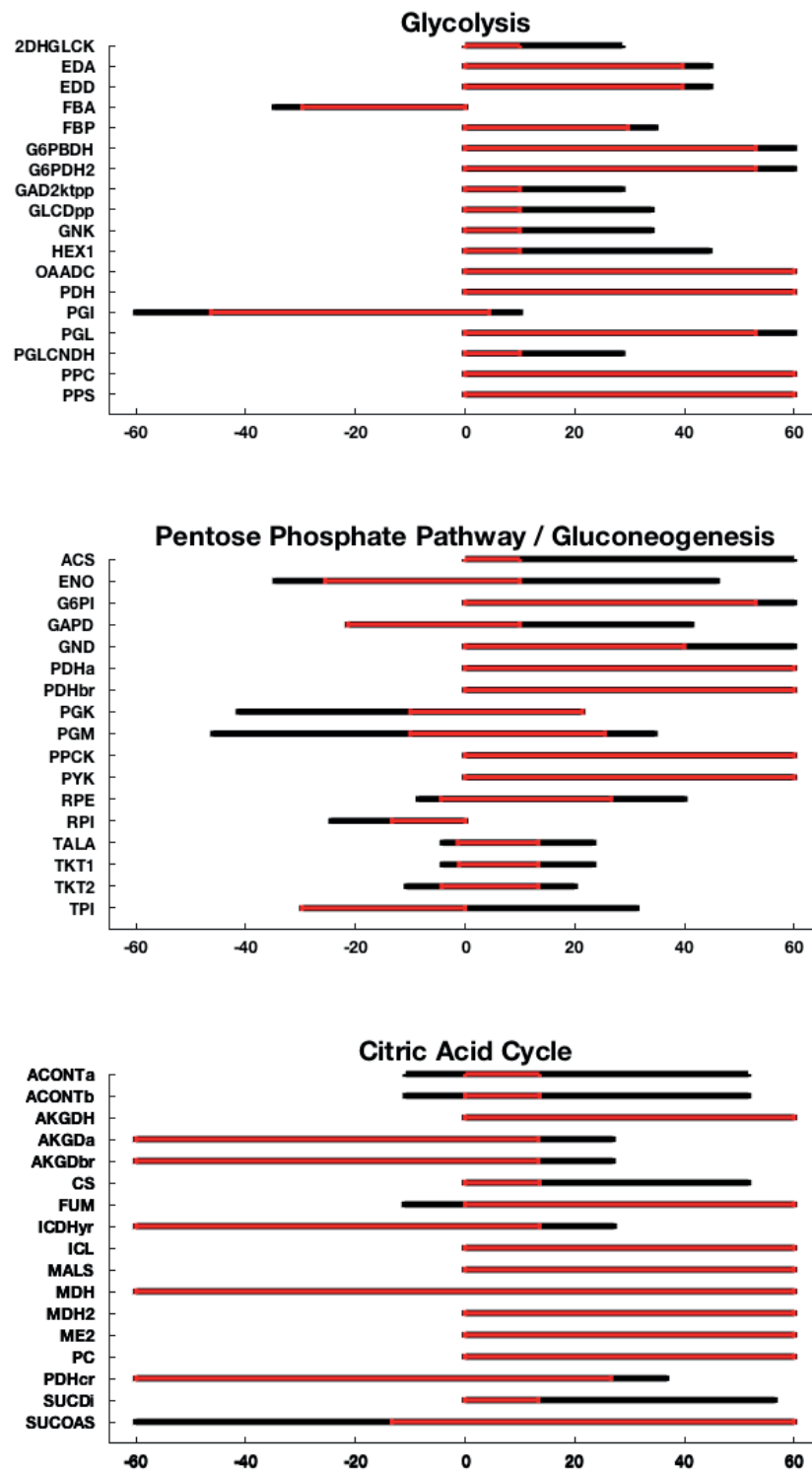


Figure 2.2 Flux variability of reactions in 4 starting subsystems in D2 and GEM
Red line FVA of reduced model, black lines FVA of GEM

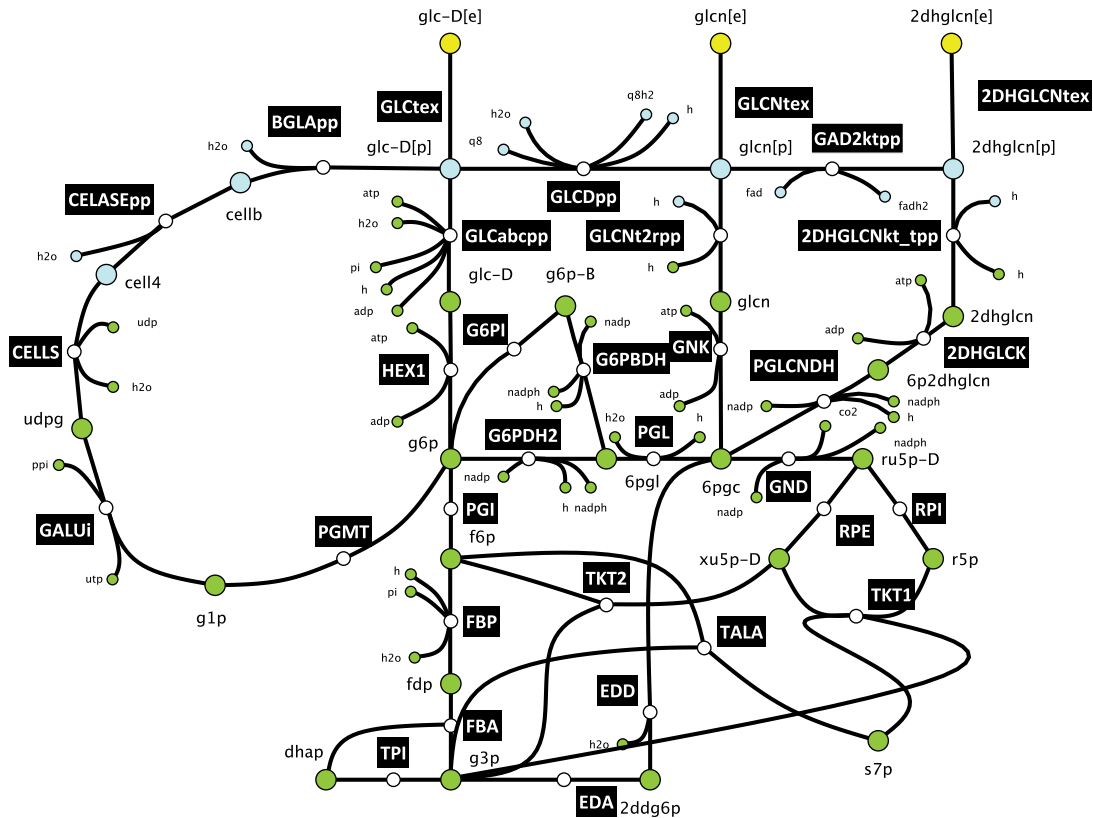


Figure 2.3 Starch and sucrose metabolism and cell envelope biosynthesis cellulose metabolism from GEM

Reactions belonging to these subsystems: PGMT (phosphoglucomutase), GALUi (UTP glucose 1 phosphate uridylyltransferase irreversible), CELS (cellulose synthase UDP forming), CELSEpp (Endo 1 4 D glucanase cellulose), BGLApp (beta glucosidase periplasmic)

2.3.5 Essentiality of genes encoding for EDA and EDD

Neither D2 nor GEM could predict experimentally observed essentiality of genes from the Entner-Doudoroff (ED) pathway. ED pathway is essential for the growth of *P. putida* on glucose, which is experimentally confirmed by the absence of the growth in mutants lacking the key enzymes 2-dehydro-3-deoxy-phosphogluconate aldolase (EDA) and 6-phosphogluconate dehydratase (EDD) [111, 116, 117]. *In silico*, these genes are not essential [89] because the model can replenish the pool of triose phosphates through pentose phosphate pathway.

We analyzed how the directionalities of reactions from the pentose phosphate pathway impact the essentiality of EDA and EDD in D2. We found that the directionalities of three reactions that have glyceraldehyde 3-phosphate (g3p) as reactant (transaldolase, TALA, and two transketolases, TKT1 and TKT2) determine if EDD and EDA are *in silico* essential. When directionality of TKT2 was imposed towards production of g3p, TALA and TKT1 became exclusively unidirectional towards consumption of g3p and production of g3p, respectfully (Fig. 2.4a), and EDA and EDD were not essential. In contrast, when TKT2 operated towards consumption of g3p EDA and EDD were essential regardless the directionality of the other two reactions (Fig 2.4b). Therefore, to ensure the consistency of *in silico* and experimentally observed gene essentiality of EDD and EDA in the subsequent studies we imposed the directionality of TKT2 towards consumption of g3p. (Fig 2.4b).

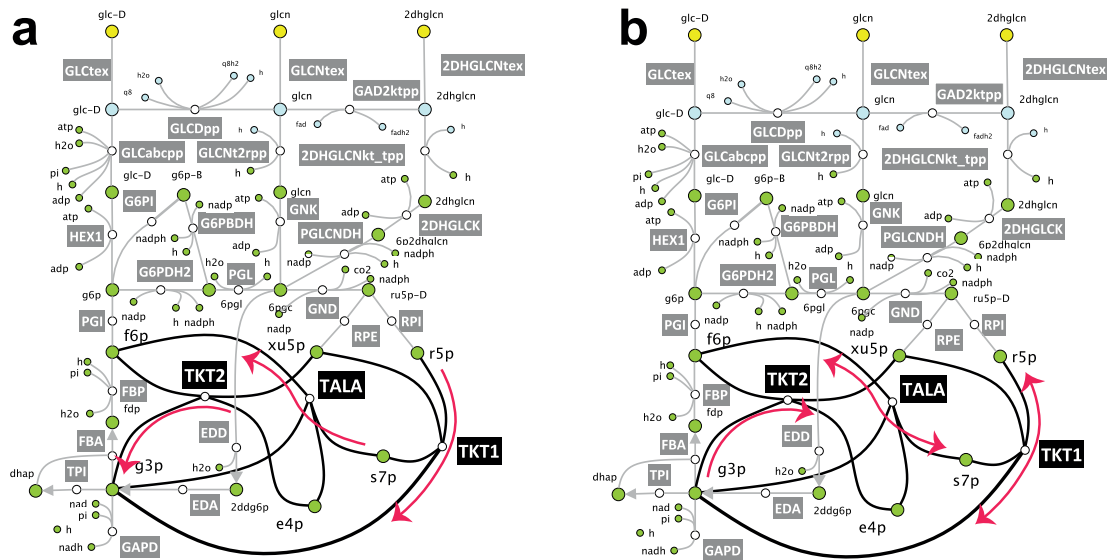


Figure 2.4 The directionality of transketolase 2 (TKT2) impacts the *in silico* essentiality of two genes encoding enzymes EDD and EDA from the Entner-Doudoroff pathway (a) if TKT2 operates towards production of g3p, then due to the stoichiometric coupling transketolase 1 (TKT1) and transaldolase (TALA) are unidirectional and EDD and EDA are not *in silico* essential. (b) if TKT2 operates towards consumption of g3p, EDD and EDA are essential irrespective of the directionalities of TKT1 and TALA.

2.4 Conclusions

In this chapter, we presented the first thermodynamically curated genome-scale model of *P. putida*. Thermodynamic curation makes the curated GEM iJN1411 amenable for integrating metabolomics data in a wide gamut of studies. The integration of thermodynamics data into models restricts the available flux and concentration spaces [53, 55] because thermodynamics determines the directionality in which reactions can operate [52, 55]. For example, Flux Balance Analysis (FBA) of iJN1411 revealed that 108 reactions could operate in both forward and reverse direction (bi-directional reactions) while still being consistent with the integrated fluxomics data [111]. We found that there were 87 bi-directional reactions when additional metabolomics data [112] were integrated with TFA, meaning that 21 reactions that were bi-directional with the stoichiometric constraints could not operate in both directions due to thermodynamic constraints.

To improve the integration of metabolomics data into stoichiometric models that have species in more than one compartment, we proposed a novel set of constraints. Current metabolomics measurement techniques do not allow for distinguishing concentrations of the same species in different compartments. Consequently, when integrating metabolomics data in constraint-based techniques that consider thermodynamics such as the energy balance analysis [118], the network-embedded thermodynamic analysis [119] and the thermodynamics-based flux analysis [51, 55-57] it is commonly assumed that the concentrations of a metabolite appearing in several compartments are identical and constrained within experimentally measured values. The novel constraints enable integration of metabolomics data without imposing this restrictive assumption thus allowing the metabolites that exist in several compartments to have different concentrations and still be consistent with the experimentally measured values for the whole cell. This way, we ensure that the set of possible metabolic outcomes predicted by model encompasses the actual cellular physiology.

Finally, we delivered three reduced models of different complexity in the completely unbiased way, which assures that these models can be used in different studies designed outside of this laboratory. These models captured experimentally confirmed essentiality

of previously reported as *in-silico* non-essential genes like triose-phosphate isomerase. We have shown on the example of trasketolase how different constraints imposed in a model can affect our conclusion about essentiality. This suggests that even published GEMs require constant improvement, in order to avoid false conclusions.

Chapter 3. Discovery and evaluation of biosynthetic pathways for the production of five methyl ethyl ketone precursors

3.1 Introduction

Limited reserves of oil and natural gas and the environmental issues associated with their exploitation in the production of chemicals sparked off current developments of processes that can produce the same chemicals from renewable feedstocks using microorganisms [8, 120, 121]. A fair amount of these efforts focuses on a sustainable production of the 2nd generation biofuels.

As it was discussed in the Introduction, these 2nd generation biofuels should outperform the currently used fossil fuels and bioethanol in terms of energy density, carbon emission and should have lower hygroscopicity. Although some of these compounds that can satisfy above-mentioned criteria, like branched-chain higher alcohols, were detected in living cells, they cannot be synthesized economically using native organisms [10, 16]. For example, *Clostridium acetobutylicum* is a native producer of 1-butanol, but its high yield production is hampered with various byproducts like butyrate or acetone [122]. In addition, *Clostridium acetobutylicum*'s complex physiology hinders possible further metabolic improvements [123].

On the other hand, the high production yield can be achieved in a more suitable host using a synthetic pathway. In this respect, *Escherichia coli* is probably the best-studied and well-characterized microorganism. Hanai *et al.* [124] engineered a synthetic pathway in *Escherichia coli* to produce isopropanol by expressing various combinations of genes from *Clostridium acetobutylicum* ATCC 824, *Escherichia coli* K-12 MG1655, *Clostridium beijerinckii* NRRL B593, and *Thermoanaerobacter brockii* HTD4. In the production phase they achieved a yield of 43.5 % mol/mol. In a similar manner, Atsumi *et al.* used *Escherichia coli* as a host for 1-butanol production [123]. Other successful examples include compounds such as (S)-3-methyl-1-pentanol [125] and 3-methyl-1-butanol [126].

For chemicals whose microbial producers are not known, novel biosynthetic pathways have to be designed [11, 12]. Pathway discovery and its experimental verification is not an easy task. Typically, it takes multiple gene insertions and deletions (usually three to four) to achieve desired phenotype [127]. Even if a production pathway already exists and it is operative in a certain organism, there is no guarantee that expression of this pathway will be successful in some other host (e.g. can violate thermodynamics). Experimental verification can be costly and time consuming. Also, there are just a few examples in which pathways are operating on the edge of the theoretical maximum [128]. Pathway could operate on theoretical maximum in one organism but in the other is not so efficient. Also, it is important to find alternatives to bypass existing patents and further reduce operating costs.

Computational approaches provide valuable assistance in the design of novel biosynthetic pathways because they allow exhaustive generation of alternative novel biosynthetic pathways and evaluation of their properties and prospects for producing target chemicals [12]. For instance, computational tools can be used to assess, prior to experimental pathway implementation, the performance of a biosynthetic pathway operating in one organism across other host organisms.

There are different computational tools for pathway prediction available in the literature [11, 129-138]. An important class of these tools is based on the concept of generalized enzyme reaction rules, which were introduced by Hatzimanikatis and co-workers [139, 140]. These rules emulate the functions of enzymes, and they can be used to predict in

silico biotransformations over a wide range of substrates [12]. Most of the implementations of this concept appear in the context of retrobiosynthesis, where the algorithm generates all possible pathways by starting from a target compound and moving backward towards desired precursors [11, 12, 121, 129, 133, 135, 138-144].

In this chapter, we used the retrobiosynthesis framework of BNICE.ch [12, 129, 139-144] to explore the biotransformation space around Methyl Ethyl Ketone (MEK). Besides acetone, MEK is the most commercially produced ketone with broad applications as a solvent for paints and adhesives and as a plastic welding agent [145]. MEK shows superior characteristics compared to conventional gasoline and ethanol in terms of its thermo-physical properties, increased combustion stability at low engine load, and cold boundary conditions, while decreasing particle emissions [146]. There is no known native microbial producer of MEK, but in the recent studies this molecule was produced in *E. coli* [147, 148] and *S. cerevisiae* [9] by introducing novel biosynthetic pathways. To convert 2,3-butanediol to MEK, Yoneda *et al.* [147] introduced into *E. coli* a B-12 dependent glycerol dehydratase from *Klebsiella pneumoniae*. Srirangan *et al.* [148] expressed in *E. coli* a set of promiscuous ketothiolases from *Cupriavidus necator* to form 3-ketovaleryl-CoA, and they further converted this molecule to MEK by expressing acetoacetyl-CoA:acetate/butyrate:CoA transferase and acetoacetate decarboxylase from *Clostridium acetobutylicum*. In *S. cerevisiae*, Ghiaci *et al.* [9] expressed a B₁₂-dependent diol dehydratase from *Lactobacillus reuteri* to convert 2,3-butanediol to MEK. Alternatively, hybrid biochemical/chemical approaches were proposed where precursors of MEK were biologically produced through fermentations and then catalytic processes were used to produce MEK [149, 150].

We used the BNICE.ch algorithm to generate a network of potential biochemical reactions around MEK, and we identified 1'325 (159 biochemical and 1'166 chemical) compounds one reaction step away from MEK (Appendix Chapter 3 Table S1). We considered as biochemical compounds the ones that we found in the KEGG [151, 152] database, and as chemical compounds the ones that we found in the PubChem [153, 154] but not in the KEGG database. A set of 154 compounds appeared in both databases. Out of these 1'325 compounds, 2-hydroxy-2-methyl-butanenitrile (MEKCNH) was the only KEGG compound connected to MEK through a KEGG reaction (KEGG R09358). For

further study, we chose MEKCNH along with three KEGG compounds: 3-oxopentanoate (3OXPNT), but-3-en-2-one (MVK) and butylamine (BuNH₂), and one PubChem compound: 1-en-2-olate (1B2OT). The latter four compounds were chosen based on two important properties: (i) their simple chemical conversion to MEK, e.g., 3OXPNT spontaneously decarboxylates to MEK; and (ii) their potential use as precursor metabolites to further produce a range of other valuable chemicals [155-157]. MVK can be converted to MEK by a 2-enoate reductase from *Pseudomonas putida*, *Kluyveromyces lactis* or *Yersinia bercovieri*, [158] however, these reactions are not catalogued in KEGG. Similarly, 3OXPNT can be decarboxylated to MEK by acetoacetate decarboxylase from *Clostridium acetobutylicum* [148]. In contrast, there are no known enzymes that can convert 1B2OT and BuNH₂ to MEK.

We have reconstructed all possible novel biosynthetic pathways (3'679'610 in total) up to a length of 4 reaction steps from the central carbon metabolites of *E. coli* towards the 5 compounds mentioned above. We evaluated the feasibility of these 3'679'610 pathways with respect to the mass and energy balance, and we found 18'662 thermodynamically feasible pathways which we further ranked with respect to their carbon yields. We identified the metabolic subnetworks that were carrying fluxes when the optimal yields were attained, and we determined the minimal sets of precursors and the common routes and enzymes for production of the target compounds.

3.2 Methods

The BNICE.ch framework [12, 129, 139-144] was employed to generate biosynthetic pathways towards 5 precursors of Methyl Ethyl Ketone: 3-oxopentanoate (3OXPNT), 2-hydroxy-2-methyl-butanenitrile (MEKCNH), but-3-en-2-one (MVK), 1-en-2-olate (1B2OT) and butylamine (BuNH₂). We tested the set of reconstructed pathways against thermodynamic feasibility and mass balance constraints, and discarded the pathways that were not satisfying these requirements [12]. Next, the pruned pathways were ranked based on the several criteria, such as yield, number of known reaction steps and pathway length. The steps of the employed workflow are discussed further (Figure 3.1)

3.2.1 Metabolic network generation

The retrobiosynthesis algorithm of BNICE.ch [12, 159] was applied to generate a biosynthetic network that contains all theoretically possible compounds and reactions that are up to 5 reaction steps away from MEK. The BNICE.ch network generation algorithm utilizes the expert-curated generalized enzyme reaction rules [139, 140, 160] for identifying all potential compounds and reactions that lead to the production of the target molecules. The most recent version of BNICE.ch includes 361 bidirectional generalized reaction rules capable of reconstructing more than 6'500 KEGG reactions [142] Starting from MEK and 26 cofactors required for the generalized enzyme reaction rules (Appendix Chapter 3 Table S2), we identified the reactions that lead to MEK along with its potential precursors [161].

Note that for studies where we need to generate a metabolic network that involves only KEGG compounds, mining the ATLAS of Biochemistry [142] is a more efficient procedure than using BNICE.ch retrobiosynthesis algorithm. The ATLAS of Biochemistry is a repository that contains all KEGG reactions and over 130'000 novel enzymatic reactions between KEGG compounds.

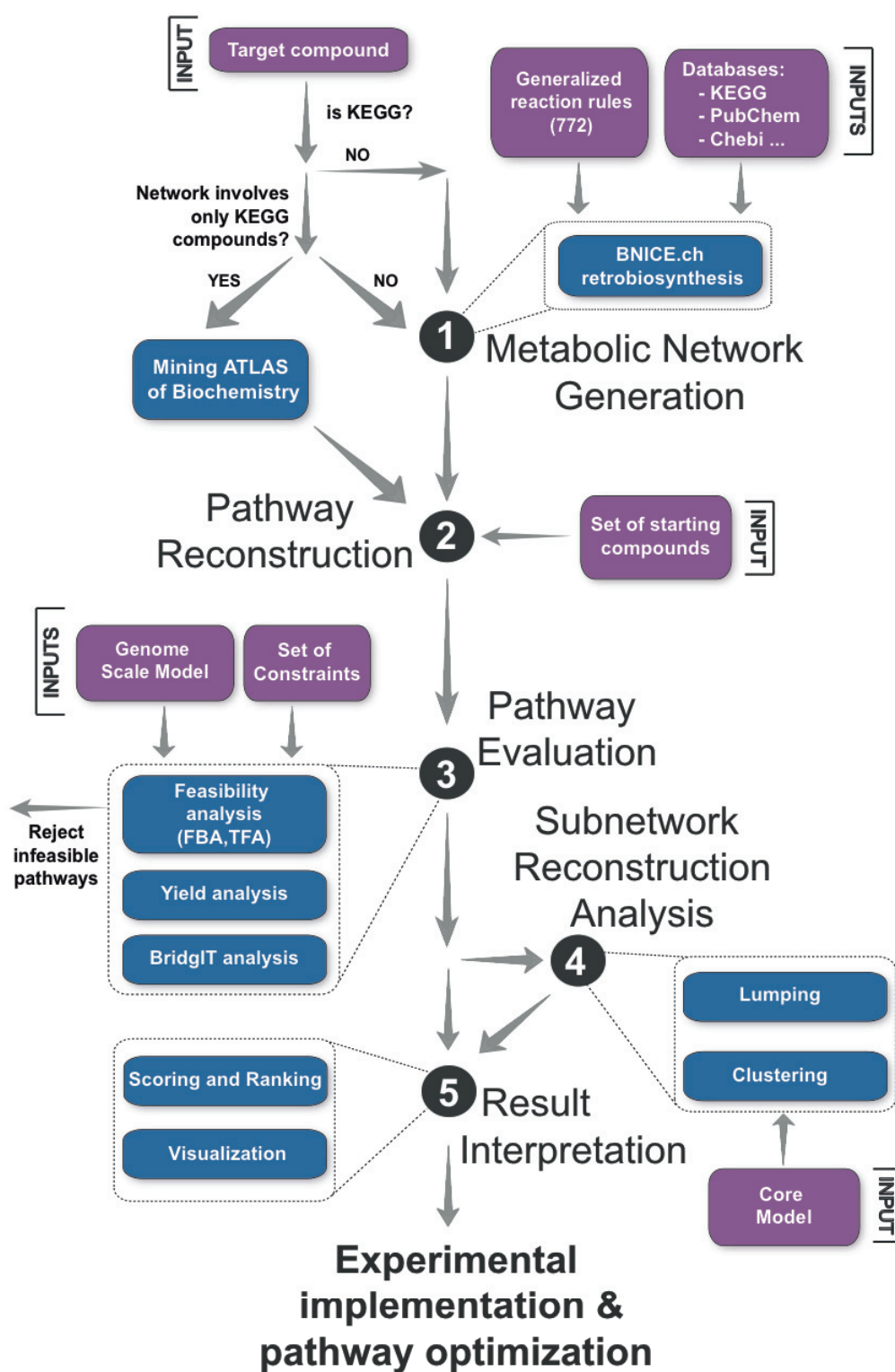


Figure 3.1 Computational pipeline for discovery, evaluation and analysis of biosynthetic pathways

3.2.2 Pathway reconstruction

The generated metabolic network was represented as a graph where the compounds were graph nodes and the reactions graph edges. A graph-based pathway search algorithm was then used to reconstruct all possible linear pathways up to the length of 4 reaction steps that connect the five target molecules with the set of 157 native *E. coli* metabolites (Appendix Chapter 3 Table S3) [29]. The search algorithm was employed as follows. Starting from a given native *E. coli* metabolite, we first searched for the shortest pathways toward a target molecule. Then, we searched for the pathways that are one reaction step longer than the shortest ones. We continued the search by gradually reconstructing the pathways of increased length, and we stopped with the 4 reaction step pathways. This procedure was repeated for all combinations of native *E. coli* metabolites and target molecules.

Note: If we were interested in pathways containing only KEGG reactions, we would perform a graph-based search over the network mined from the ATLAS of Biochemistry [142].

3.2.3 Pathway evaluation

It is crucial to identify and select, out of a vast number of generated pathways, the ones that satisfy physico-chemical constraints, such as mass balance and thermodynamics, or the ones that have an economically viable production yield of the target compounds from a carbon source. Evaluation of pathways is context-dependent, and it is important to perform it in an exact host organism model and under the same physiological conditions as the ones that will be used in the experimental implementation. Both Flux Balance Analysis (FBA) [162] and Thermodynamic-based Flux Analysis (TFA) [51-53, 55-57] were performed to evaluate the pathways. We have also used BridgIT [163] to identify candidate sequences for protein and evolutionary engineering in implementing the pathways. The availability of such sequences for the novel reactions and the ability to engineer them should also serve as a metric in ranking the feasibility of the pathways.

Flux balance and thermodynamic-based flux balance analysis. The generated pathways were embedded one at the time in the genome-scale model of *E. coli*, iJO1366 [29] and FBA and TFA were performed for each the resulting models. In these analyses, we distinguished the following types the reactions: (R_1) known and novel reactions for which have no information about their directionality; (R_2) reactions that have preassigned directionality in iJO1366; and (R_3) reactions that involve CO_2 as a metabolite. It was assumed that the only carbon source was glucose and the following two types of constraints on reaction directionalities were applied:

(C1) The preassigned reaction directionalities [164] from the iJO1366 model (R_2 reactions) were removed with the exception of ATP maintenance (ATPM), and it was assumed that the reactions that involve CO_2 (R_3 reactions) are operating in the decarboxylation direction. The lower bound on ATPM was set to 8.39 mmol/gDCW/hr. The remaining reactions (R_1 reactions) were assumed to be bi-directional for FBA, whereas for TFA the directionality of these reactions was imposed by thermodynamics. The purpose of removing preassigned reaction directionalities was to investigate alternative hypotheses about the catalytic reversibility of the enzymes. The catalytic reversibility or irreversibility of enzymes could be altered through protein and evolutionary engineering and enzyme screening [165].

(C2) The preassigned directionalities of the R_2 reactions were kept and the directionality of the R_3 reactions was fixed towards decarboxylation. R_1 reactions are left unconstrained.

Since FBA is computationally less expensive than TFA, we first performed FBA as a prescreening method to identify and discard the pathways: (i) that are not satisfying the mass balance, e.g., pathways that need co-substrates not present in the model; and (ii) that have a yield from glucose to the target compounds lower than a pre-specified threshold. In this work we used the pre-specified threshold of 0.1 mol/mol, however, this value can be chosen based on various criteria such as the economic viability of pathways. TFA was then performed on the reduced set of pathways to identify the pathways that

are bio-energetically favorable and their yields from glucose to 5 target compounds were computed under thermodynamic constraints.

BridgIT analysis. We used BridgIT [163] to associate genes to novel reactions appearing in the feasible pathways. BridgIT compares the similarity of a novel reaction to the KEGG reactions with annotated protein sequences using the information about the structures of their substrates and products, and then assigns genes of the most similar known reactions as candidates for catalyzing the novel one. BridgIT integrates the information about the structures of substrates and products of a reaction into reaction difference fingerprints [166]. These reaction fingerprints contain the information about chemical groups in substrates and products that were modified in the course of a reaction. BridgIT compares the reaction fingerprints of novel reactions to the ones of known reactions and quantifies this comparison with the Tanimoto similarity score [167]. The Tanimoto score of 1 signifies that two compared reactions had a high similarity, whereas the Tanimoto score values close to 0 signify that there was no similarity. This score was used to rank the reactions identified as similar to each of the novel reactions. The gene and protein sequences of the highest ranked reactions were proposed as candidates for either a direct experimental implementation or enzyme engineering.

3.2.4 Subnetwork reconstruction

Once the biologically feasible pathways were identified and ranked, the parts of the metabolism that carry fluxes when the target compounds are produced from glucose were analyzed. We considered that the active parts of metabolism consisted of: (i) the core metabolic network (Figure 3.4a), which included the central carbon pathways, such as glycolysis, pentose phosphate pathway, tricarboxylic acid cycle, electron transport chain; and (ii) the active anabolic subnetworks (Figure 3.4a), which contain the reactions that would carry fluxes when a target molecule is produced but did not belong to the core metabolic network. We also defined the connecting precursors as metabolites that are connecting the core and the active anabolic subnetworks (Figure 3.4.a).

The core metabolic network was derived from the genome-scale reconstruction iJO1366 [29] using the redGEM algorithm [76], and the lumpGEM [110] algorithm was then used to identify active anabolic subnetworks and to compute their lumped reactions. The

analysis of lumped reactions allowed us to identify connecting precursors of the target chemicals. We then performed clustering to uncover connecting precursors, common enzymes, and intermediate metabolites of the anabolic subnetworks leading to the production of the target chemicals.

Identification and lumping of active anabolic subnetworks. The lumpGEM algorithm was applied to identify the comprehensive set of smallest metabolic subnetworks that were stoichiometrically balanced and capable of synthesizing a target compound from a defined set of core metabolites. The set of core metabolites belongs to the core metabolic network, and it includes also cofactors, small metabolites, and inorganic metabolites (Appendix Chapter 3 Table S4). Then, for each target compound and for each identified subnetwork, we used lumpGEM to generate a corresponding lumped reaction. Within this process, the stoichiometric cost of core metabolites for the biosynthesis of these target compounds was also identified.

Clustering of subnetworks. To better understand the chemistry that leads towards the target compounds, we performed two types of clustering on the identified subnetworks:

- Clustering based on the structural similarity between the connecting precursors and byproducts of the lumped reactions. For each lumped reaction, we removed all non-carbon compounds, such as H₂, O₂, and phosphate, and the cofactor pairs, such as ATP and ADP, NAD⁺ and NADH, NADP⁺ and NADPH, flavodoxin oxidized and reduced, thioredoxin oxidized and reduced, ubiquinone and ubiquinol. This way, a set of substrates (connecting precursors) and byproducts of interest was created for each lumped reaction. We then used the *msim* algorithm from the RxnSim[168] tool to compare the lumped reactions based on individual similarities of their connecting precursors and byproducts. We finally used the obtained similarity scores to perform the clustering.
- Clustering based on the structural similarity between reactions that constitute the anabolic subnetworks. BridgIT was used to compute structural fingerprints of reactions that constitute the anabolic subnetworks, and we then performed a pairwise comparison of the anabolic subnetworks as follows.

For a given pair of anabolic subnetworks, a pairwise comparison of their reactions was carried out. As a comparison metric we used the Tanimoto distance of the

reaction fingerprints [167]. Based on this comparison, the pair of the most similar reactions in two subnetworks was found and the corresponding distance score was stored. This pair of reactions was then removed from comparison, and the next pair of the most similar reactions was found and their distance score was stored. We continued with this procedure until all pairs of reactions in two subnetworks were found. Whenever the number of reactions in two subnetworks was unequal, the unmatched reactions were ignored. The distance score between two compared subnetworks was formed as the sum of the distance scores of compared pairs of reactions. This procedure was repeated for all pairs of subnetworks.

We then used the computed distance scores to perform the subnetworks clustering.

3.2.5 Ranking and visualization of *in silico* pathways

In this step, we identified the pathways that were most likely to produce the target molecules. We defined the following criteria: (i) minimal number of reaction steps without promising candidate enzymes for catalyzing them; we consider that a novel reaction step has a promising candidate enzyme if the BridgIT algorithm has found its most similar known reactions with the similarity score higher than 0.3 [163]; (ii) minimal number of novel reaction steps in a pathway; (iii) maximal yield from glucose to the target molecules; (iv) minimal number of reaction steps in the production pathway; and (v) highest average similarity scores of novel reaction steps from BridgIT. For scoring and ranking the biologically meaningful pathways, we used criterion (i) as the primary ranking. Then, equally ranked pathways from the primary ranking were further ranked based on criterion (ii). Analogously, we performed the tertiary, quaternary and quinary ranking based on criteria (iii), (iv) and (v), respectively.

An expert opinion is important in choosing the pathways for implementation. One can use other ranking criteria or a different prioritization of the criteria. For example, the pathways can be first ranked based on a maximum yield and then based on a minimal number of novel reactions. The complete set of pathways is provided on <http://lcsb-databases.epfl.ch/GraphList/ProjectList>, and the readers can rank the pathways according to their own rules.

3.2.6 Experimental implementation and pathway optimization

The highest ranked candidate pathways can then be experimentally implemented in the host organism and can further be optimized through the Design-Built-Test-(Learn) cycle of metabolic engineering [169-171].

3.3 Results and Discussion

3.3.1 Generated metabolic network around Methyl Ethyl Ketone

We used the retrobiosynthesis algorithm of BNICE.ch to reconstruct the biochemical network around MEK. BNICE.ch [12, 129, 139-144] is a computational framework that takes advantage of the biochemical knowledge derived from the thousands of known enzymatic reactions to predict all possible biotransformation pathways from known compounds to desired target molecules. We applied BNICE.ch and generated all compounds and reactions that were up to five reaction steps away from MEK (Figures 3.2a and 3.2b).

To start the reconstruction procedure, we provided the initial set of compounds that contained 26 cofactors along with MEK (Appendix Chapter 3 Table S2). In the first BNICE.ch generation, we produced 6 biochemical and 25 chemical compounds connected through 48 reactions to MEK. Interestingly, among these reactions were also the ones proposed by Yoneda *et al.* [147], Srirangan *et al.* [148] and Ghiaci *et al.* [9]

After five generations, a total of 13'498 compounds were generated (Figure 3.2a). Out of these, 749 were biochemical and the remaining 12'749 were chemical compounds. We could also find 665 out of the 749 biochemical compounds in the PubChem database. All generated compounds were involved in 65'644 reactions, out of which 560 existed in the KEGG database and the remaining 65'084 were novel reactions (Figure 3.2b). A large majority of the predicted reactions (67%) were oxidoreductases, 15.4% were lyases, 8.6% were hydrolases, 4.3% transferases, 3.6% isomerases and only 0.72% ligases (Figure 3.2c). Out of 361 bidirectional generalized enzyme reaction rules of BNICE.ch, 220 were required to generate the metabolic network around MEK with the size of 5 reaction

steps. As expected from the statistics on the predicted reactions, most of these rules (38%) described the oxidoreductase biotransformation (Figure 3.2d).

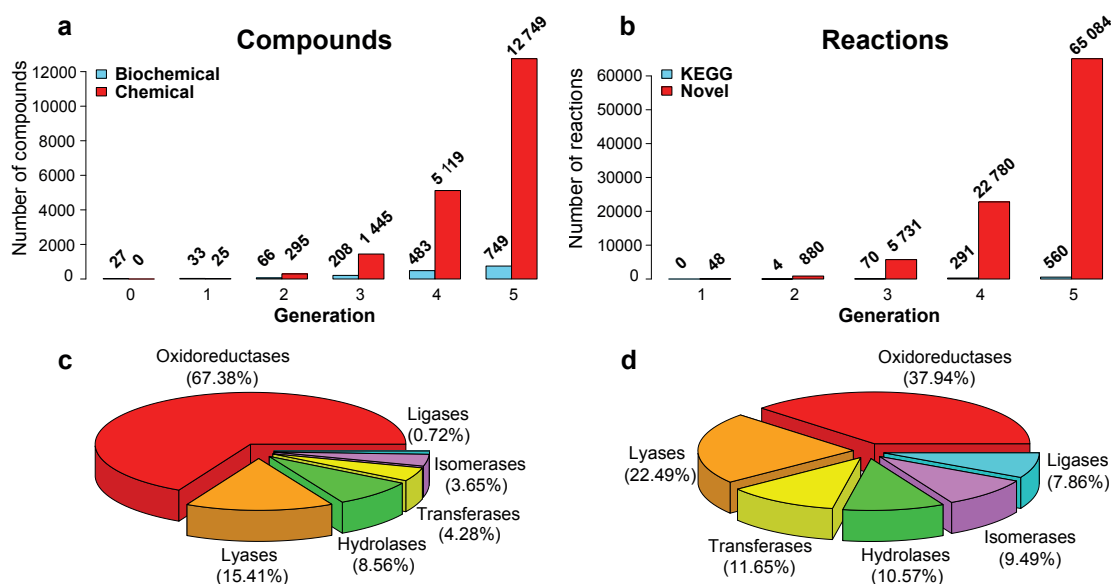


Figure 3.2 Growth of the generated metabolic network over 5 generations

(a) The BNICE.ch retrobiosynthesis algorithm generated 749 biochemical (blue) and 12'749 chemical (red) compounds. (b) Generated compounds participated in 560 KEGG (blue) and 65'084 novel (red) reactions. Categorization of the predicted reactions (c) and utilized generalized enzyme reaction rules (d) on the basis of their Enzymatic Commission[172], EC, classification.

Though MEK participated in a total of 1'551 reactions (Appendix Chapter 3 Table S5) only one reaction, which connected MEK to MEKCNH, was catalogued in the KEGG database (KEGG R09358). These 1'551 reactions connected MEK to 1'325 compounds (159 biochemical and 1'166 chemical), which could be potentially used as MEK precursors (Appendix Chapter 3 Table S1). Reaction steps for a biochemical production of MEK from the five precursors (3OXPNT, MVK, BuNH₂, 1B2OT, and MEKCNH) together with their most similar KEGG reactions can be consulted in Appendix Chapter 3 Table S6.

3.3.2 Pathway reconstruction towards five target compounds

In the pathway reconstruction process, we used as starting compounds 157 metabolites selected from the generated network, which were identified as native *E. coli* metabolites using the *E. coli* genome-scale model iJO1366 [29] (Appendix Chapter 3 Table S3). We performed an exhaustive pathway search on the generated metabolic network, and we reconstructed 3'679'610 pathways towards these five target compounds with pathway lengths ranging from 1 up to 4 reaction steps (Table 3.1). The reconstructed pathways combined consist of 37'448 reactions, i.e., 57% of the 65'644 reactions reproduced from the BNICE.ch generated metabolic network.

More than 58% of the discovered pathways were towards BuNH₂, while only 3.8% of the reconstructed pathways were towards 1B2OT, which was the only PubChem target compound (Table 3.1). Only 33 reconstructed pathways were of length one, and 28 out of them were towards BuNH₂ and none towards 1B2OT. The majority of reconstructed pathways (> 97%) were of length four. These results suggest that the biochemistry of enzymatic reactions favors smaller changes of a molecule structure over several steps.

Table 3.1 Reconstructed pathways towards five target compounds.

Target compounds	Reconstructed pathways	Reaction steps				Feasible pathways	
		1	2	3	4	FBA	TFA
3-oxopentanoate (3OXPNT)	641'493	1	198	12'222	629'072	361'187	11'145
but-3-en-2-one (MVK)	438'889	1	136	7'554	431'198	57'173	4'117
Butylamine (BuNH ₂)	2'146'890	28	1'236	53'573	2'092'053	27'211	916
but-1-en-2-olate (1B2OT)	140'779	0	53	2'905	137'821	30'689	1'826
2-hydroxy-2-methyl- butanenitrile (MEKCNH)	311'559	3	94	6'546	304'916	11'151	658
	3'679'610	33	1'717	82'800	3'595'060	487'411	18'662

3.3.3 Evaluation of reconstructed pathways

We performed a series of studies of the 3'679'610 generated pathways to assess their biological feasibility and performance (Methods 3.2.3). The feasibility of the pathways depends on the metabolic network of the chassis organism. Therefore, we embedded each of the reconstructed pathways in the *E. coli* genome-scale model iJO1366 and performed flux balance analysis (FBA) [162] and thermodynamics-based flux analysis (TFA) [51-53, 55]. The directionality of the reactions is an important factor in FBA and TFA [55], and in our studies, unless stated otherwise, for FBA and TFA we applied the *C₁* constraints on reaction directionalities where we constrained the reactions that involve CO₂ to operate in the decarboxylation direction (Methods 3.2.3).

Flux balance analysis. We used FBA as a prescreening method to reject pathways that were incompatible with the host organism (Methods 3.2.3). If an FBA model formed by embedding a pathway in iJO1366 can produce the target compound, then the pathway is considered as FBA feasible. Out of all reconstructed pathways, only 13.24% (487'411) were FBA feasible (Table 3.1). Though the largest number of reconstructed pathways were towards BuNH₂, only 1.27% (27'211) of these were FBA feasible. The number of FBA feasible pathways for MEKCNH was also low (3.59%). In contrast, more than 56% of pathways towards 3OXPNT were FBA feasible.

Thermodynamics-based flux analysis. We used TFA to identify 18'662 thermodynamically feasible pathways (0.5% of all generated pathways, or 3.8% of the FBA feasible pathways). A pathway is considered TFA feasible if a TFA model formed by embedding the pathway in iJO1366 can produce the target compound under thermodynamic constraints. The set of TFA feasible pathways involved 3'166 unique reactions. These results demonstrate that TFA is important for pathway evaluation and screening.

Table 3.2 Number of known reaction steps versus all reaction steps in the predicted pathways.

Table should be read as follows: e.g., among the predicted pathways of length 3 towards 3-oxopentanoate, there were 371 pathways with all novel reaction steps (no known steps), 118 pathways with 1 known and 2 novel reaction steps, and no pathways with 2 known and 1 novel reaction steps. Pathways with one novel reaction step are marked in red. All shown pathways are TFA feasible.

		Total reaction steps			Feasibility	
		2	3	4	TFA	
Number of known reaction steps in a pathway	0	14	371	7'059	7'444	3-oxopentanoate (3-oxopentanoate)
	1		118	2'956	3'074	
	2			627	627	
	0	4	72	3'196	3'272	but-3-en-2-one (MVK)
	1	1	13	703	717	
	2		2	110	112	
	3			16	16	
	0	2	27	752	781	Butylamine (BuNH ₂)
	1		7	108	115	
	2			20	20	
	0		23	1'576	1'599	but-1-en-2-olate (1B ₂ OT)
	1		10	196	206	
	2			21	21	
	0		50	380	430	2-hydroxy-2-methyl-butanenitrile (MEKCNH)
	1		2	202	204	
	2			24	24	
		21	695	17'946	18'662	

We found BuNH₂ to have the lowest rate of TFA feasible pathways with 0.04% of reconstructed pathways being TFA feasible (Table 3.1). The highest rate of TFA feasible pathways was again for 3OXPNT (1.74 %). The shortest TFA feasible pathways consisted of 2 reaction steps (21 pathways), whereas a majority of TFA feasible pathways had length 4 (Table 3.2). All pathways contained novel reaction steps, and only 19 pathways had one novel reaction step (Table 3.2). All of these 19 pathways were towards MVK, and they all had as intermediates 2-acetolactate and acetoin. The final reaction step converting acetoin to MVK was novel in all of them.

Yield analysis. We used TFA to assess the production yield of the feasible pathways from glucose to the target compounds (Appendix Chapter 3 Table S7). We identified pathways for all target compounds that could operate without a loss of carbon from glucose. More than a half of the pathways towards 3OXPNT (57%) could operate with the maximum theoretical yield of 0.774 g/g, i.e., 1Cmol/1Cmol (Appendix Chapter 3 Table S7). In contrast, only 4% (25 out of 658) pathways towards MEKCNH could operate with the maximal theoretical yield of 0.66 g/g (Appendix Chapter 3 Table S7). We found that pathway yields were distributed into several distinct sets rather than being more spread and continuous, i.e., we obtained eleven sets for 3OXPNT, four sets for MEKCNH, 11 sets for BuNH₂, nine sets for 1B2OT and ten sets for MVK (Appendix Chapter 3 Table S7). Interestingly, a discrete pattern in pathway yields was also observed in a similar retrobiosynthesis study for the production of mono-ethylene glycol in *Moorella thermoacetica* and *Clostridium ljungdahlii* [159].

Analysis of alternative assumptions on reaction directionalities. Since we found that the directionality of reactions in the network impacts yields, we investigated how the type of alternative constraints C₂ affected the yield distribution (Methods 3.2.3). The C₂ constraints contain the preassigned reaction directionalities from the iJO1366 model together with the C₁ constraints. As expected, these additional constraints reduced flexibility of the metabolic network and some pathways even became infeasible (Appendix Chapter 3 Table S8). With the C₂ constraints, the yields were in general reduced and their distribution was more spread compared to the one obtained using the C₁ constraints. For example, we found with both sets of constraints three alternative

pathways for the production of 3OXPNT from acetate via two intermediate compounds: 2-ethylmalate and (3S)-3-hydroxypentanoate. The three alternative pathways had three different cofactor pairs in the final reaction step that converts (3S)-3-hydroxypentanoate to 3OXPNT (Figure 3.3). With the C_1 constraints, the three pathways had an identical yield of 0.642 g/g. In contrast, with the C_2 constraints, the pathway with NADH/NAD cofactor pair in the final step had a yield of 0.537 g/g, the one with NADPH/NADP had a yield of 0.542 g/g, and the one with H_2O_2/H_2O had a yield of 0.495 g/g. These differences in yields are a consequence of the different costs of cofactor production upon adding supplementary constraints.

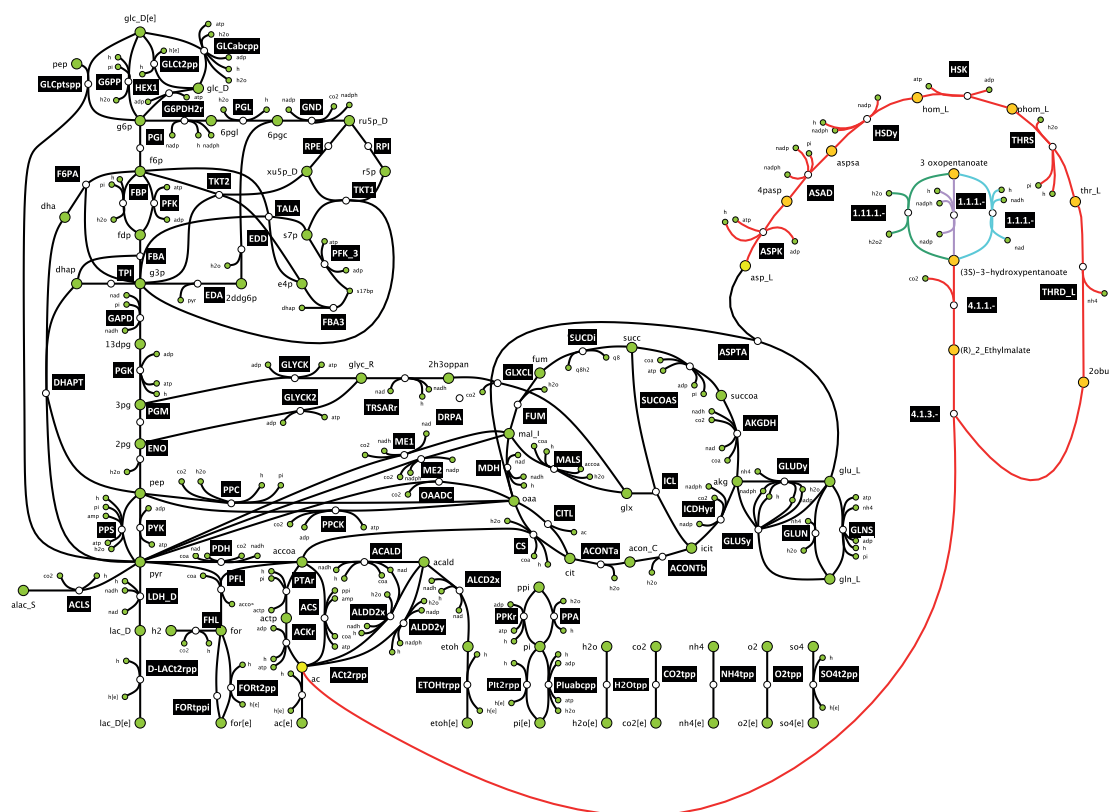


Figure 3.3 Three alternative ways to produce 3OXPNT from acetate through 2 intermediate metabolites: 2-ethylmalate and 3-hydroxypentanoate

BridgIT analysis. For each novel reaction from the feasible pathways, we identified the most similar KEGG reaction whose gene and protein sequences were assigned to the novel reaction (Methods). The BridgIT [163] results can be consulted at <http://lcsb-databases.epfl.ch/GraphList/ProjectList> upon subscription.

3.3.4 Identification and analysis of anabolic subnetworks capable of synthesizing target molecules

In pathway reconstruction, we identified the sequence of the main reactions required to produce the target molecules from precursor metabolites in the core network. However, these reactions require additional co-substrates and cofactors that should become available from the rest of the metabolism. In addition, these reactions produce also side products and cofactors that must be recycled by the genome-scale metabolic network in order to have a biologically feasible and balanced subnetwork for the production of the target molecules. Therefore, we identified the active metabolic subnetworks required to synthesize the corresponding target molecule (Methods 3.2.4). The active metabolic subnetworks were then divided into the *core metabolic network*, which included central carbon metabolism pathways [173, 174], and the active *anabolic subnetwork* (Figure 3.4.a, and Methods 3.2.4). On average there were more than three alternative anabolic subnetworks per pathway due to the redundant topology of metabolism (Table 3.3). For example, we identified 35'013 alternative anabolic subnetworks for 11'145 feasible pathways towards 3OXPNT. Overall, for the 18'662 TFA feasible pathways, 55'788 active anabolic subnetworks were identified.

Table 3.3 Alternative anabolic subnetworks for 5 target compounds together with their lumped reactions and precursors.

Target compounds	Feasible pathways	Alternative anabolic subnetworks	Unique lumped reactions	Overlapping sets of precursors	Unique precursors
3-oxopentanoate (3OXPNT)	11'145	35'013	4'517	281	40
but-3-en-2-one (MVK)	4'117	10'162	1'762	126	32
Butylamine (BuNH ₂)	916	2'858	974	102	30
but-1-en-2-olate (1B2OT)	1'826	5'339	1'536	97	30
2-hydroxy-2-methyl- butanenitrile (MEKCNH)	658	2'416	792	37	17
<hr/>					
	18'662	55'788	9'581		
<hr/>					

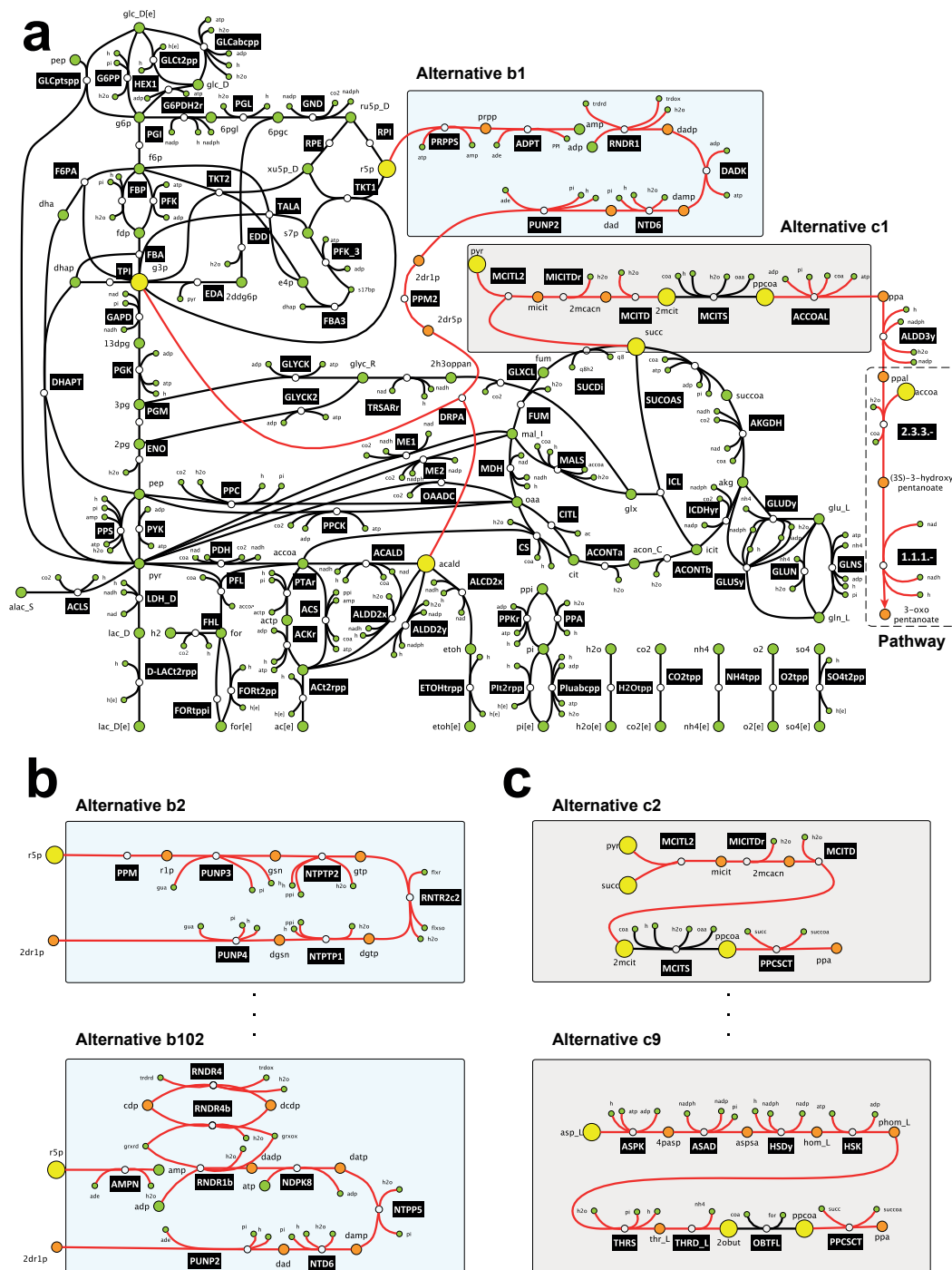


Figure 3.4 Alternative ways of producing 3-oxo-pentanoate from glucose.

(a) Schematic representation of the metabolic network producing 3-oxo-pentanoate from glucose. Reactions pertaining to the core metabolic (black) and the active anabolic (red) subnetworks. Metabolites of the core metabolic (green) and the active anabolic (orange) subnetworks together with connecting precursors (yellow), i.e., metabolites that connect the core and active anabolic subnetworks. (b) Alternative pathways connecting ribose-5-phosphate, r5p, with 2-deoxy-D-ribose-1-phosphate, 2dr1p. (c) Alternative pathways connecting the core metabolites with propanal, Ppal.

Next, we computed a lumped reaction for each of the alternative subnetworks (Methods 3.2.4). Similar to previous findings from the analysis of the biomass building blocks in *E. coli* [110], only 9'581 out of the 55'788 computed lumped reactions were unique (Table 3.3). This result suggests that the overall chemistry and the cost to produce the corresponding target molecule are the same for many different pathways. Since the cost of producing a target molecule depends of the host organism, this implies that the choice of the host organism is important. On the other hand, the multiple alternative options could also provide useful degrees of freedom for synthetic biology and metabolic engineering design.

The largest diversity in alternative subnetworks per lumped reaction was found for 3OXPN_T, where on average more than seven alternative subnetworks had the same lumped reaction (Table 3.3). In contrast, we observed the smallest diversity for BuNH₂ with approximately three alternative subnetworks per lumped reaction (Table 3.3). An illustrative example of multiple pathways with the same lumped reaction is provided in Figure 3.5.

Interestingly, the 35'013 active anabolic networks towards the production of 3OXPN_T were composed of only 394 unique reactions. Out of these 394 reactions, 132 were common with the pathways leading towards the production of all biomass building blocks (Appendix Chapter 3 Table S9) except chorismate, phenylalanine, and tyrosine. This finding suggests that biomass building blocks could be competing for resources with 3OXPN_T and that they could affect the production of this compound.

Origins of diversity of alternative anabolic subnetworks. To better understand the diversity in alternative anabolic subnetworks, we performed an in-depth analysis of the two-step pathway from acetyl-CoA and propanal to 3OXPN_T, which presented the largest number of alternative anabolic networks (185) among all reconstructed pathways (Figure 3.4a). The smallest anabolic subnetwork of the 185 alternatives consisted of 14 enzymes, whereas the largest one comprised 22 enzymes (Appendix Chapter 3 Table S10). All 185 subnetworks shared five common enzymes: the two enzymes from the reconstructed pathway converting propanal via (3S)-3-hydroxypentanoate to 3OXPN_T (with the BNICE.ch assigned third level Enzymatic Commission [172], EC, numbers

2.3.3.- and 1.1.1.-), two enzymes involved in acetyl-CoA production (phosphopentomutase deoxyribose (PPM2), and deoxyribose-phosphate aldolase (DRPA)), and aldehyde dehydrogenase (ALDD3y) that converts propionate to propanal (Figure 3.4).

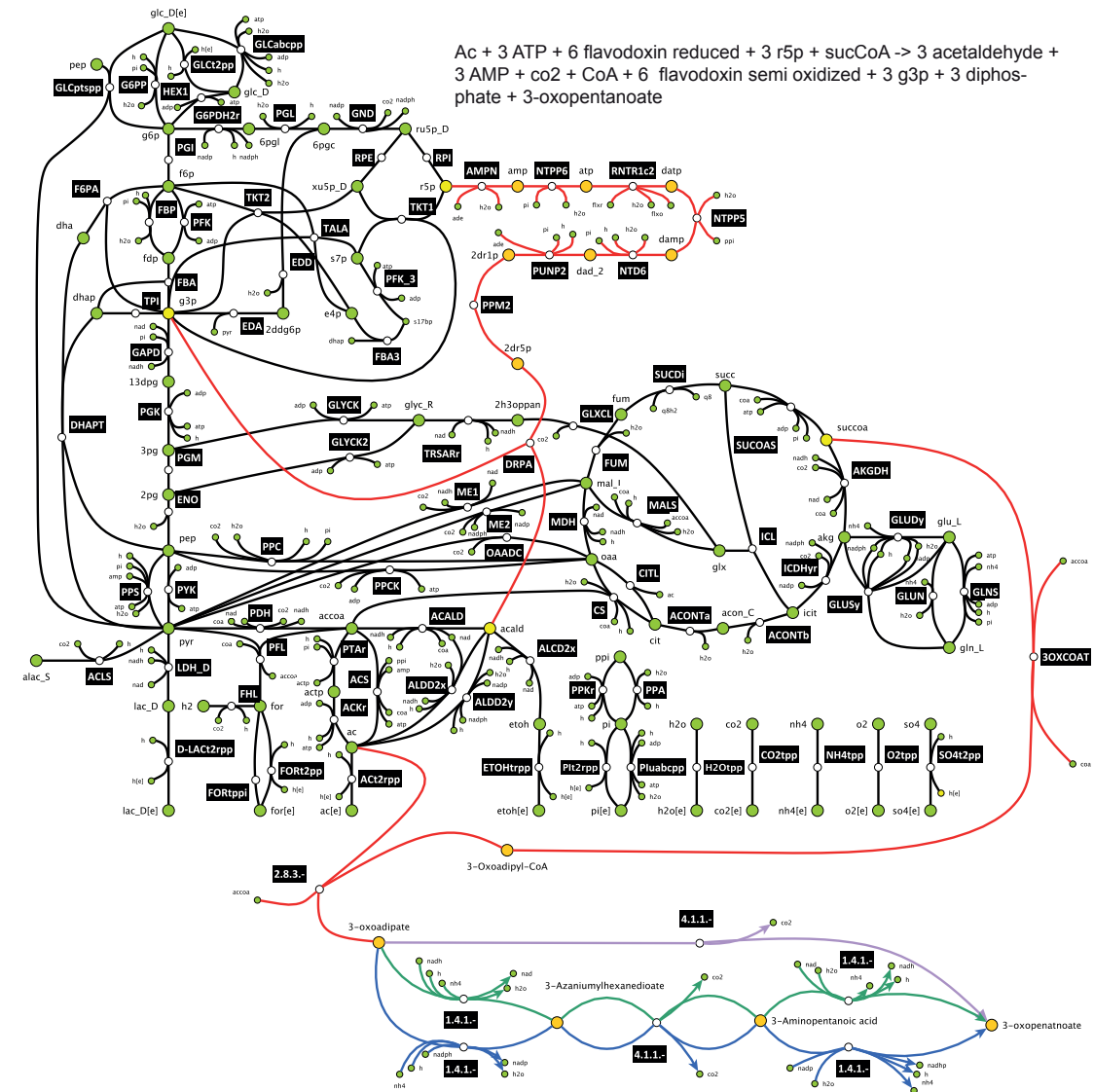


Figure 3.5 Three different pathways from acetate to 3-oxopentanoate sharing the same lumped reaction. Ac (acetate), r5p (D-ribose 5-phosphate), sucCoA (succinyl CoA), g3p (glyceraldehyde 3-phosphate).

The multiplicity of ways to produce acetyl-CoA and propionate contributed to a large number of alternative subnetworks: there were 102 alternative ways of producing acetyl-CoA from ribose-5-phosphate (r5p) via 2-deoxy-D-ribose-1-phosphate (2drip) (Figure 3.4.b) and 9 different ways of producing propionate (Figure 3.4.c).

There were two major routes to produce 2drip within the 102 alternatives. In the first route with 50 alternatives, r5p is converted either to ribose-1-phosphate (in 31 alternatives) or to D-ribose (in 19 alternatives), which are intermediates in producing nucleosides such as adenosine, guanosine, inosine and uridine. These nucleosides are further converted to deoxyadenosine (dad), deoxyguanosine (dgsn) and deoxyuridine (duri) that are ultimately phosphorylated to 2drip. In 26 of the remaining 52 alternatives of the second route, r5p is converted to phosphoribosyl pyrophosphate (prpp), which is followed by a transfer of its phospho-ribose group to nucleotides such as AMP, GMP, IMP and UMP. These nucleotides are then converted to 2drip by downstream reaction steps. In the remaining alternatives for the second route, r5p is first converted to AMP in one reaction step, and then to 2drip via dad and dgsn.

There were 9 alternative routes to produce propionate. In 4 of these, this compound was produced from pyruvate and succinate (Figure 3.4.a and 3.4.c), in 3 routes it was produced from aspartate (Figure 3.4.c), and in 2 routes it was produced from 3-phosphoglycerate and glutamate.

Connecting precursors of five target compounds. An abundant availability of precursor metabolites is crucial for an efficient production of target molecules [175]. Here, we defined as *connecting precursors* the metabolites that connect the core to the active anabolic subnetworks (Figure 3.4.a). We analyzed the different combinations of connecting precursors that appeared in the alternative subnetworks. Our analysis revealed that the majority of subnetworks were connected to the core network through a limited number of connecting precursors. We found that all 35'013 alternative subnetworks for the production of 3OXPNP were connected to the core network by 281 sets of different combinations among 40 unique connecting precursors (Table 3.3). We ranked these sets based on their number of appearances in the alternative networks. The top ten sets appeared in 24'210 subnetworks, which represented 69% of all identified

subnetworks for this compound (Table 3.4). Moreover, the metabolites from the top set (acetyl-CoA, propionyl-CoA, pyruvate, ribose-5-phosphate, and succinate) were the precursors in 8'510 (24.3%) subnetworks for 3OXPNP (Table 3.4). Ribose-5-phosphate appeared in 9 out of the top ten sets, and it was a precursor in 32'237 (92%) 3OXPNP producing subnetworks.

Table 3.4 Top ten connecting precursor combinations for the production of 3OXPNP. Connecting precursors: acetate (ac), acetyl-CoA (acCoA), aspartate (asp-L), dihydroxyacetone phosphate (dhap), propionyl-CoA (ppCoA), pyruvate (pyr), ribose-5-phosphate (r5p), succinate (succ), succinyl-CoA (succCoA).

ac	acCoA	asp-L	dhap	ppCoA	pyr	r5p	succ	succCoA	No. of sub-networks	No. of feasible pathways
	✓			✓	✓	✓	✓		8'510	624
		✓		✓		✓			5'409	2790
				✓	✓	✓	✓		3'463	920
✓						✓		✓	1'344	672
					✓	✓	✓		1'049	382
	✓					✓	✓		965	191
	✓					✓		✓	956	478
				✓	✓		✓		915	460
	✓		✓			✓			834	419
			✓			✓			765	387
									24'210	7323

3.3.5 Clustering of feasible pathways

The repeating occurrences of connecting precursors and lumped reactions in the alternative anabolic subnetworks motivated us to identify common patterns in connecting precursors, enzymes, and intermediate metabolites required to produce the target molecules. To this end, we used the feasible pathways from acetate to 3OXPNT as the test study, and we performed two types of clustering on these 115 pathways.

Clustering based on connecting precursors and byproducts of lumped reactions.

We identified 242 alternative anabolic networks for 115 pathways from the test study, and we computed the corresponding 242 lumped reactions. We chose the first lumped reaction returned by the solver for each of the 115 pathways, and we clustered the pathways based on the structural similarity between the connecting precursors and byproducts of the lumped reactions (Methods 3.2.4).

The clustering separated the 115 pathways in eleven groups, B₁-B₁₁ (Figure 3.6a and 3.6b, Appendix Chapter 3 Table S₁₁). The main clustering condition among the 115 pathways was the presence or absence of thioesters, such as AcCoA, in the set of connecting precursors. There were 56 pathways with CoA-related precursors (groups B₁-B₅) and 59 pathways that did not require CoA (groups B₆-B₁₁). The pathways from groups B₁-B₅ were further clustered subject to the presence of: the precursor succCoA and the byproduct CO₂ (group B₁); the precursor succCoA and the byproduct malonate (group B₂); the precursor ppCoA (group B₃); the precursors acCoA and ppCoA (group B₄); and the precursor acCoA (group B₅). The pathways that did not require CoA were further clustered depending on if they had as precursors dhap (groups B₆ and B₇) or formate (B₈) or not (B₉-B₁₁).

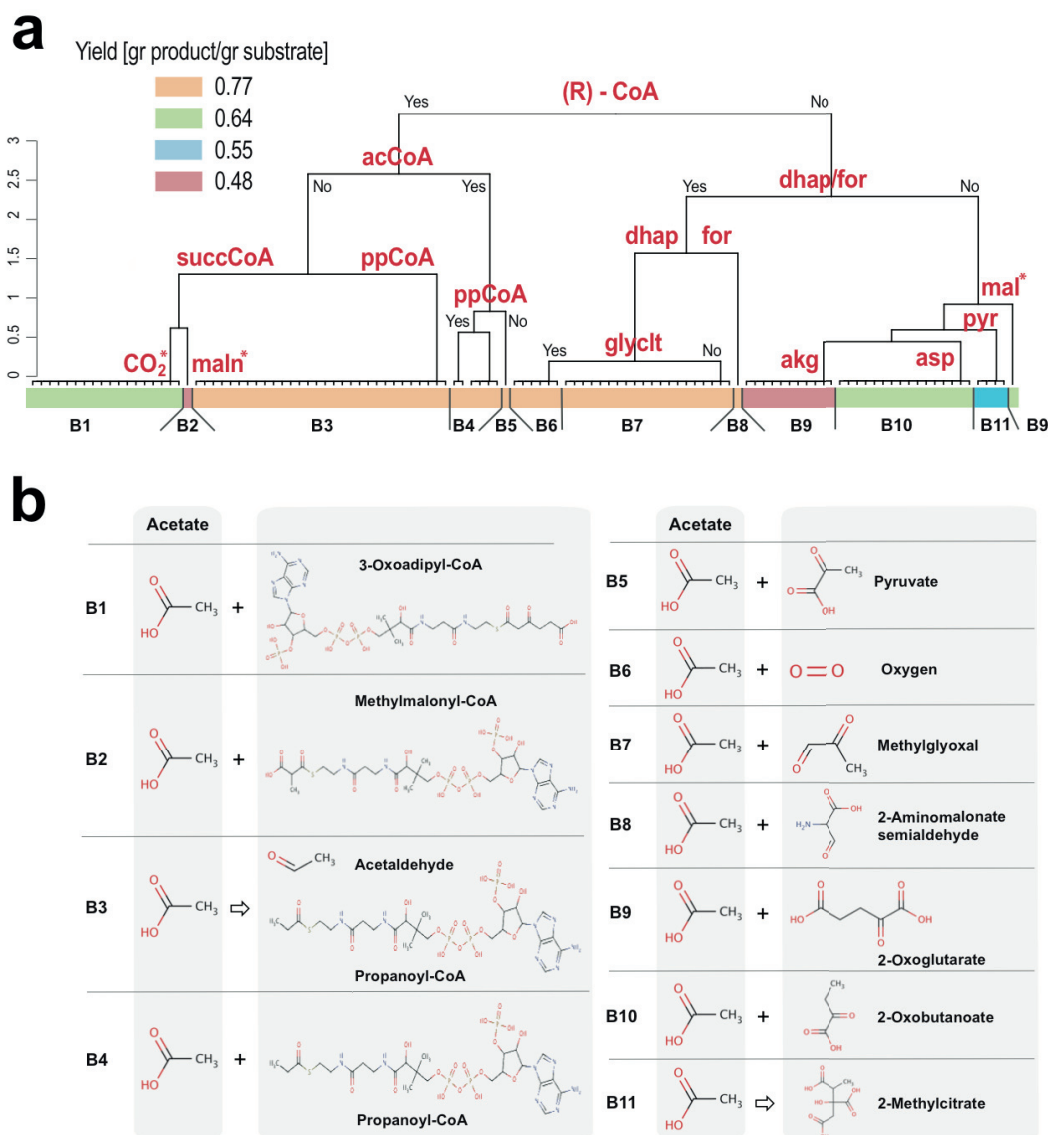


Figure 3.6 Clustering of the 115 reconstructed pathways from acetate to 3OXPNP.

(a) Pathways were classified in eleven groups (B1-B11) based on connecting precursors and byproducts of their lumped reactions. The byproducts are denoted with an asterisk (*). The pathway yields were consistent within each of the groups and distinctly separated between them. The color-coding of the groups corresponds to the yields of the involved pathways. (R)-CoA denotes the group of thioesters. Abbreviations: 2-oxoglutarate (akg), acetyl-CoA (acCoA), aspartate (asp), dihydroxyacetone phosphate (dhap), formate (for), glycolate (glyclt), malate (mal), malonate (maln), propionyl-CoA (ppCoA), pyruvate (pyr), succinyl-CoA (succCoA). (b) Nine out of eleven groups of reconstructed pathways (B1-B2 and B4-B10) are characterized by the connecting precursors and their co-substrates in the first reaction step of the pathways. Group B3 is characterized by intermediate metabolites acetaldehyde and propionyl-CoA involved in the novel reaction step EC 2.3.1.-. Group B11 pathways have as an intermediate 2-methylcitrate.

Pathways from group B₃ utilized different co-substrates, such as ATP and crotonoyl-CoA, along with acetate to produce acetaldehyde in the first reaction step. All these pathways shared a common novel reaction step with acetaldehyde and propionyl-CoA as substrates (EC 2.3.1.-).

Finally, group B₁₁ contained the pathways with the intermediate 2-methylcitrate, which was produced from pyruvate (pyr).

The presented clustering analysis has been shown to be very powerful in identifying the features of the large number of pathways. The classification can further guide us to identify the biochemistry responsible for the properties of pathways. Such deeper understanding can provide further assistance for the design and analysis of novel synthetic pathways.

Clustering based on involved enzymes. Although the clustering based on the connecting precursors and byproducts provided an insight of the chemistry underlying the production of 3OXPN_T from acetate, lumped reactions conceal the identity of the enzymes involved in the active anabolic subnetworks. We analyzed the 115 active subnetworks corresponding to 115 pathways, and we found that five enzymes were present in all of them: AMP nucleosidase (AMPN), 5'-nucleotidase (NTD6), purine-nucleoside phosphorylase (PUNP₂), PPM₂ and DRPA, which participated in the production of acetaldehyde from r5p (Figure 3.7b).

To find common enzyme routes in these subnetworks, we performed a clustering based on the structural similarity between their constitutive reactions (Methods). The clustering separated 115 subnetworks in two groups depending on the existence (47 subnetworks) or not (68 subnetworks) of a sequence of six enzymes starting with aspartate kinase (ASPK) and ending with L-threonine deaminase (THRD_L), whose product 2-oxobutanoate was converted downstream to 3OXPN_T (Figures 3.7a and 3.7b).

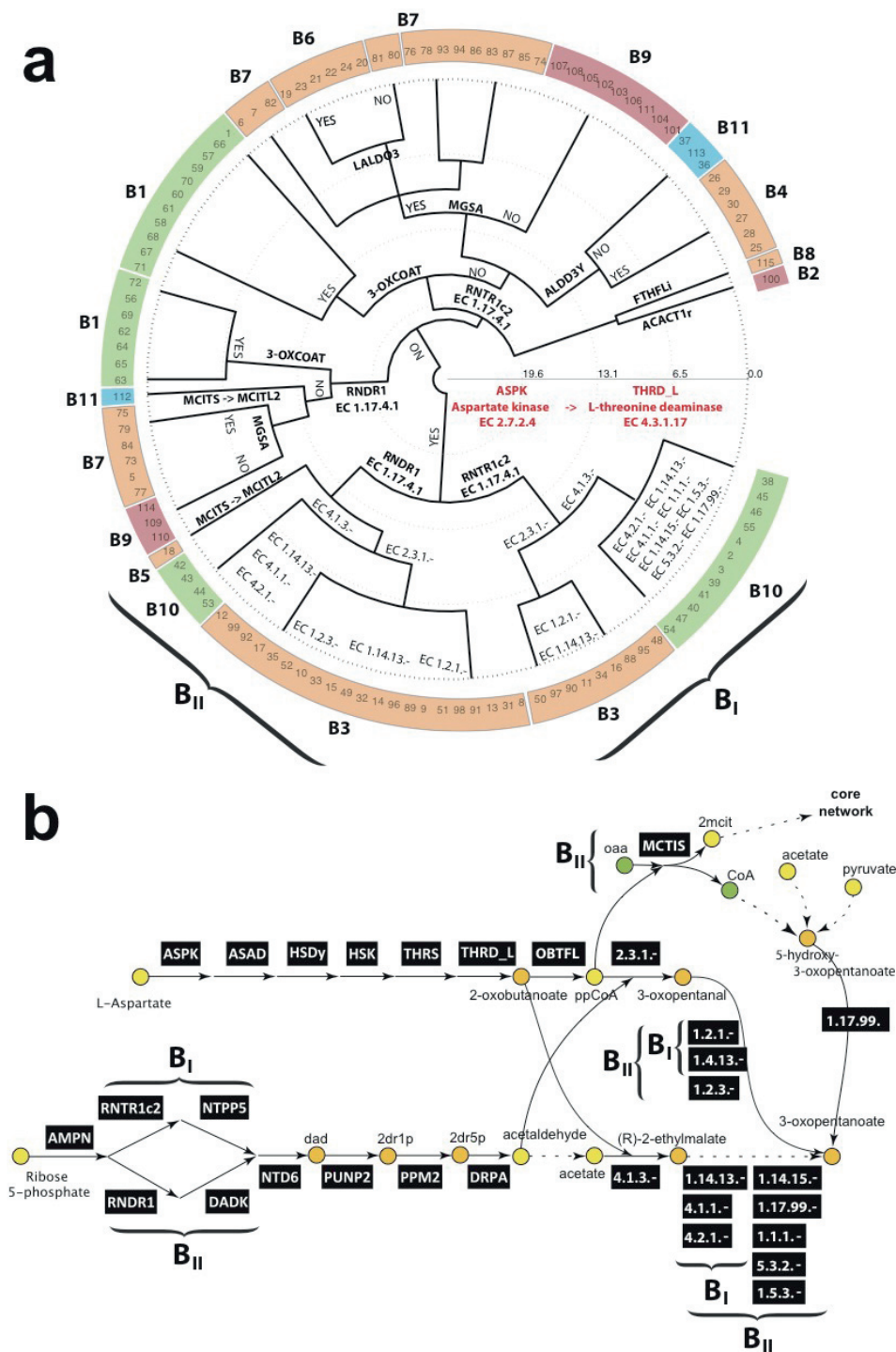


Figure 3.7 Clustering of 115 active subnetworks corresponding to 115 reconstructed pathways from acetate to 3OXPNP.

(a) Subnetworks were clustered based on enzymes they involved. The groups B1-B11 are color-coded as in Figure 3.6.a. (b) Structure of 47 subnetworks containing a sequence of six enzymes starting with aspartate kinase (ASPK) and ending with L-threonine deaminase (THRD_L) (groups B_I and B_{II} in a). The core metabolites are marked in green, the connecting metabolites in yellow, while the metabolites from the active anabolic networks are marked in orange.

Both groups were further clustered based on a set of enzymes required to produce deoxyadenosine and the downstream metabolite acetaldehyde (Figures 4a and 4b). The first subgroup of enzymes, i.e. ribonucleoside-diphosphate reductase (RNDR₁), deoxyadenylate kinase (DADK) and NTD6, converted adp to deoxyadenosine. In the second subgroup, atp was transferred to deoxyadenosine via ribonucleoside-triphosphate reductase (RNTR_{1c2}), nucleoside triphosphate pyrophosphorylase (NTPP₅) and NTD6 (Figure 3.7b). Then, for both subgroups, deoxyadenosine was converted to 2-deoxy-D-ribose 5-phosphate (2dr5p) that was further transformed to acetaldehyde via PPM2 and DRPA (Figures 3.4 and 3.7b).

The clustering based on enzymes allowed us to identify enzymatic routes corresponding to different yields (Figure 3.7a, and Appendix Chapter 3 Table S7). For example, all pathways that include ASPK and novel reaction steps with the third level EC class 2.3.1.- and 1.2.1.- (group B₃), would provide the maximal theoretical yield of 0.774 g/g (Figure 4a). Similarly, pathways that contain ALDD_{3Y} (group B₄), methylglyoxal synthase (MGSA) (groups B₆ and B₇), and ASPK, RNDR₁, and methylisocitrate lyase (MCITL₂) (group B₅), would also provide the maximal theoretical yield. In contrast, the clustering also permitted us to identify key enzymes participating in pathways with a reduced yield. For example, pathways that contained 3-OXCOAT had a yield of 0.644 g/g.

Furthermore, the clustering based on enzymes allowed us to clarify the link between the precursors and the corresponding sequence of enzymes that needed to be active for producing the target molecule. For example, pathways from group B₁, which had succCoA as a connecting precursor and CO₂ as a byproduct, had the common reaction step 3-OXCOAT (Figure 3.7a). Similarly, all pathways from group B₄ with connecting precursors ppCoA and acCoA contained ALDD_{3Y}.

3.3.6 Ranking of biosynthetic pathways and recommendations

We further ranked the corresponding feasible pathways according to number of reaction steps and enzymes that could be directly implemented or needed to be engineered, their yield, and the BridgIT score of the novel reaction steps (Methods 3.2.5). As we saw earlier (e.g. in Appendix Chapter 3 Table S7), there are several distinct maximum yield values that can be achieved with all these alternatives rather than a continuous distribution of

yields. The clustering analysis suggests that the reason for the discreet distribution is the loss of the carbon atoms in specific steps along the pathways. We obtained the top candidate pathways for each of the target molecules that were likely to produce these compounds with economically viable yields (Appendix Chapter 3 Table S12-S16). For each of the target molecules, the highest ranked candidate pathways could operate with their maximum theoretical yields (Figure 3.7). Furthermore, the BridgIT results suggest that the novel reactions appearing in these pathways can be catalyzed by the known enzymes (Figure 3.7). The highest ranked candidate pathway among all feasible pathways was from pyruvate to 3OXPNP, and it consisted of two novel reactions of the third level EC class 2.3.3- and 1.1.1.- (Figure 3.7a). The BridgIT[163] analysis identified KEGG R00472 as the most similar reaction to 2.3.3.-. KEGG reports that R00472 can be catalyzed by EC 2.3.3.9. Similarly, KEGG R01361 was identified as the most similar to 1.1.1., and according to the KEGG database this reaction is catalyzed by EC 1.1.1.30. Interestingly, there is a reaction that involves CO₂ in the top pathways for MVK and MEKCNH (Figure 3.7b and 3.7e). Although in the literature this reaction is reported to operate in the decarboxylation direction, TFA allows it to operate in the opposite direction as well. Without TFA this information would stay hidden.

The pathways were visualized and can be consulted at <http://lcsb-databases.epfl.ch/GraphList/ProjectList> upon subscription.

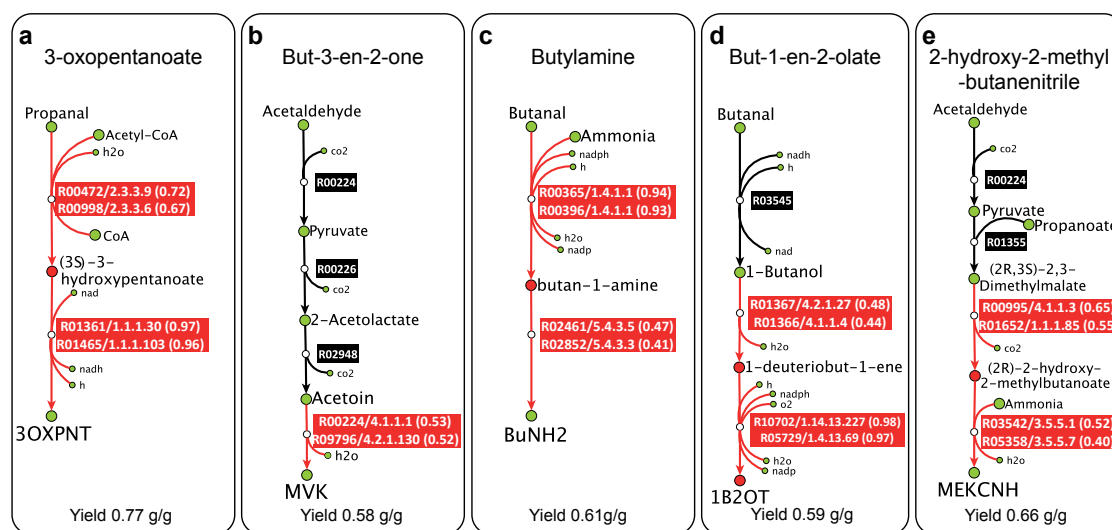


Figure 3.8 The highest ranked candidate pathways for the production of (a) 3OXPN, (b) MVK, (c) BuNH₂, (d) 1B2OT and (e) MEKCNH.

Known non-orphan KEGG reactions are the black and novel reactions the red boxes. KEGG compounds are denoted as the green and PubChem compounds as the red circles. For each of the novel reactions, the KEGG IDs, catalyzing enzyme, and the BridgIT similarity scores of the two most similar non-orphan KEGG reactions are provided.

3.3.7 Further experimental implementation and pathway optimization

After ranking of the top candidate pathways, the experts can choose the most amenable ones for experimental implementation in the host organism. The implemented pathways typically need to be optimized further for economically viable production titers and rates. The optimization is performed through the Design-Built-Test-(Learn) cycle of metabolic engineering [169-171] where stoichiometric [177-179] and kinetic models [68, 69, 180-185], genome editing [186, 187] and phenotypic characterization [188] are combined to improve recombinant strains for production of biochemicals.

3.4 Conclusions

In this chapter, we used BNICE.ch to reconstruct, evaluate and analyze more than 3.6 million biosynthetic pathways from the central carbon metabolites of *E. coli* towards five precursors of Methyl Ethyl Ketone (MEK), a 2nd generation biofuel candidate. Our evaluation and analysis showed that more than 18'000 of these pathways are biologically feasible. We ranked these pathways based on process- and physiology-based criteria, and we identified gene and protein sequences of the structurally most similar KEGG reactions to the novel reactions in the feasible pathways, which can be used to accelerate their experimental realization. Implementation of the discovered pathways in *E. coli* will allow the sustainable and efficient production of five precursors of MEK (3OXPNT, MVK, 1B2OT, BuNH₂, and MEKCNH), which can also be used as precursors for the production of other valuable chemicals [155-157].

The pathway analysis methods developed and used in this work offer a systematic way for classifying and evaluating alternative ways for the production of target molecules. They also provide a better understanding of the underlying chemistry and can be used to guide the design of novel biosynthetic pathways for a wide range of biochemicals and for their implementation into host organisms.

The present study shows the potential of computational retrobiosynthesis tools for discovery and design of novel synthetic pathways for complex molecules, and their relevance for future developments in the area of metabolic engineering and synthetic biology.

Abbreviations

Abbreviation	Reaction	EC number
3OXCOAT	3-oxoadipyl-CoA thiolase	2.3.1.174
ACACT1r	Acetyl-CoA C-acetyltransferase	2.3.1.9
ALDD3y	Aldehyde dehydrogenase (propanal, NADP)	1.2.1.4
AMPN	AMP nucleosidase	3.2.2.4
ASPK	Aspartate kinase	2.7.2.4
DADK	Deoxyadenylate kinase	2.7.4.3
DRPA	Deoxyribose-phosphate aldolase	4.1.2.4
FTHFLi	Formate-tetrahydrofolate ligase	6.3.4.3
LALDO3	L-Lactaldehyde:NADP+ 1-oxidoreductase	1.1.1.283, 1.2.1.49
MCITD	2-methylcitrate dehydratase	4.2.1.79
MCITL2	Methylisocitrate lyase	4.1.3.30
MGSA	Methylglyoxal synthase	4.2.3.3
MMM	Methylmalonyl-CoA mutase	5.4.99.2
NTD6	5'-nucleotidase (dAMP)	3.1.3.89
NTTP5	Nucleoside triphosphate pyrophosphorylase	3.6.1.19
PPM2	Phosphopentomutase 2 (deoxyribose)	5.4.2.7
PUNP2	Purine-nucleoside phosphorylase	2.4.2.1
RNDR1	Ribonucleoside-diphosphate reductase (ADP)	1.17.4.1
RNTR1c2	Ribonucleoside-triphosphate reductase (ATP)	1.17.4.2
THRD_L	L-threonine deaminase	4.1.1.19

Abbreviation	Compound	Alternative name	Database IDs	
			KEGG	PubChem
3OXPNT	3-Oxopentanoate	3-Oxopentanoic acid	C02233	5297
MVK	But-3-en-2-one	Methyl Vinyl Ketone	C20701	172232421
BuNH ₂	Butylamine	Butanamine	C18706	124489380
1B2OT	But-1-en-2-olate	1-Butene-2-olate	X	54444500
MEKCH	2-Hydroxy-2-methylbutanenitrile	Methyl Ethyl Ketone Cyanohydrin	C18796	124489470
2drp	2-deoxy-D-ribose-1-phosphate	2-Deoxy-alpha-D-ribose 1-phosphate	C00672	3941
2dr5p	2-deoxy-D-ribose phosphate	5- 2-Deoxy-alpha-D-ribose 5-phosphate	C00673	3942
dad	Deoxyadenosine	2'-Deoxyadenosine	C00559	3839
dgsn	Deoxyguanosine	2'-Deoxyguanosine	C00330	3624
duri	Deoxyuridine	2'-Deoxyuridine	C00526	3809
prpp	Phosphoribosyl pyrophosphate	5-Phospho-alpha-D-ribose 1-diphosphate	C00119	3419
ac	Acetate	Acetic acid	C00033	3335
acCoA	Acetyl-CoA	Acetyl coenzyme A	C00024	3326
akg	2-oxoglutarate	2-Ketoglutaric acid	C00026	3328
asp-L	Aspartate	L-Aspartic acid	C00049	3351
dhap	Dihydroxyacetone phosphate	3-Hydroxy-2-oxopropyl phosphate	C00111	3411
ppCoA	Propionyl-CoA	Propionyl coenzyme A	C00100	3400
pyr	Pyruvate	2-Oxopropanoate	C00022	3324
r5p	Ribose-5-phosphate	alpha-D-Ribose 5-phosphate	C03736	6499
succ	Succinate	Butanedionic acid	C00042	3344
succCoA	Succinyl-CoA	Succinyl coenzyme A	C00091	3391

Chapter 4. Comparative study on *Escherichia coli* and *Pseudomonas putida* metabolic capabilities

4.1 Introduction

Escherichia coli is probably the best studied and the most comprehensively understood prokaryotic organism [189], especially popular among metabolic engineers due to its simplicity, easy manipulation and cultivation. It was discovered and later named after Theodor Escherich in 1884. Since then, it was used as a model organism in many studies. More recently, it was used regularly as a host for the production of the second generation of biofuels [14, 16, 123, 124, 126, 190-192].

Successful application of genetically modified organisms for biofuel production depends on organism ability to produce biofuels in industrial scale [189]. In a report in Nature Biotechnology, Stephen del Cardayre, vice president of research and development at LS9 of South San Francisco, California stated: "To be competitive a fermentation needs to produce around 100 grams per liter of end product; its productivity should exceed two grams per liter per hour; and its anaerobic yield should stand at ~95% of the theoretic yield" [193]. Toxicity is recognized as the major obstacle towards obtaining high titers and yields [4, 14]. The natural tolerance of non-native systems is typically one to two order of magnitude lower than some of the native producers [194], e.g., isobutanol reduces *E. coli* growth already at concentrations > 1% vol/vol [192]. While *Pseudomonas putida* represents one of the most promising alternatives as a host for the biofuel production due to its superior tolerance to compounds such as: benzene, toluene, ethylbenzene, xylene and as well as other hydrocarbons (e.g., n-hexane, cyclohexane), there are not many examples in literature of *P. putida* usage as a biofuel producing host. Nikel and de Lorenzo [38] recruited genes for ethanol biosynthesis from *Zymomonas*

mobilis to activate EtOH synthesis pathway in *P. putida* and they showed that *P. putida* outperformed *E. coli* in every ethanol tolerance test conducted. Nielsen *et al.* [195] expressed butanol biosynthesis pathway from *Clostridium acetobutylicum* in *P. putida* and obtained titer of 50 ± 6 mg/l with glucose and 122 ± 12 mg/l with glycerol as the carbon source.

The main advantage of *P. putida* comes from its glucose catabolism, which favors NADPH production, an important counteractor against different environmental stresses [116]. Glycolysis in *P. putida* is known to occur almost solely through Entner-Doudoroff (ED) pathway [109, 196]. The first step in the ED pathway is the conversion of glucose to gluconate with enzyme glucose dehydrogenase localized in the periplasmic space. Gluconate is further converted into 2-ketogluconate or imported to the cytosol and converted to the key ED metabolite 6-phospho-D-gluconate. Nickel *et al.* [116] showed that about 90% of consumed glucose is converted to directly to 6-phospho-D-gluconate, confirming that other possible ways of glucose conversions, like conversion to glucose 6-phosphate are only minor shunts in *P. putida* [196]. 6-phospho-D-gluconate is broken down further to the ED pathway end-products pyruvate and glyceraldehyde 3-phosphate via EDD (6-phosphogluconate dehydratase) and EDA (2-dehydro-3-deoxy-phosphogluconate aldolase). As a result, the ED pathway yields one molecule of ATP, one of NADH and one of NADPH per one molecule of glucose consumed (Figure 4.1a). Chavarria *et al.* [109] tried to activate Embden-Meyerhoff-Parnas (EMP) pathway in *P. putida* by transferring *pfkA* (phosphofructokinase) gene from *E. coli*, but they failed to redirect carbon from the ED to EMP pathway.

In contrast to *P. putida*, *E. coli* has functional both EMP and ED pathway (Figure 4.1b). However, the glucose metabolism in this organism relies mainly on the EMP pathway, while the ED pathway remains mostly inactive [197]. Hollinshead *et al.* [197] showed that about 90 % of flux in *E. coli* is channeled through EMP pathway while the flux through ED pathway was negligible (less than 1% of the consumed total carbon). They overexpressed EDD and EDA enzymes and they reported that 20 % of the consumed total carbon was directed through the ED pathway. In a mutant strain lacking *pfkA* (phosphofructokinase) in combination with the overexpression of EDD and EDA, about 72 % of the flux is channeled through the ED pathway [197]. Since the EMP pathway

Similarly, Nielsen *et al.* [195] reported that the butanol biosynthesis in *E. coli* was followed with the secretion of ethanol, succinate, lactate, acetate, and formate, whereas *P. putida* secreted only small amounts of acetate during the butanol biosynthesis. When growing on glucose *P. putida* can accumulate relatively small amounts of gluconate and 2-ketogluconate, but these two metabolites can be re-imported back into the cell once glucose is depleted [111].

Despite many shared characteristics, it is clear that these two organisms have also many distinctive metabolic features. In this chapter, we compared these two organisms from different aspects. First, we focused on the similarities and differences of their core carbon networks. With redGEM and lumGEM we derived three different reduced models of *E. coli* and we compared them against their *P. putida* counterparts discussed in Chapter 2. Next, we compared their: (i) capability for the production of different cofactors; (ii) displacements from the thermodynamic equilibrium; and (iii) capability for the production of 5 Methyl-Ethyl Ketone precursors.

4.2 Methods

4.2.1 Reduced *P. putida* and *E. coli* models

For the purpose of comparison, we used redGEM [76] and lumpGEM [110] (see Section 2.2.4) to derive three different core reduced models of *E. coli* referred here as D₁ (smallest complexity), D₂ (middle complexity) and D₃ (the most complex model) of *E. coli*'s iJO1366 [29]. Similar to the case of *P. putida*, the reduced models were built around the following subsystems: glycolysis/gluconeogenesis, TCA cycle, pentose phosphate pathway, pyruvate metabolism and oxidative phosphorylation. Lumped reactions are obtained from the $S_{\min}+2$ subnetworks, keeping all the possible alternatives of $S_{\min}+2$ size (Section 2.2.4). *P. putida* models of the same complexity were described in Chapter 2.

4.2.2 Displacements from the thermodynamic equilibrium

To calculate displacement from the thermodynamic equilibrium we used standard TFA [51, 53-55, 57]. For any reaction, the displacement from the thermodynamic equilibrium is a function of the equilibrium constant K_{eq} , defined as a ratio:

$$\Gamma = \frac{1}{K_{eq}} \frac{\prod P_j}{\prod S_i} \quad (4.1)$$

where P_j and S_i are the concentration of the involved products and substrates.

The equilibrium displacement is related to the Gibbs free energy as follows:

$$\Delta G = RT \ln \Gamma \quad (4.2)$$

Reaction can operate (i) strictly far from the equilibrium ($0 < \Gamma < 0.1$), (ii) with the middle displacement ($0.1 < \Gamma < 0.9$) or (iii) strictly near equilibrium ($0.9 < \Gamma < 1$).

Each bi-directional reaction can lead to two distinct flux directionality profiles. We sampled the flux solution space, and we used the principle component analysis to the representative flux directionality profile. For the chosen flux directionality profile, we sampled the space of metabolite concentrations and we computed a population of 10'000 vectors of reaction displacements from thermodynamic equilibrium. In our analysis, we assumed that if 75 % of the displacements of a reaction were lying between 0.9 and 1, that the reaction was considered to be close to the equilibrium. In contrast, if 75 % of samples were between 0 and 0.1, the reaction was considered to be far from the equilibrium. For some reactions, the displacements belonged to more than one of the above-defined groups. In order to have fair comparison of the reaction displacements between the two organisms, we imposed the same directionalities for the bi-directional reactions belonging to both organisms.

4.2.3 Pathway evaluation

We embedded the discovered pathways in the genome-scale model iJN1411, and we performed the pathway evaluation as already discussed in section 3.2.3. We used C1 set of constraints, with the exception of ATP maintenance whose value was set to 0.92 mmol/gDCW/h and it is *P. putida* specific [24].

4.3 Results and Discussion

4.3.1 Stoichiometric comparison of reduced models

Comparison of core networks. Latest GEM of *P. putida*, iJN1411 [91] consists of 2860 reactions separated in 90 subsystems across two compartments (cytosol and periplasm) and extracellular space. In 5 core carbon subsystems (glycolysis, gluconeogenesis, TCA cycle, pentose phosphate pathway, pyruvate metabolism, and oxidative phosphorylation) there were 90 reactions and 99 metabolites (Do core network, for more details see section 2.2.4). In *E. coli* GEM iJO1366 [29], 2585 reactions are separated into 37 subsystems. In the same 5 core carbon subsystems of *E. coli*, there were 128 reactions and 113 metabolites (Do core network). The major difference between the two came from the oxidative phosphorylation (35 reactions in *P. putida*, 70 reactions in *E. coli*) because *P. putida* is lacking menaquinones and demethylmenaquinones, and from glycolysis, because *E. coli* has phosphofructokinase and glycogen related metabolic pathways whereas *P. putida* has not.

Different topologies of *E. coli* and *P. putida* GEMs and differences among central carbon core reactions resulted in differences among reduced models of three different sizes (Table 4.1).

Table 4.1 Comparison of reduced models of *P. putida* and *E. coli*First number represents *P. putida*, second *E. coli*

	D1	D2	D3
Reactions	828 / 857	704 / 866	750 / 903
Core	278 / 342	307 / 435	343 / 465
Lumped	550 / 515	397 / 431	407 / 438
Metabolites	286 / 303	306 / 348	336 / 364
Cytosolic	156 / 160	174 / 188	200 / 199
Periplasmic	70 / 80	71 / 91	74 / 93
Extracellular	60 / 63	61 / 69	62 / 72

The smallest model of *E. coli*, D1EC, consists of 857 reactions and 303 metabolites. Compared to *P. putida* D1 model, D1PP, D1EC has more core reactions (342 compared to 278). We observed that D1EC has 64 core reactions more than D1PP, and D3EC has 122 core reactions more than D1PP. This suggests the better connectivity in the *E. coli* core networks than in the *P. putida* ones.

Comparison of subnetworks for production of biomass building blocks. A different topology of GEMs resulted in differences among the subnetworks and the corresponding lumped reactions for the production of biomass building blocks (BBBs). To produce 86 BBBs of *P. putida*, we generated 550 unique lumped reactions from 1'463 alternative subnetworks (Appendix Chapter 4, Table S1). In contrast, for 78 BBBs of *E. coli*, we generated 1399 alternative subnetworks and 515 unique lumped reactions (Appendix Chapter 4, Table S2). The *P. putida* subnetworks contained 638 unique reactions, while the *E. coli* subnetworks contained 531 unique reactions. Common between these two sets were 379 reactions, 191 were unique for the *P. putida* subnetworks and 152 unique for the *E. coli* subnetworks. The main difference came from the

organism-specific BBBs such as lipopolysaccharide in *P. putida* (48 alternative subnetworks and 16 lumped reactions Appendix Chapter 4, Table S1). For some shared BBBs like lipoate, we generated 179 alternative subnetworks and 37 lumped reactions in *P. putida*, while in *E. coli* we could generate only 25 alternative subnetworks and 22 lumped reactions (Appendix Chapter 4, Table S1 and Table S2).

Comparison of bidirectional reactions. D1PP model has 21 bidirectional reactions (BDRs) when we require at least 90 % of the optimal growth (Appendix Chapter 4, Table S3). In D1EC there are 19 BDRs under the same conditions (Appendix Chapter 4, Table S4). While some of BDRs are common in two organisms, especially in pentose phosphate pathway transaldolase (TALA), both transketolases (TKT1 and TKT2) and ribulose 5-phosphate 3-epimerase (RPE) and in TCA succinyl CoA synthetase (SUCCOAS), some BDRs are specific to only one organism. As a direct consequence of predominate ED pathway in *P. putida*, fructose-bisphosphate aldolase (FBA) and triose-phosphate isomerase (TPI) are unidirectional in D1PP, while in D1EC they are BDRs. Since *E. coli* is using EMP pathway for glucose catabolism, glyceraldehyde-3-phosphate dehydrogenase (GAPD), enolase (ENO), phosphoglycerate kinase (PGK) and phosphoglycerate mutase (PGM) are unidirectional in D1EC, while in D1PP they are BDRs.

4.3.2 Cofactor production costs

Cofactors play a significant role in the production of reduced compounds like biofuels. Here we compared the costs of production of different cofactors in *P. putida* and *E. coli*. As a basis for comparison, we took the minimal network and its alternatives if they exist for the production of a cofactor from the core metabolites belonging to the simplest metabolic network, D1.

The smallest network for the NADPH production in *P. putida* consists of 26 reactions (Table 4.2). There is one additional alternative subnetwork of the same size. Both of the subnetworks generate their own unique lumps. In the first alternative, the cost of production of one molecule of NADPH is 5.5 molecules of ATP (Appendix Chapter 4, Table S5). In this alternative, one molecule of CO₂ is produced as well. The cost of the production of one molecule of NADPH in the second alternative is six molecules of ATP. In both alternatives, the generation of NADPH is followed by the generation of glycine.

The standard Gibbs free energy of these reactions is -42.81 kcal/mol and -42.38 kcal/mol, respectively.

Compared to *P. putida*, the smallest subnetwork for NADPH production in *E. coli* is 27 reactions (Appendix Chapter 4, Table S5). There are 4 alternative subnetworks and each of them generates its own unique lumped reaction (Table 4.2). In two alternatives, it is possible to obtain same ratio ATP/NADPH like in *P. putida*, but without generating CO₂ (the standard Gibbs free energy of these reactions is -44.75 kcal/mol and -44.32 kcal/mol, respectively). Also, the NADPH generation is accompanied by the generation of glycine. The other two alternatives are much more energetically expensive. They require 12 (the standard Gibbs free energy -52.23 kcal/mol) and 13 molecules of ATP per molecule of NADPH (the standard Gibbs free energy -51.83 kcal/mol), respectively, but in this case, glycine is not produced (Appendix Chapter 4, Table S5).

Table 4.2 Comparison between *P. putida* and *E. coli* alternative subnetworks for production of different cofactors.

The first number in the table cells corresponds to *P. putida* and the second one to *E. coli*.

Cofactor	Size of the minimal subnetwork	Number of alternatives	Number of unique lumps
NADPH	26 27	2 4	2 4
Malonyl CoA	52 44	5 6	5 2

Another cofactor, important in the biodiesel production, is Malonyl Coenzyme A. Biodiesel is composed of fatty acid alkyl esters, including fatty acid methyl esters (FAMES) and fatty acid ethyl esters (FAEEs) [200], which are produced in the fatty acid metabolism. In this metabolism, Malonyl CoA plays a notable role [201, 202]. *P. putida*'s minimal subnetwork size for production of CoA is 52 (Table 4.2). There are 5 different alternatives with 5 unique lumped reactions. Production of one molecule of malonyl CoA can be achieved with consumption of 18, 20, 21, 23 or 25 ATP molecules and 5 NADPH. The production of NADPH is accompanied by the formation of 3 or 4 molecules of CO₂. In *E. coli*, the minimal network size is 44, and there are 6 alternative networks

with 2 unique lumped reactions. Production of one molecule of malonyl CoA requires either 19 or 21 ATP molecules, 5 NADPH molecules and one molecule of CO₂ is secreted.

For some important biochemical intermediates like chorismate, *P. putida* and *E. coli* share their production routes. Kuepper *et al.* used chorismite pathway to produce anthranilate from glucose in *P. putida* [203]. Yu *et al.* used the same intermediate pathway for production of para-hydroxy benzoic acid in *P. putida* [204]. For the chorismate production, *P. putida* and *E. coli* share their subnetwork of 7 reactions and one overall lumped reaction (Appendix Chapter 4, Tables S1 and S2).

A detailed comparison of subnetworks for production of each biomass building block can be consulted in Appendix Chapter 4, Tables S1 and S2.

4.3.3 Displacements from the thermodynamic equilibrium

One of the features the thermodynamics-based flux balance analysis brought to the constrained-based modelling is the possibility to calculate reaction displacements from the thermodynamic equilibrium. It is important to calculate the equilibrium displacements because the enzymes that operate near thermodynamic equilibrium have no control on fluxes in a cellular network [68]. Analysis of equilibrium displacements can give insight also about representative regulatory enzymes in the pathway. Kümmel *et al.* [119] identified pfk and pyk as potential regulatory enzymes of glycolysis in *E. coli* since they operated far from equilibrium.

We analyzed the distribution of equilibrium displacements of reactions in a wild-type optimally grown *E. coli*, and we discovered that only 5 cytosolic reactions operate strictly near thermodynamic equilibrium: pyruvate synthase (POR5), flavodoxin reductase (FLDR2), aspartate transaminase (ASPTA), aconitase (ACONTa) and NAD transhydrogenase (NADTHRD) (Fig 4.2). Phosphofructokinase (PFK) and pyruvate kinase (PYK), which have been identified as regulatory enzymes in the previous studies [119, 205], and also the enzymes from the ED pathway were among 513 reactions (169 non-lumped and 344 lumped reactions) that operated far from the thermodynamic equilibrium (Fig 4.2).

In a wild-type optimally grown *P. putida*, only one reaction was operating strictly near equilibrium: Kdo 2 hexa acyl lipid A hydroxylase (LIPAH) (Fig 4.3). Further analysis of the equilibrium displacements revealed that majority of enzymes from the ED pathway as well as hexokinase may serve as potential glycolytic regulatory targets. Sudarsan *et al.* [174] reported upregulation of the regulatory genes (*ptxS* and *lacI*), which control the utilization of glucose via the ketogluconate and gluconate pathways, during the shift from benzoate to glucose. They also observed the upregulation of GltR which controls the glucokinase pathway. Moreover, they reported the induction of the *oadA* gene, encoding the oxaloacetate decarboxylase (OAADC), which was identified to operate far from thermodynamic equilibrium in our analysis.

We compared the equilibrium displacements of reactions in *P. putida* and *E. coli*, and the major observed difference was in glycolytic/gluconeogenic enzymes phosphoglycerate kinase (PGK), phosphoglycerate mutase (PGM) and enolase (ENO). In *P. putida*, PGK and ENO were strictly far from equilibrium and PGM was either far from equilibrium or with the middle displacement. In contrast, in *E. coli* only PGM was strictly far from equilibrium, whereas PGK was operating with the middle displacement and ENO was operating either close to equilibrium or with the middle displacement (Fig 4.2 and Fig 4.3).

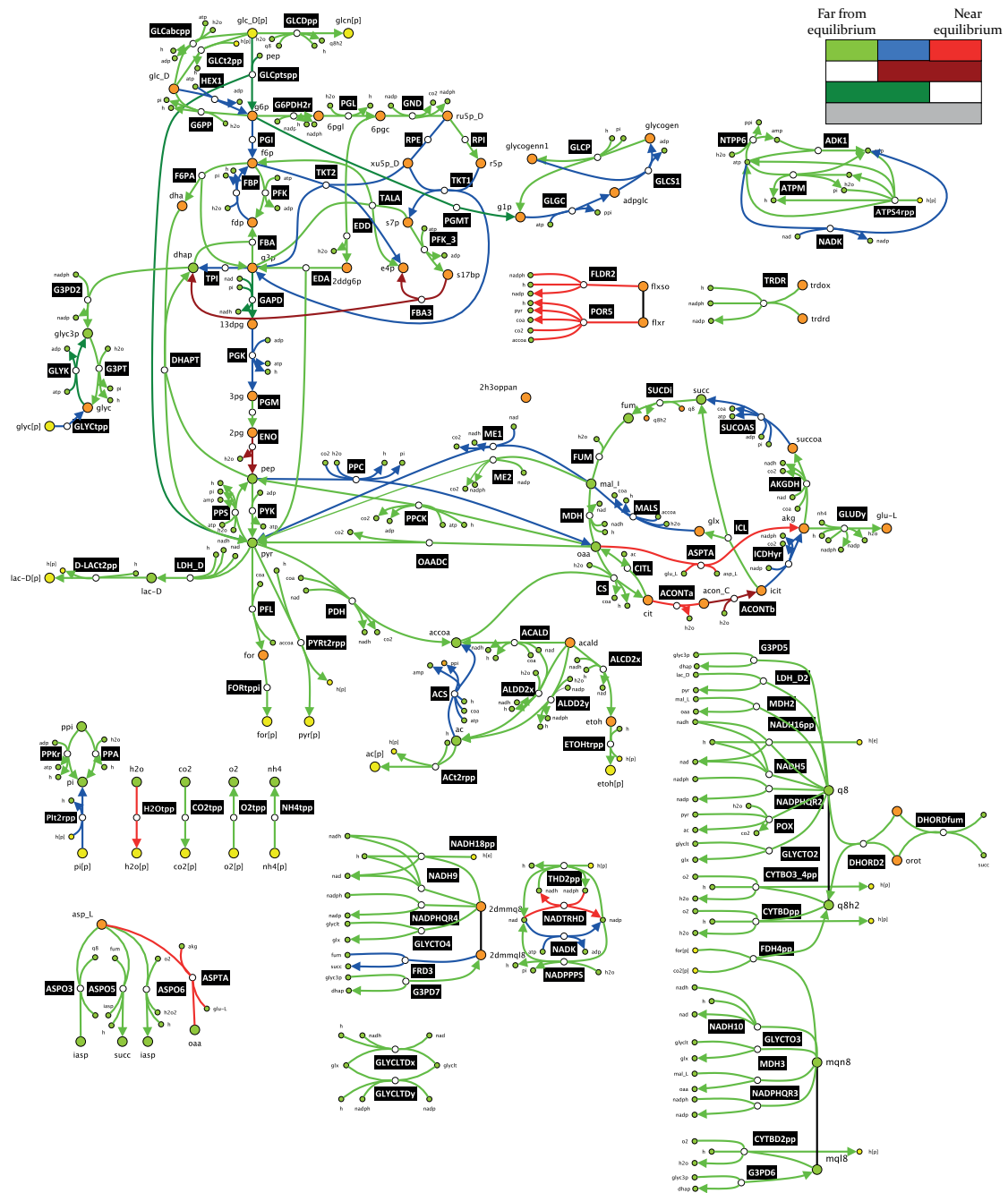


Figure 4.2 Displacements from the thermodynamic equilibrium for *E. coli*
Color-coding of the reactions denotes the distance from the thermodynamic equilibrium

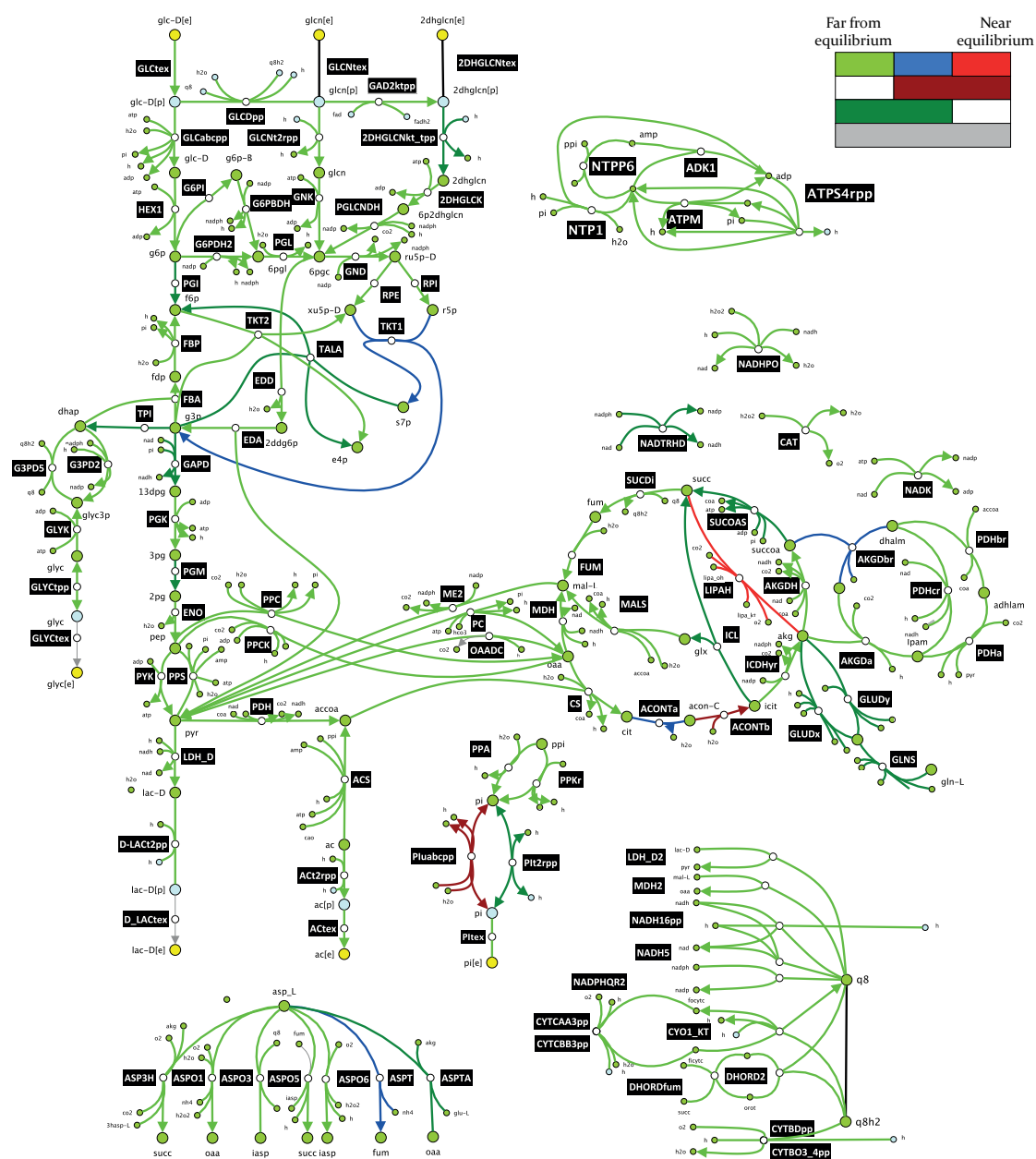


Figure 4.3 Displacements from the thermodynamic equilibrium for *P. putida*
Color-coding of the reactions denotes the distance from the thermodynamic equilibrium

4.3.3 *Pseudomonas putida* as a potential host for production of 5 methyl ethyl ketone precursors

Table 4.3 Reconstructed pathways towards 5 compounds used as a basis for comparison of *P. putida* and *E. coli* capabilities

Target compounds	Reconstructed pathways	Reaction steps				TFA feasible pathways	
		1	2	3	4	<i>P. putida</i>	<i>E. coli</i>
3-oxopentanoate (3OXPNT)	606'181	1	186	11'675	594'319	11'027	10'559
but-3-en-2-one (MVK)	412'737	1	128	7130	405'478	4'699	4'016
Butylamine (BuNH ₂)	1'988'174	28	1'196	49'873	1'937'077	1'053	873
but-1-en-2-olate (1B2OT)	131'093	0	48	2'679	128'366	2'232	1'803
2-hydroxy-2-methyl-butanenitrile (MEKCNH)	311'559	3	94	6'546	304'916	946	658
	3'449'744	33	1'652	77'903	3'370'156	19'957	17'909

Feasibility analysis. To assess *P. putida* capability as an alternative host for production of the second-generation biofuels, we embedded biosynthetic pathways for 5 methyl ethyl ketone precursors reconstructed with BNICE.ch (see Chapter 3) in the thermodynamically curated GEM iJN1411 (see Chapter 2). To ensure a fair comparison of pathways in two organisms, from the initial set of pathways reconstructed from the central carbon metabolites of *E. coli*, we removed the ones starting from 16 metabolites

that are not part of the iJN1411 (Appendix Chapter 4, Table S6). In total, we removed 249'155 pathways from the initial set of 3'679'610 reconstructed biosynthetic pathways in *E. coli* (Chapter 2), and we also removed the corresponding FBA and TFA feasible pathways (Appendix Chapter 4, Table S6). In total, we removed 753 pathways from 18'662 *E. coli* TFA feasible pathways (Table 4.3).

The highest rate of host-specific pathways was for 3-oxopentanoate, 43.7 % (2'061 out of 4'711) for *P. putida* and 59.8 % (1'593 out of 2'663) for *E. coli* (Table 4.4). Out of 4'711 TFA feasible only in *P. putida*, 1'729 (36.7%) were FBA feasible in both organisms (Table 4.4). We observed also that 2'015 out of 2'663 (75.6%) TFA pathways feasible only in *E. coli* were FBA feasible in both organisms (Table 4.4). These findings further confirm our previous statements from Chapter 3 that TFA is a must for reliable biosynthetic pathway evaluation procedures [93].

Yield analysis. We analyzed yields for 15'246 pathways which were TFA feasible in both organisms. In the case of *P. putida*, for three compounds (3OXPNT, MVK, and 1B2OT), we identified pathways that could operate without the loss of carbon from glucose (Table 4.5). We identified such pathways in *E. coli* for all five compounds. For 4 compounds (3OXPNT, MVK, BuNH₂, and 1B2OT) we could find the pathways with the same yield in both organisms. The exception is MEKCNH, where all pathways had higher yield in *E. coli*.

Also, we identified pathways with the higher yield in *P. putida* and there were pathways with a higher yield in *E. coli*. For example, out of 8'966 feasible pathways for production of 3OXPNT in both organisms, 677 had higher yield in *P. putida*, 7'012 in *E. coli*, and 1'277 pathways had the same yield in both organisms (Table 4.5). Moreover, the lowest obtained yield in *P. putida* for this compound was 0.15 g/g, while in *E. coli* it was 0.28 g/g.

Table 4.4 Comparison of TFA feasible host specific pathways

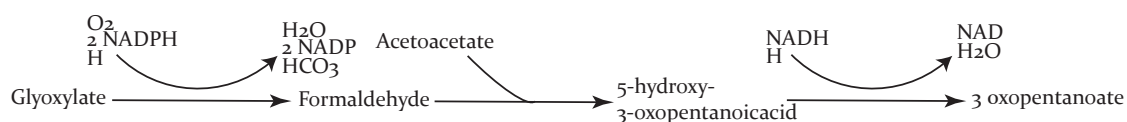
Out of TFA feasible pathways in only one organism, the number in the brackets represent the number of FBA feasible pathways in both organisms. For example, there were 2'061 TFA feasible pathways towards 3OXPNT in *P. putida*. Out of them, 916 were FBA feasible in both *E. coli* and *P. putida*.

Target compounds	TFA feasible pathways only in <i>P. putida</i>	TFA feasible pathways only in <i>E. coli</i>
3-oxopentanoate (3OXPNT)	2'061 (916)	1'593 (1'432)
but-3-en-2-one (MVK)	1'438 (529)	755 (527)
Butylamine (BuNH ₂)	319 (177)	139 (23)
but-1-en-2-olate (1B2OT)	602 (83)	173 (30)
2-hydroxy-2-methyl- butanenitrile (MEKCNH)	291 (24)	3 (3)
	4'711 (1'729)	2'663 (2'015)

Table 4.5 Yield dependency with the respect to the chosen host

Target compounds	TFA feasible pathways in both organisms	Yield higher in <i>P. putida</i>	Yield same in both	Yield higher in <i>E. coli</i>	Yield ranges in <i>P. putida</i> [g/g]	Yield ranges in <i>E. coli</i> [g/g]
3-oxopentanoate (3OXPNP)	8'966	677	1'277	7'012	0.15-0.77	0.28-0.77
but-3-en-2-one (MVK)	3'261	1'196	396	1'669	0.13 – 0.58	0.19 – 0.58
Butylamine (BuNH ₂)	734	40	158	536	0.10 - 0.41	0.17 – 0.61
but-1-en-2-olate (1B2OT)	1'630	251	89	1'290	0.2 - 0.59	0.2 – 0.59
2-hydroxy-2-methyl-butanenitrile (MEKCNH)	655	/	/	655	0.32 - 0.44	0.41 - 0.66

Subnetwork analysis. An important question in understanding the differences in metabolic capabilities of *E. coli* and *P. putida* is why certain pathways have different yields in the two organisms. As the first glimpse into this question, we analyzed the anabolic subnetworks (see 3.2.4) corresponding to a three-step pathway that converts glyoxylate to 3OXPNP, with the maximal predicted yield of 0.72 g/g in *P. putida* and 0.66 g/g in *E. coli*:



We first computed the minimal anabolic subnetwork for this pathway in *P. putida* and its alternatives. The minimal network contained 37 reactions, with 24 alternatives of the

same size. The 24 alternatives had only 2 unique lumped reactions. In *E. coli*, the minimal anabolic subnetwork contained 13 reactions, with two alternatives that had its own unique lumped reaction (Appendix Chapter 4, Table S7). The main differences between the two sets of anabolic networks (one set in *P. putida* and one set in *E. coli*) came from the inability of *P. putida* to phosphorylate nucleosides to 2 deoxy-D-ribose-1-phosphate (in *E. coli* this pathway is described in Chapter 3, Fig 3.4). Consequently, *P. putida* and *E. coli* are using different sets of core precursors. *P. putida* is using pyruvate, L-aspartate, D-erythrose 4 phosphate and phosphoenolpyruvate, whereas *E. coli* is using D-ribose 5-phosphate and acetate (Appendix Chapter 4, Table S7). In Chapter 3, we have demonstrated that yields depend on the precursors (Fig 3.6), which suggests the answer to the addressed research question. A similar question was addressed in the work by Asplund-Samuelsson *et al* [93].

4.4 Conclusions

The selection of the suitable host organism is one of the critical points in metabolic engineering. While *E. coli* was for a long time, and in many cases still is, the first choice, recent studies in metabolic engineering and synthetic biology demonstrated that some other organisms can be as suitable hosts as *E. coli*, especially for the production of highly toxic compounds such as 2nd generation biofuels. Analysis of reduced models derived by redGEM [76] and lumGEM [110] allowed us to analyze fundamental differences in metabolic capacities between *E. coli* and *P. putida*. We showed that differences exist already at the level of central carbon pathways and that routes for production of BBBs are different for the two organisms. Analysis of thermodynamics displacements revealed differences in potential regulatory targets.

We have already shown in Chapter 3 and in Tokic *et al.* [92] that thermodynamic feasibility of individual pathways cannot be studied outside the metabolic network. In order to assess the thermodynamic feasibility of reaction steps within a pathway, we need to embed it into a genome-scale model of the host organism. In the study conducted in this chapter, we discovered that there are pathways with the higher yield in *P. putida* than in *E. coli* for four target molecules. We demonstrated that simple stoichiometric models are not enough and that pathway evaluation process has to be

performed in the thermodynamically curated models. Moreover, we have shown that thermodynamics is a host specific. There is no guarantee that if a pathway is thermodynamically feasible in one host, it will be feasible in the other host. Furthermore, if a pathway is feasible in multiple organisms, it might well be that it will not have the same yield due to, e.g., the different cofactor production costs.

Chapter 5. Large-scale kinetic metabolic models of *Pseudomonas putida* for consistent design of metabolic engineering strategies

5.1 Introduction

Kinetic models couple dynamics of metabolic concentrations and metabolic fluxes to enzyme concentrations and they allow us to take into consideration allosteric regulation and regulation at post-translational level [206]. Although the predictive capabilities and potential of kinetic models compared to their stoichiometric counterparts are promising, they come at a price. These dynamical descriptions of metabolic pathways are typically built in a bottom-up manner, wherein for each reaction a description of kinetic rate expression along with its corresponding parameter values is required. This results in model structures with large number of parameters. Due to the absence of large-scale experimental assays that could provide the required extent and diversity of measurements for the rigorous parameterization of these models, researchers incorporate the needed information from different sources: (i) literature; (ii) databases such as Brenda and SABIO-RK [207, 208]; or (iii) they perform experimental measurements themselves [74, 209, 210]. Whenever the model parameters are not experimentally measured, parametric estimation methods [211] or Monte Carlo methods are used [69, 181, 212, 213]. In the latter, the parameters are characterized within well-defined bounds that are consistent with the studied conditions and physicochemical laws. The available experimental values of kinetic parameters are often uncertain due to

variations stemming from different experimental conditions and set-ups and due to measurement and estimation errors [78]. As a consequence, many existing kinetic models are of a limited scope, often with *ad hoc* stoichiometry, they cover one or a few metabolic pathways, and frequently they neglect the whole network dynamics as observed in [206, 214].

Recent efforts have been made towards building large- and genome-scale kinetic models [34, 54, 80, 181, 215-219]. In the quest for the models with a large- or genome-scale scope, one must ensure that the increased size and scope is attained without sacrificing the consistency with physicochemical laws and the necessary mechanistic details.

5.1.1 Issues in building kinetic models

When building kinetic models, we are given a set of observation and we seek to identify the set of kinetic parameters that best describe the observations. In metabolic kinetic models we usually start with a set of metabolic fluxes (extracellular and intracellular) and a set of metabolite concentrations (extracellular and intracellular). We also assume that the stoichiometry of the metabolic network and the thermodynamic properties of the reactions in the network are known. The basic problem then is to identify a kinetic model or a set of kinetic models that consistently describe the experimental observations. We discuss here briefly the main issues in building kinetic models.

5.1.1.a Uncertainty

Uncertainty is widely recognized in the literature as the main, persisting challenge in kinetic modeling of biological systems [69, 206, 214]. The dynamic behavior of metabolism is a result of complex interactions of metabolite concentrations, through kinetics and thermodynamics, and uncertainty in these interactions propagate to the structure and parameters of the kinetic models.

- **Uncertainty in kinetic properties of enzymes**

We can distinguish two types of uncertainty in kinetic properties of enzymes: structural and quantitative [69, 78-80]. *Structural uncertainty* is associated to the missing information about kinetic mechanisms and allosteric regulations. *Quantitative uncertainty* refers to the lack of knowledge about the values of kinetic

data and inconsistencies in the available kinetic data. While the databases that collect and organize the information about kinetic parameters are growing in size [207, 208], the available kinetic data are not standardized. Indeed, kinetic data are coming from different sources and they are measured under different experimental conditions. As a result, the values of available kinetic parameters often range within several orders of magnitude. Furthermore, important factors that impact the values of kinetic parameters such as temperature or pH are frequently not reported in the databases or in the literature. An additional issue is a very disputed question if the values of the kinetic parameters that are quantified *in vitro* and for each enzyme separately can represent well the behavior of a multitude of enzymes interacting in a crowded *in vivo* environment [80, 209, 220, 221].

- **Uncertainty in metabolic fluxes**

Despite the availability of abundant fluxomics data, complex topology of metabolic networks, e.g. existence of a large number of branching points, prohibits determination of the exact values and directionality of intracellular metabolic fluxes [53, 56]. This translates into the existence of multiple alternative flux profiles that are consistent with the measured data but with uncertainty in determining a *unique* flux profile.

- **Uncertainty in metabolite concentration levels and thermodynamic properties**

The introduction of Second Law of Thermodynamics in the context of flux balance analysis (FBA) allows coupling the directionality of the fluxes and the levels of metabolite concentrations [51, 53, 56]. Using the expressions for the Gibbs energy of reactions to formulate the additional set of thermodynamic constraints on feasible solution space we can eliminate thermodynamically infeasible flux directionalities. At the same time, this enables integration of metabolomics data in the constraint-based analyses [54]. However, the thermodynamic properties (Gibbs free energies) of many reactions are not measured, and instead, they are estimated using group contribution methods [58]. These estimates contain both measurement and estimation errors, and together with uncertainties in metabolite concentration

measurements they can affect the conclusions about directionalities of reactions and eventually the conclusions about cellular physiology.

5.1.1.b Size and content of metabolic networks

As the main purpose of the models is the understanding of system-wide properties, we need large models in order to capture the interactions determining the behavior of the system as a whole. The size of a model introduces a tradeoff between the accuracy of the models that comes from the description of all possible and important interactions and the number of unknown and uncertain parameters introduced into the model. However, there are issues to be considered when large- and genome-scale kinetic models are constructed.

- **Large number of unknown parameters, sloppiness and overfitting**

As the size of the metabolic network increases, the portion of available experimentally measured values for kinetic parameters is rapidly going down. Consequently, a large number of unknown parameter values have to be determined. A common way to find the missing parameter values is to use the parameter estimation techniques [211]. However, due to a large number of parameters to be estimated, uncertainty in available data, and the intrinsic sloppiness of parametric models in systems biology [222, 223] there are multiple sets of parameters consistent with data, i.e. it is impossible to compute unique parameter values. Considering that available data in biology are usually scarce, and when the number of parameters is large relative to the number of observations, the obtained models tend to describe measurement errors rather than functional relationships within the modeled process (overfitting). As a result, poor predictions are obtained when these models are validated against independent data sets.

- **Issues with parameter estimation methods**

Parameter estimation methods use optimization procedures to obtain the values of parameters. Depending on the underlying formulation, network structure and employed optimization technique parameter estimation might become computationally intractable for large metabolic networks [224].

- **Issues with Monte Carlo methods**

In Monte Carlo based methods the admissible parameter space is constrained with physicochemical and thermodynamic laws along with the constraints obtained from available measurements, and then a population of alternative parameter sets is drawn from such a reduced solution space [54, 69, 181, 212, 213, 217, 225, 226]. Sampling of such constrained admissible parameter space is a computationally daunting task for large metabolic networks. Another important challenge is that an efficient sampling necessitates well-defined bounds on kinetic parameters such as Michaelis constants, and these bounds are rarely known. To address these issues a new tailor-made formulation and a new sampling technique were proposed in [80].

- **Stiffness of metabolism dynamics**

Large- and genome-scale kinetic models of metabolism provide an increased level of details about dynamics of metabolism. As such, they are stiff systems of ordinary differential equations (ODE) since they span over metabolic reactions with a wide range of rate dynamics. The stiffness of these systems and the intrinsic nonlinearities of the kinetic rate expressions will require advanced computational tools for robust simulation of these models [227].

5.1.2. Standard Requirements in Kinetic Modeling

The structure and the complexity of a kinetic model should be adjusted to the modeling goal, to the characteristics of the organism and to the physiological conditions of the system under study. However, there is a set of conditions and terms that a kinetic model must follow in order to fulfill the required quality requirements. These conditions will preserve the most of the prior knowledge about biochemistry and cellular physiology, and they will avoid biases that can be introduced during the assembly of the kinetic model.

5.1.2.a Consistent pathway/network stoichiometry

The lack of knowledge about kinetic parameters and the difficulties in parameter estimation for large-scale networks have led to kinetic models that are constrained to few pathways, and often with a reduced level of details both in stoichiometry and in

kinetic rate expressions. As discussed in [228], since the metabolism is a highly interconnected network by disregarding certain parts of a network one potentially neglects dynamics that is crucial for the behavior of the whole system. In addition, ignoring so-called “small metabolites” has also important consequences on modeling conclusions. For instance, there are studies that demonstrate the importance of considering phosphate in the kinetic models [229, 230]. Thus, in balancing between the complexity of the model and its predictive capabilities one has to ensure that the elemental and charge balance is maintained irrespectively of the size of the modeled metabolic pathways.

One of the most important questions in metabolic modeling, and especially in building kinetic models of metabolism, is how the model size impacts its quality. While small models might have provided some insights [231], the bias introduced by the *ad hoc* choice of model size can always contaminate the results and limit the predictive strength of the models as well as the confidence and reliability of the conclusions and learning obtained using these models. Physiologically relevant kinetic models should be built on a scaffold of a **context specific stoichiometric models** with the same stoichiometric detail and consistency.

A critical quality check for a kinetic model is how well it represents the stoichiometric model it has used as a scaffold and how well it accounts for those parts of the stoichiometric models the modeler has chosen to omit from the stoichiometric description. Most of the current kinetic models use the stoichiometry without accounting consistently for many reactions around the pathways they attempt to model [228, 232-234]. One could argue that such models simulate the dynamics of mutant organisms with the knockouts of the reactions omitted from their model.

Researchers should consider an elementally and charge balanced stoichiometry that focuses on the studied metabolic pathway(s), and simultaneously it is consistent with the genome-scale metabolic reconstructions (GEMs) it was derived from. Along this direction, Ataman *et al.* [76] have developed an algorithm called redGEM that is capable to systematically produce reduced context specific stoichiometric models from GEMs.

5.1.2.b Consistency with physicochemical laws and experimental data

The computed kinetic models have to be consistent with the observed experimental data from flux and metabolite measurements. While the consistency with stoichiometric models will guarantee elemental and species balance, we must ensure the thermodynamic consistency of the integrated metabolomics data that will also determine the directionalities of the reactions. This way, the uncertainty in models is reduced and reaction directionalities that are not consistent with the observed physiology are discarded [53]. The conservation of the physicochemical laws is also very critical for the performance of the models since they introduce fundamental physicochemical constraints similar to the typical metabolite balance constraints. Palsson and Lee have demonstrated that neglecting electro-neutrality and osmotic balances for the case of a red blood cell model can lead to erroneous interpretation and analysis of the studied physiology [235].

5.1.2.c Appropriate kinetic descriptions and regulation

The purpose of the model, the computational complexity, and prior knowledge and experience of the researchers determine the choice of kinetic rate laws, e.g. mass action vs. generalized mass action, and the mechanistic details, e.g. allosteric regulation. Simplified rate expressions are more frequently encountered in large-scale kinetic models as they normally require less parameters compared to more complex kinetic descriptions. On the other hand, as it is anticipated that kinetic models with oversimplified rate laws and neglected allosteric regulations can have limited predictive capabilities [231, 236]. Within this context, Chakrabarti *et al.* have recently shown that for a specific concentration and flux profile in aerobically grown *E. coli*, it was not possible to find parameters for a kinetic model with the mass-action kinetics for the overall reaction that do account for the enzyme maximum activity. However, when they considered kinetic models with detailed mechanistic rate laws, they found a large population of models that could describe the observed physiology [181]. These studies suggest that the first and foremost requirement is to use kinetic rate laws that are able

to model enzyme saturations. The second important consideration will be to model the allosteric regulation whenever possible [209, 231].

Analyses of a large class of nonlinear systems have shown that only a reduced set of parameters determines the system behavior. However, the actual parameters that constitute this set depend on the particular state of the system, and these can change if the system moves away from the reference condition. An approach in modeling metabolic networks is to model in detail only the kinetics of reactions which have been shown to determine the system behavior, and to use simple approximate kinetic laws for the remaining reactions [34, 206]. While this idea is appealing, we have to know in advance which reactions are important for a particular physiological condition. Therefore, we suggest that we use an efficient method to generate a representative class of nonlinear models, perform next such dynamic model reduction, and then generate a larger population of reduced models, which can be a combination of kinetic models with stoichiometric, flux balanced, subsystems.

Such kinetic models that couple kinetic and stoichiometric models have appeared recently [226]. However, they are based on rather strong and potentially misleading assumptions because they use an *ad hoc* reduction of the kinetic scale without prior knowledge about the steady-state and dynamic properties of the overall system they attempt to model. As we discussed earlier, one must first construct a detailed mechanistic model, show that some parts of the model operate near steady state (quasi-steady state), and then to model these parts by stoichiometry only. Therefore, hybrid models that combine subsystems of kinetic models with subsystems of stoichiometric models should be used with caution and they must follow a rigorous model reduction procedure.

5.1.2.d Compartmentalization

In eukaryotic cells many metabolic pathways involve reactions from more than one compartment, and many metabolites are produced in one compartment and then transported to other compartments [237]. The activity of these pathways that span across several compartments depend on energy, redox and cofactor availability in each

of involved compartments [238]. Although researchers are tempted to remove compartments from the models, it was demonstrated that compartmentalization in constraint based models is an important feature [239]. Specifically, it was shown that removal of compartments from the model had significant effects on energy-related pathways due to disruption of concentration gradients between compartments. While the modeling of the cellular compartments must be done during the development of the stoichiometric models, the modeling of the kinetics of the process across compartments is critical for the performance of the kinetic model.

5.1.3 Current scope, level of details and consistency in kinetic models

The majority of recent studies that involves kinetic models are focused on *S. cerevisiae* [34, 37, 54, 67, 209, 210, 226, 240, 241] and *E. coli* [35, 181, 213, 217, 219, 225, 228, 242-248], and more recently kinetic models for other organisms such as *B. subtilis* [234], *P. falciparum* [74], *P. knowlesi* [233], *Z. mobilis* [232], *L. Lactis* [212, 223, 229], *S. pyogenes* [229], *C. acetobutylicum* [249], *T. brucei* [230], CHO cells [241], *C. thermocellum* [184], *P. pastoris* [250] and human cell [251, 252] have appeared. Most of these models are focused around glycolysis [37, 74, 209, 210, 229, 232] or in combination with either pentose phosphate pathway and/or citric acid cycle [35, 226, 230, 242-244, 246, 247, 249]. Some models are focused around a specific pathway of interest, e.g. Entner-Doudoroff pathway [232], phospholipid [233] or riboflavin [234] synthesis, and only a few models have a broader scope [34, 54, 67, 181, 184, 217, 219, 225, 248, 252]. However, the choice of the reactions and the assembly of the stoichiometric networks are based on *ad hoc* choices and almost all of the models are missing important interactions. For example, many of these models include the dynamic mass balances of two metabolites, and they omit some reactions that connect *directly* these metabolites without providing proper justification.

There are attempts to reduce complexity by combining kinetic and stoichiometric models [226, 233], or alternatively the number of model parameters is reduced by considering the concentrations of co-factors and small molecules such as CO₂ and phosphate to be constant in the rate expressions [245, 251]. In some studies these

molecules are completely overlooked in kinetic descriptions [241, 243, 249]. Another attempt to reduce the complexity of the metabolic models of eukaryotes is to neglect the reactions and species across and within compartments. Indeed, only very few of the models [54, 206, 230, 251, 252] that are used to study eukaryote organisms include compartments for the different organelles. However, all these attempts reduce complexity at the cost of elemental balances and thermodynamic consistency, and only a few kinetic modeling approaches provide thermodynamically consistent models [34, 54, 67, 181, 217, 219, 225, 248, 252].

5.1.4 Review of published kinetic models of *P. putida*

Multiple small-scale kinetic models of *P. putida* metabolism used either of Monod, Haldane and Andrews kinetics for modeling the growth and changes in extracellular concentrations [253-262]. Bandyopadhyay *et al.* used a simple Monod model to study the effect of phenol degradation [255]. Wang and Loh modelled the co-metabolism of phenol and 4-chlorophenol in the presence of sodium glutamate [262], and their kinetic model accounted for cell growth, the toxicity of 4-chlorophenol, and cross-inhibitions among the three substrates. Other models were used for studying growth during benzene [253], toluene [253, 257-259, 261] and phenol biodegradation [253], growth and biosynthesis of medium-chain-length poly-(3-hydroxyalkanoates) [254] and dibenzothiophene desulfurization [256, 260].

More recently, Sudarsan *et al.* developed a kinetic model of the β -ketoadipate pathway that contained mass balance equations for both extracellular and intracellular metabolites described by mechanistic rate expressions based on in vitro investigation of the participating enzymes [263]. Chavarria *et al.* modelled the dynamics of fructose uptake while taking into account the dynamics of gene expression, protein stability, enzymatic activity and the concentrations of intracellular and extracellular metabolites [264].

In all these cases, the models were of limited size and with *ad hoc* stoichiometry, which accentuated a need for developing large-scale kinetic models capable of reliably identifying metabolic engineering targets for production of the desired chemicals [180]. However, construction of large-scale kinetic models remains a challenging task. Each

reaction in a kinetic model requires a matching kinetic rate expression along with values of kinetic parameters, which are frequently unknown. Moreover, even if the values of kinetic parameters are available in the literature and databases, their ranges are often reported within several orders of magnitude. Additionally, partial experimental fluxomic and metabolomic data and estimation errors in related thermodynamic properties [180] hinder us from determining unique steady-state metabolic fluxes and metabolite concentrations. As a consequence, we are unable to find a unique model capable of describing the observed physiology. Instead, to overcome this issue, a population of kinetic models is constructed, and statistical methods are used to analyze and predict the metabolic responses in the system.

In this chapter, we used the ORACLE framework [67, 69, 80] to generate a population of large-scale kinetic models of *P. putida* around the computed steady state, and we employed these models in two studies. In the first study, for wild-type strain of *P. putida* KT2440 grown under aerobic conditions using glucose as a carbon source, we evaluated and validated the predictions of the generated kinetic models against a collection of experimental single-gene knockouts [265]. In the second study, we analyzed the capacity of *P. putida* to adapt to increased energy demand [24], and we identified potential metabolic engineering targets for improved resistance of this organism to stress conditions.

5.2 Methods

5.2.1 Configuring stoichiometric model for kinetic studies of wild-type physiology

We used D2 model described in Chapter 2 as a scaffold for constructing a population of thermodynamically feasible kinetic models. We first expanded D2 by allowing free diffusion to extracellular space for all intracellular metabolites that: (i) have less than 10 carbon atoms and do not contain phosphate or CoA; and (ii) do not have an existing transport reaction in the model. This was done to model a possibility that small amounts of these metabolites were produced during fermentation but in insufficient quantities

for experimental detection. The expanded model contained 752 reactions and 331 metabolites across cytosol, periplasm, and extracellular space.

Based on the data provided in del Castillo *et al.* [111], we integrated into the model the experimentally measured rates of glucose uptake and biomass growth and we forced the secretion of D-gluconate and 2-dehydro-D-gluconate by putting a lower bound on their exchange reactions to 0.3 mmol/gDCW/hr. For the remaining carbon-based by-products, we allowed only their basal secretion by constraining their transport rates to the extracellular space (10^{-6} - 10^{-3} mmol/gDCW/hr) following the common observation in the literature that *P. putida* can break the carbon down almost without any by-product formation [24].

We next integrated 57 experimentally measured intracellular metabolite concentrations [112]. In the model, 12 out of the 57 measured metabolites appear in both cytosol and periplasm. The concentration values of these 12 metabolites were measured per cell and not per compartments, and therefore to integrate this information for each species in the two compartments 2 additional constraints per metabolite were added in TFA (Methods 2.2.2). Overall, these 57 measurements provided constraints for 69 metabolite concentrations in the model.

Furthermore, we configured the model with several additional assumptions: (i) TCA cycle was complete [24, 116]; (ii) two glutamate dehydrogenases (GLUDx and GLUDy), were operating towards production of L-glutamate, (iii) dihydrolipoamide S-succinyltransferase was generating NADH from NAD⁺ [266]. Since kinetic models in ORACLE framework [67, 69, 78-80, 181, 248] are built around uniquely described flux profile, additional assumptions for the remaining 8 bi-directional reactions (Figure 5.1) are obtained from the result of the two step optimization in which biomass yield is maximized subject to minimization of sum of the fluxes.

Each of the bidirectional reactions yields two distinct flux directionality profiles (FDPs), in one a reaction is operating in the forward direction, in the second it is operational in the backward direction. In theory, a maximal number of FDP is equal to $2^{\text{number of bidirectional reactions}}$, but usually, this number is smaller, considering different stoichiometric and flux coupling [267, 268]. In this case, 8 BDRs yielded 48 different

FDPs. For each FDP, we first calculated minimal sum of the fluxes. This sum is further imposed as an additional constraint in the TFA. We then performed additional optimization in which we maximized biomass growth. Out of the resulting 48 flux directionality profiles (FDPs), we chose the one with the minimal sum of fluxes (Figure 5.2). The directionality of the resulting flux profile was imposed into the stoichiometric model: (i) acetaldehyde dehydrogenase (ACALD) was producing acetaldehyde, (ii) ribulose 5-phosphate 3-epimerase (RPE) was converting D-ribulose 5-phosphate to D-xylulose 5-phosphate; (iii) transaldolase (TALA) was operating in the direction of consuming glyceraldehyde 3-phosphate while transketolase 1 (TKT1) was operating towards production of glyceraldehyde 3-phosphate, (iv) adenylate kinase (ADK1) and nucleoside-diphosphate kinase (NDPK1) were consuming ATP; (iv) GTP-dependent adenylate kinase (ADK3) was consuming GTP, and (v) sodium antiporter (Nat3_15) was importing sodium to the cell.

We performed TFA with so configured model, and all reaction directionalities within the obtained thermodynamically feasible steady-state flux and metabolite concentration profile were in agreement with the pre-assigned directionalities from iJN1411 [91].

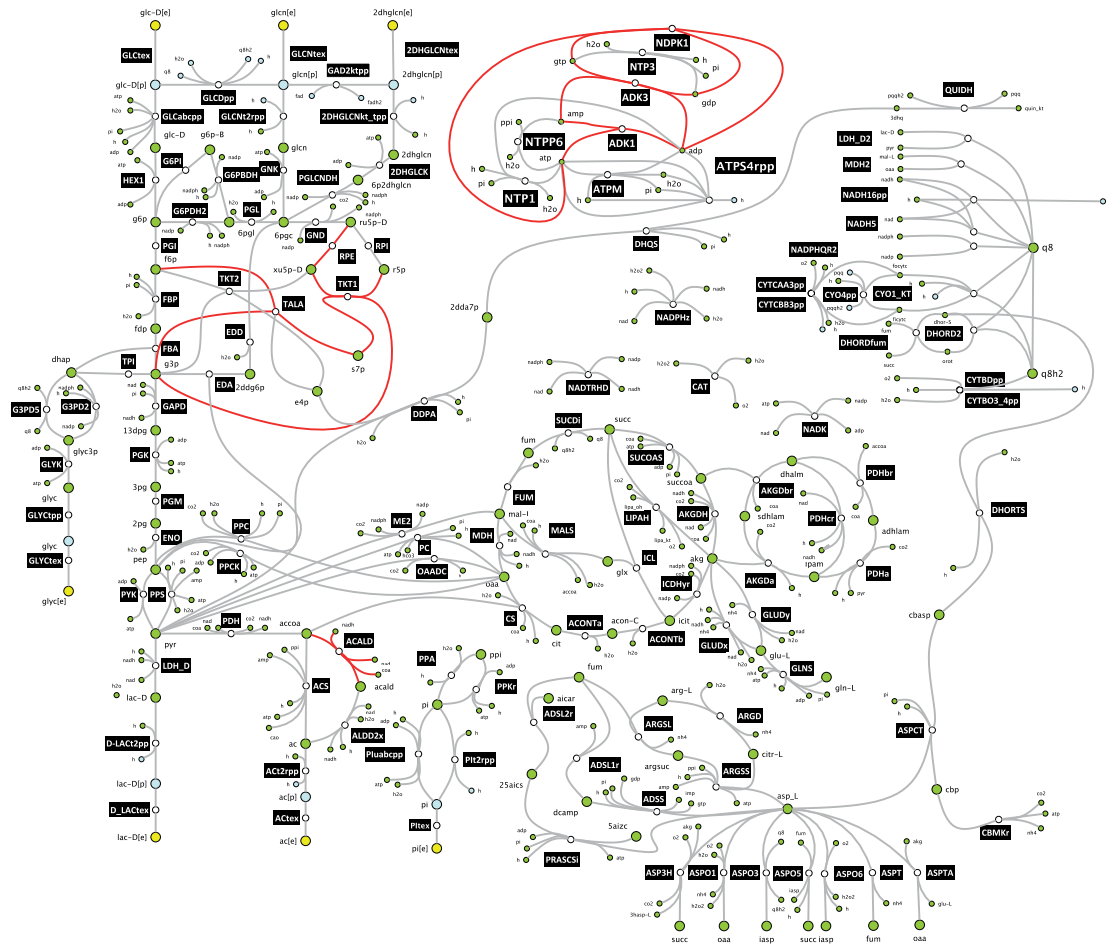


Figure 5.1 8 bidirectional reactions (BDRs) from D2 used to compute 48 feasible flux directionality profiles

BDRs: acetaldehyde dehydrogenase (ACALD), ribulose 5-phosphate 3-epimerase (RPE), transaldolase (TALA), transketolase 1 (TKT1), adenylate kinase (ADK1), nucleoside-diphosphate kinase (NDPK1), GTP-dependent adenylate kinase (ADK3), sodium proton antiporter (NAT3_1pp).

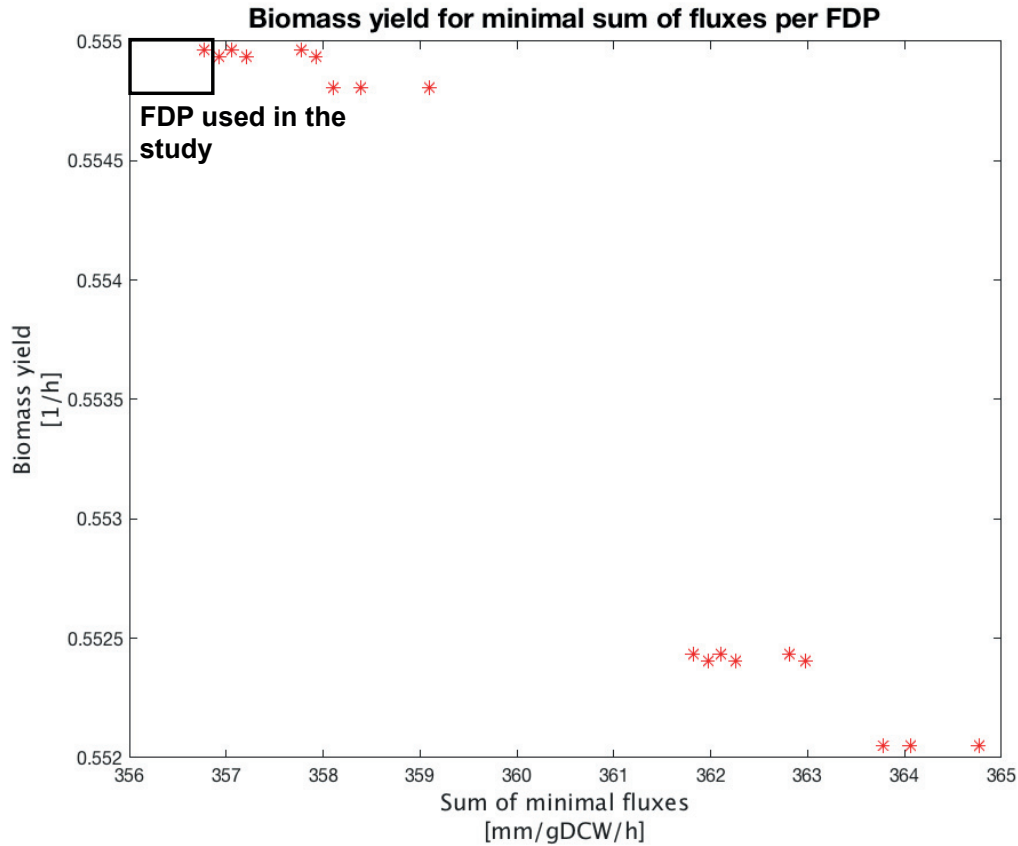


Figure 5.2 Biomass yield for minimal sum of fluxes per FDP

Each red star represents one or multiple FDPs. In total 48 FDPs are plotted.

5.2.2 Configuring stoichiometric model for kinetic studies of stress conditions

5.2.2.a Sensitivity analysis of a model response to changes in the oxygen uptake rates and flux changes in ATP synthesis

Depending on physiological conditions, maximal rates of oxygen uptake and ATP synthase in *P. putida* can take a wide range of values. For instance, in optimally grown *P. putida*, oxygen uptake rate is about 15 mm/gDCW/h [83], while in the stress conditions it can go above 50 mm/gDCW/h [24]. To investigate the effects of the maximal rates on model predictions, we constrained upper bound on biomass growth

to 0.73 1/h and we performed multiple TFAs for different combinations of maximal allowed rates of oxygen uptake and ATP synthesis.

We varied the allowed maximal oxygen uptake between 30 and 70 mmol/gDCW/h (the range between 40 and 60 mmol/gDCW/h was reported in [24]), and the allowed maximal flux through ATP synthase between 40 to 100 mmol/gDCW/h. For each combination of oxygen uptake/ATP synthase maximal rates, we computed changes of minimal required glucose uptake with the respect to changes in flux through ATP hydrolysis (Figure 7).

For the allowed maximal oxygen uptake of 30 mmol/gDCW/h, the peak of the minimal glucose uptake rate was at 10.22 mmol/gDCW/h, which is slightly under the value reported in Ebert *et al.* [24] (11.6 ± 1.2 mmol/gDCW/h) (Figure 5.3). For the allowed maximal oxygen uptake of 40 mmol/gDCW/h, the peak of the minimal glucose uptake rate was at 11.89 mmol/gDCW/h which was within the bounds reported in [24], whereas for the allowed maximal oxygen uptake of 50 mmol/gDCW/h, the peak of minimal glucose uptake rate was above the experimental values (13.56 mmol/gDCW/h). Consequently, we used the bound on allowed maximal oxygen uptake rate of 40 mmol/gDCW/h for our kinetic studies.

Interestingly, the constraint on the allowed maximal ATP synthase rate did not have an effect on the magnitude of the peak value of the minimal glucose uptake rate. Instead, it affected the position of the peak with the respect to the ATP hydrolysis flux (Fig. 7). The higher the ATP synthase rate, the higher ATP hydrolysis flux was required to attain the peak value of the minimal glucose uptake. For example, in the case of the allowed maximal oxygen uptake of 30 mmol/gDCW/h, the ATP hydrolysis flux of 9 and 19 mmol/gDCW/h was required to attain the peak of the minimal glucose uptake of 10.22 mmol/gDCW/h for the allowed maximal ATP synthase rates of 40 and 50 mmol/gDCW/h, respectively. Based on these observations and comparison with the experimental data, one can equally consider values of 50, 60 or 70 mmol/gDCW/h for the upper bound on ATP synthase since all three values describe qualitatively well the experimental data [24] (Fig. 5.3 and 5.8). We set the upper bound of ATP synthase to 70 mmol/gDCW/h to keep the maximal flexibility in the model.

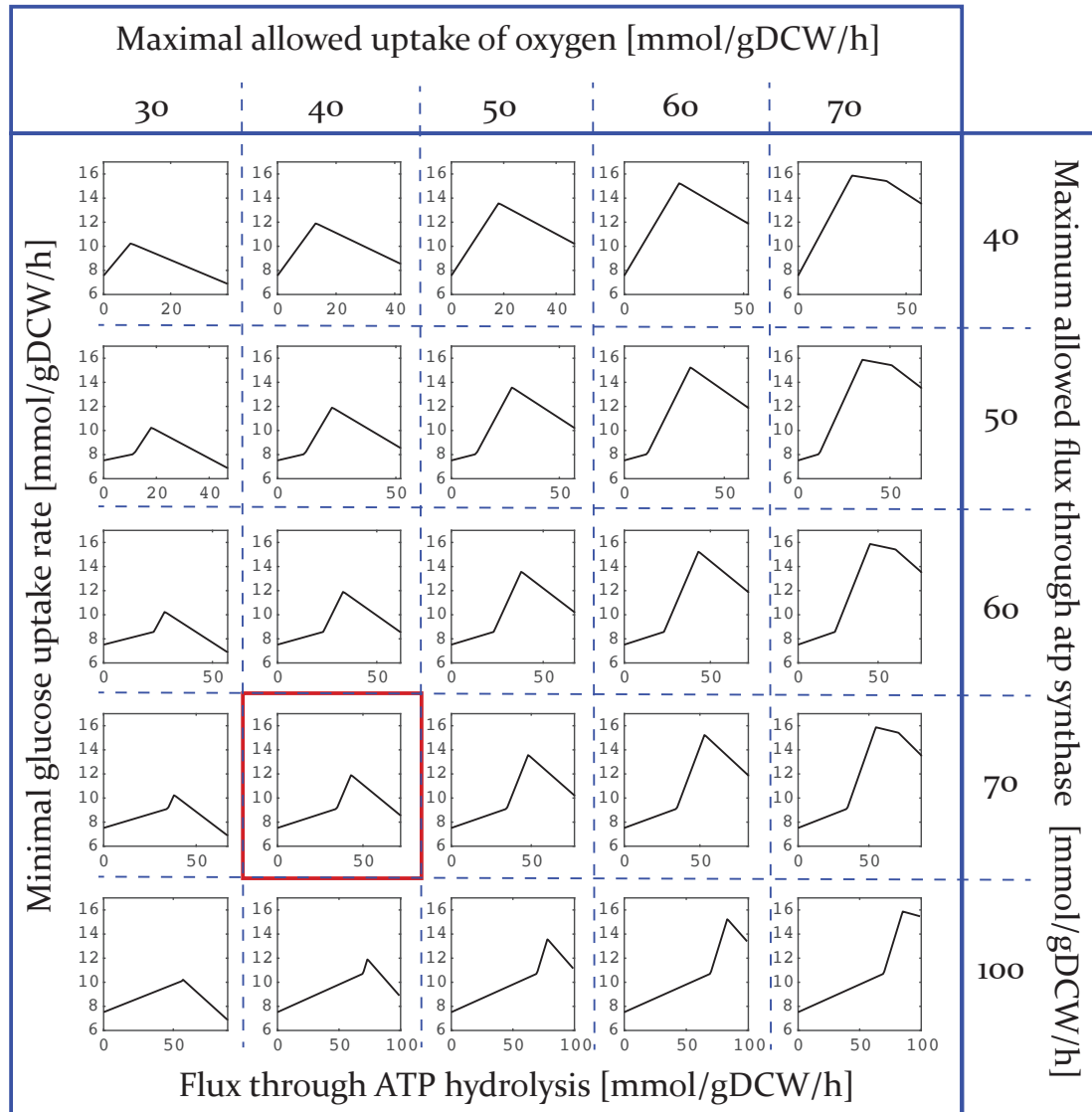


Figure 5.3 Minimal glucose uptake rate as a function of ATP hydrolysis flux for different combinations of allowed maximal rates of the oxygen uptake and ATP synthesis. The sensitivity analysis indicates that models with the maximal oxygen uptake rate of 40 mmol/gDCW/h and the ATP synthesis rate of 70 mmol/gDCW/h (red box) are providing the best qualitative agreement with the experimental data [24] while maintaining the model flexibility.

5.2.2b Configuration of a model

The stoichiometric model was reconfigured in the following way: (i) we constrained the specific growth rate in the range 0.43 ± 0.2 1/h and the glucose uptake in the range 11.6 ± 1.2 mmol/gDCW/h. These values correspond to the concentration of 700 mg/liter of

DNP in the experimental study or 44 mmol/gDCW/h in the simulation study (Figure 5.8d); (ii) the directionalities of 26 reactions from the glycolysis, gluconeogenesis, PPP and TCA were constrained by putting lower and upper bounds from Ebert *et al.* [24]. Interestingly, the reported directionality of TKT2 in this physiological condition was opposite than it was assumed in the study of wild-type physiology and in 2.3.5; (iii) two glutamate dehydrogenases were operating towards production of L-glutamate; (iv) dihydrolipoamide S-succinyltransferase was operating towards production of NADH from NAD⁺ [266].

TFA performed with so configured stoichiometric model revealed that six reactions (acetaldehyde dehydrogenase acetylating, adenylate kinase, adenylyate kinase GTP, sodium proton antiporter, nucleoside diphosphate kinase ATP:GDP and phosphate reversible transport via symport periplasm) could operate in both directions while still satisfying the integrated data. To fix the directionalities of these six reactions, we performed an TFA optimization where we minimized the sum of the fluxes in the metabolic network under the constraint that at least 99% of the observed specific growth rate should be attained.

5.2.3 Construction of the large-scale kinetic models

The application of ORACLE [67, 69, 78-80, 181, 248] for calculation of control coefficients requires the following steps

Step 1. Definition of stoichiometry and thermodynamic constraints, along with incorporation available flux and concentration data. TFA analysis is performed in order to compute thermodynamically feasible flux profiles.

Step 2. Integration of experimentally measured concentrations into the models. For the estimation of concentrations for which we do not have data, a Monte Carlo sampling method is used. The space of metabolite concentrations is sampled consistently with the flux directionalities computed in Step 1. The displacement of enzymatic reactions from the thermodynamic equilibrium is also computed at this step.

Step 3. We integrate the kinetic properties of enzymes from the literature or databases (BRENDA and SABIO-RK). For those enzymes with missing information, we sample either enzyme states or the degree of saturation.

Step 4. We define the rate laws for the reaction kinetics and we formulate the kinetic model for the complete network. The information acquired in this step allows us to construct wide spectrum of kinetic models of metabolism including nonlinear models.

Step 5. We reject models that are either inconsistent with experimental measurements and expert knowledge or that are unstable in terms of local stability characteristics.

Step 6. We can perform different computational analysis (GSA analysis, dynamic simulations and MCA analysis). For each accepted model we compute a population of control coefficients.

Step 7. We analyze the population of CCs to define the statistics of the potential responses of population and to postulate hypotheses about system to genetic perturbations

5.3 Results

5.3.1 Kinetic study of wild type *P. putida* physiology

5.3.1a Predicting model responses to six single-gene knockouts

In the process of the construction of kinetic models, we removed the mass balances for the extracellular metabolites from the stoichiometry because we consider the concentrations of extracellular metabolites as parameters. The mass balances for water and the corresponding transport reactions were also removed. We then assigned a kinetic mechanism to each of the enzyme catalyzed reactions in the model, and we integrated experimental values for 21 Michaelis constants (K_m 's) that we found for the *Pseudomonas* genus in the Brenda database [269-272]. Next, we used ORACLE [67-69, 78-80, 181] to construct a population of 50'000 nonlinear kinetic models around the computed steady-state flux and concentration profile (Methods 5.2.3). The resulting structure of kinetic models consisted of 775 enzymatic reactions and 245 mass balances for metabolites distributed over cytosol and periplasm.

As a test for evaluating predictive capabilities of the constructed models, we computed the flux control coefficients [273, 274] of glucose uptake and specific growth rate with respect to six enzymes (glucose dehydrogenase (GLCDpp), hexokinase (HEX1), gluconokinase (GNK), EDA, EDD, and phosphogluconate 2-dehydrogenase (PGLCNDH)), and compared them with the experimentally measured responses of the glucose uptake and specific growth rate to single-gene knockouts of these six enzymes [111].

The computed control coefficients for the glucose uptake and specific growth rate were in a qualitative agreement with the data reported by del Castillo *et al.* [111] (Appendix Chapter 5 Table S1), i.e., a decrease in the enzyme activity of the six enzymes would lead to a decrease in both the glucose uptake and specific growth rate (Fig. 5.4a and 5.4b). Nevertheless, a closer inspection of the flux control coefficients of glucose uptake revealed that for four enzymes (GNK, EDD, EDA and PGLCNDH) the error bars were spread around zero values (Fig. 5.4a). This meant that there was a subpopulation of models with inconsistent predictions with some of the six knockouts. In fact, only 4 999 (~10%) out of 50 000 computed models were able to correctly predict responses to all 6 knockouts reported in del Castillo *et al.* [111] due to the large uncertainty in the assigned values of the kinetic parameters. This type of uncertainty remains one of the major difficulties that limit the predictive strength of kinetic models [180].

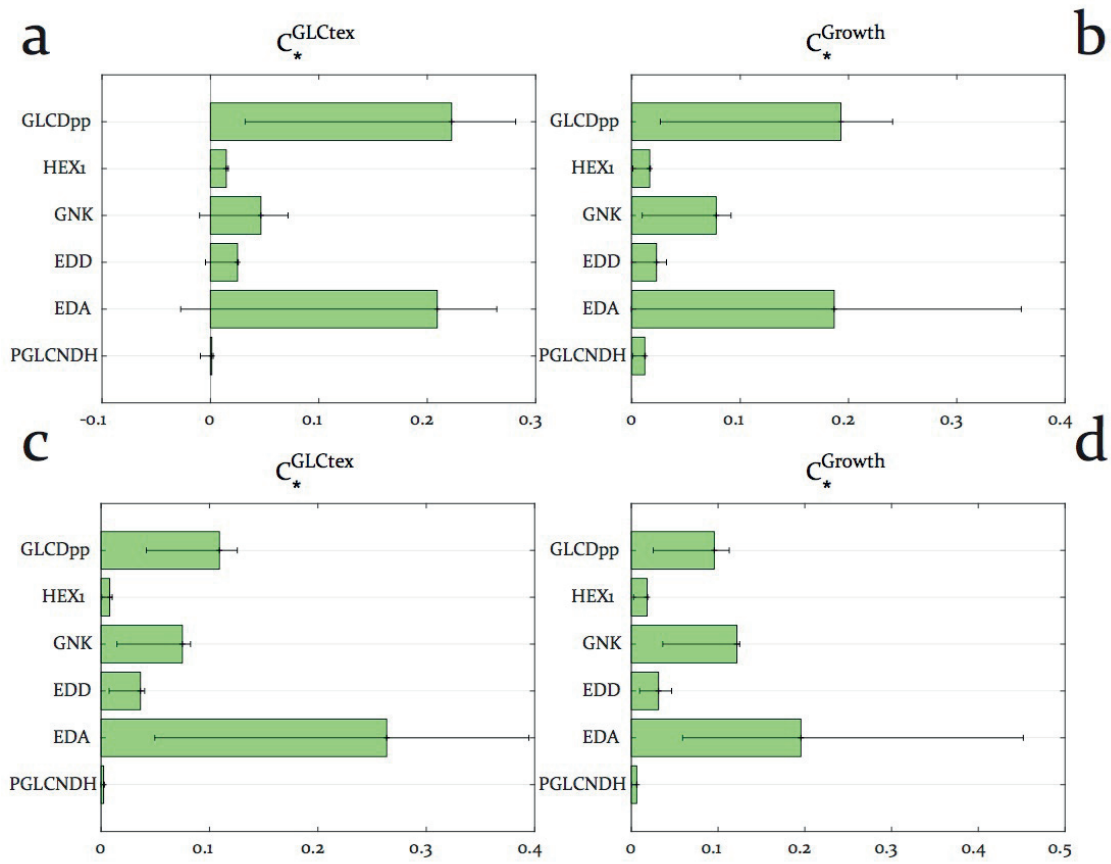


Figure 5.4 Distribution of the control coefficients of glucose uptake and specific cell growth in wild-type *P. putida* cells

The control coefficients of glucose uptake and specific growth were first computed using an unbiased sampling in ORACLE (a) and then further refined using the machine learning methodology iSCHRUNK (b). The green bars are the mean values of the control coefficients, whereas the errorbars correspond to the 25 and 75 percentiles of the distributions.

5.3.1b Refinement of model responses to six single-gene knockouts

To eliminate the inconsistencies with the experimental data observed for some of the predicted responses, we employed a machine learning method iSCHRUNK [275]. The method allowed us to identify the kinetic parameters and their ranges that ensure the consistency of model responses with the experimental observations. Out of 50 000 computed models, 4'999 were consistent with the results from del Castillo *et al.* [111], and we used the parameters of these models to train the machine learning algorithm. The method allowed us to identify seven kinetic parameters and their ranges that ensure

the consistency of model responses with the experimental observations, and interestingly, all parameters were related with the ED pathway (Table 5.1).

Table 5.1 Ranges of important parameters predicted by the iSCHRUNK algorithm
Abbreviations: 2DHGLCNtex, ketogluconate transport via diffusion extracellular to periplasm, GAD2ktp, gluconate 2 dehydrogenase periplasm, GLCDpp, glucose dehydrogenase ubiquinone 8 as acceptor periplasm, GLCNT2rpp, D-gluconate transport via proton symport reversible periplasm, GNK, gluconokinase, 2dhgln, 2-dehydro-D-gluconate, 6pgc, 6-phospho-D-gluconate, adp, ADP, glcn, D-gluconate, q8, ubiquinone-8.

Parameter	Range (mM)
$K_{m,2dhgln}^{2DHGLCNtex}$	$6.83 \cdot 10^{-5} - 2.34 \cdot 10^{-3}$
$K_{m,2dhgln}^{GAD2ktp}$	$6.83 \cdot 10^{-5} - 0.133$
$K_{m,q8}^{GLCDpp}$	$3.81 \cdot 10^{-3} - 0.899$
$K_{m,glcn}^{GLCDpp}$	$0.01 - 5.76$
$K_{m,glcn}^{GLCNT2rpp}$	$6.54 \cdot 10^{-4} - 9.49 \cdot 10^{-4}$
$K_{m,adp}^{GNK}$	$3.84 \cdot 10^{-2} - 20$
$K_{m,6pgc}^{GNK}$	$4.26 \cdot 10^{-5} - 8.37 \cdot 10^{-2}$

We generated a novel population of kinetic models with ORACLE with constrained ranges of these seven parameters as defined by iSCHRUNK, and we computed the distributions of corresponding control coefficients for the glucose uptake and specific growth rate. Out of 50'000 models, 29'979 (~60%) models correctly predicted the changes in the glucose uptake rate to six single-gene knockouts [111] (Fig. 5.4c), while 35'955 (~72%) models agreed with the experimental data for the specific growth rate (Fig. 5.4d). In total, 26'120 (~52%) models were consistent with both the experimental data for the glucose uptake and the specific growth rate.

We discovered with iSCHRUNK that operating regimes of only a few enzymes determine metabolic responses to multiple single-gene knockouts. This emphasizes the significance of accurately determining the kinetic parameters of such important enzymes in order to obtain model responses consistent with the experimental

observations. This also implies that we have to consider complex kinetic phenomena such as crowding when modeling kinetic properties of certain enzymes [276].

5.3.1c Assessment of estimated kinetic parameters

In the ORACLE framework, we employ the Monte Carlo sampling technique to compute the saturation states of enzymes. We then use these quantities to back-calculate the unknown values of Michaelis constants (K_m 's) [78-80]. To obtain an unbiased assessment of the accuracy of our estimates, we recomputed 50'000 models without imposing the experimentally available values of K_m 's from the BRENDA database [269-272]. Comparison of our estimates against available values of K_m 's from BRENDA showed that ORACLE could capture the ranges for 17 out of 21 K_m 's (Fig. 5.5). Considering that in the estimation process we did not use any kinetic parameters values and that the underlying system is undetermined, this result is remarkable because it indicates that ORACLE with integrated fluxomics and metabolomics data together with the physico-chemical laws is capable to provide consistent estimates for a large number of kinetic parameters. This means that ORACLE estimates can be used as hypothetical values for studies where the unknown kinetic parameters are required.

For the four remaining parameters such as Michaelis constant of L-Threonine in threonine aldolase or isocitrate in isocitrate lyase, ORACLEs underestimated experimental values up to one and half orders of magnitude (Fig. 4). The discrepancies between the estimated and measured values of these parameters can originate from different sources: (i) the K_m values from BRENDA were measured on several different species from the *Pseudomonas* genus, whereas our K_m values were estimated using a *P. putida* model and the experimental data were acquired on *P. putida* (fluxomics data) and *P. taiwanensis* (metabolomics data); and (ii) large uncertainty in available and partially available experimental data. In general, the more experimentally measured data are available for integration in the models by ORACLE, the better their predictive capability will be.

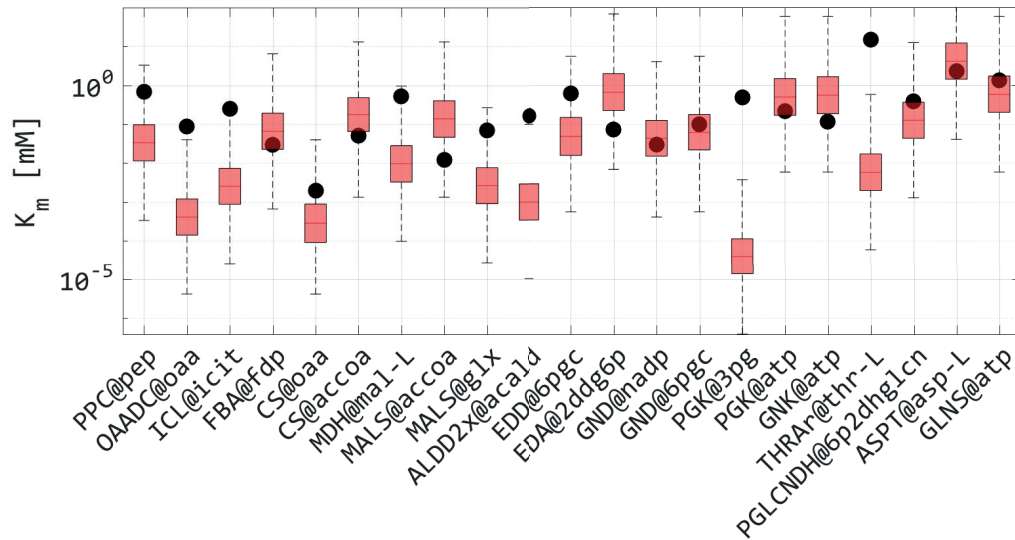


Figure 5.5 Ranges of K_m 's predicted by ORACLE

Red boxplots, distribution of K_m 's computed with ORACLE without imposing experimental values from BRENDA, black dots K_m 's reported in BRENDA. Whiskers represent minimal and maximal value predicted by ORACLE. Full name of reactions is provided in Appendix Chapter 5 Table S2

Similarly, we compared ranges of V_{max} 's computed with ORACLE against data reported on the literature. In the absence of V_{max} 's for *P. putida* in the literature, we compared them against the data for *E. coli* [277]. 15 of 19 experimentally reported values of V_{max} 's were within *in silico* predicted ranges, while 4 were laying outside of the predicted ranges (Fig 5.6). Although for majority of kinetic parameters reported in the literature, we could predict their ranges, the uncertainty of kinetic parameters remain one of the major issues in building kinetic models widely recognized in the literature [180, 275, 278].

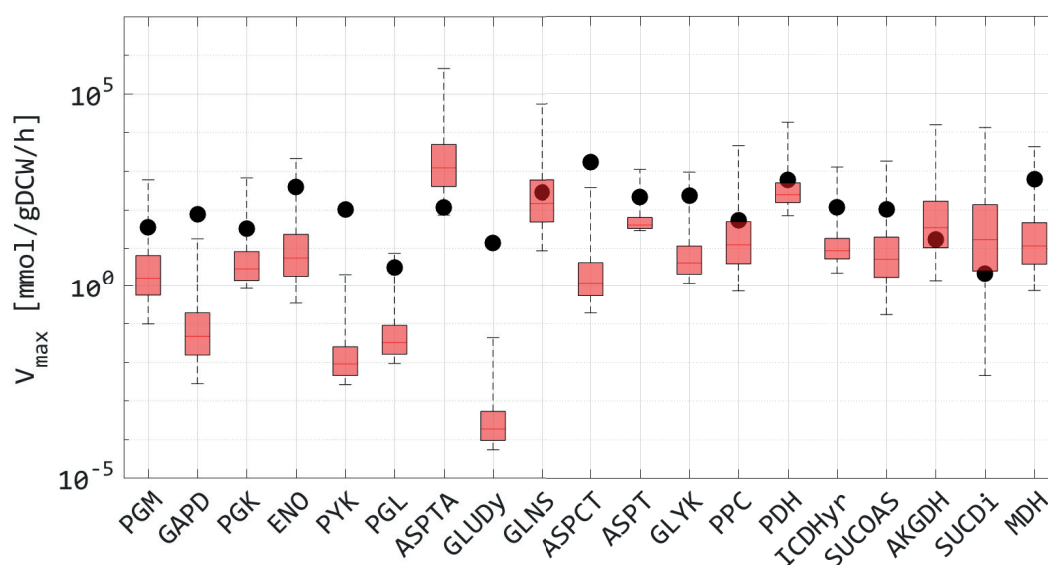


Figure 5.6 Ranges of V_{\max} 's predicted by ORACLE

Red boxplots, distribution of V_{\max} 's computed with ORACLE, black dots V_{\max} 's reported in Albe *et al.* [277] Whiskers represent minimal and maximal value predicted by ORACLE. Full name of reactions is provided in Appendix Chapter 5 Table S2

5.3.2 Kinetic study of increased ATP demand in *P. putida*

One of the main advantages of *P. putida* over *E. coli* is a buffering capacity to accommodate and counterbalance different environmental challenges without showing a distinct phenotype. Ebert *et al.* [24]. investigated the impact of increased ATP hydrolysis on the *P. putida* metabolism by titration of 2,4-dinitrophenol (DNP). They demonstrated that DNP concentrations below 300 mg/l did not impact the specific growth rate of *P. putida*. In comparison, *E. coli* showed a significant reduction in the specific growth rate already at the concentrations of 138 mg/l [279]. Above the concentration of 300 mg/liter, DNP caused a significant reduction of *P. putida*'s specific growth rate and increase of the glucose uptake (Figure 5.7a and 5.7b). At the concentration of 700 mg/liter of DNP, glucose uptake reached the maximum of ca. 11 mmol/gDCW/h. For larger values of DNP concentration, both glucose uptake and the specific growth rate declined.

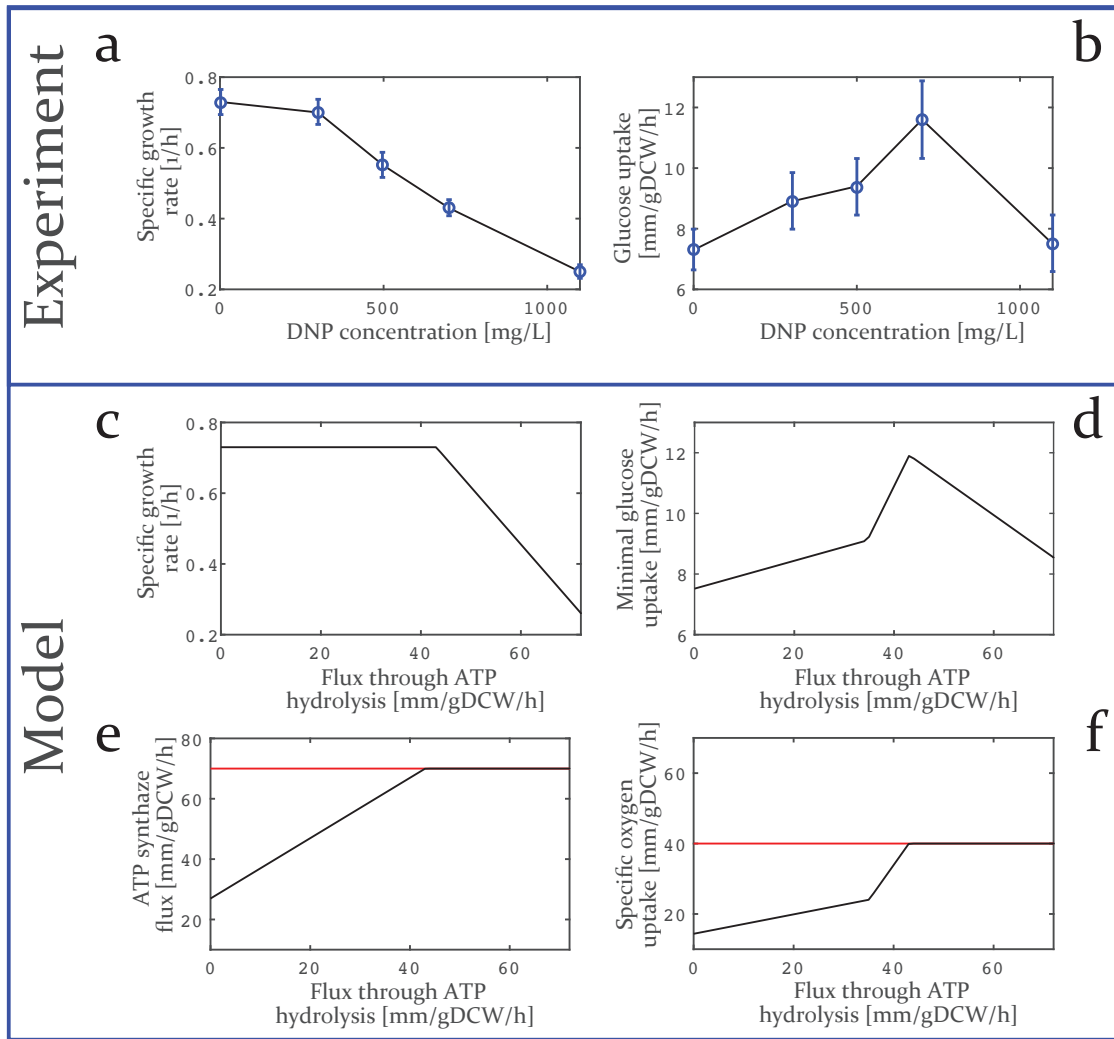


Figure 5.7 Fermentation profile of *P. putida* metabolism under increased ATP demand. Experimentally measured specific growth rate (a) and glucose uptake rate (b) of *P. putida* as the ATP demand induced by titration of 2,4 dinitrophenol (DNP) increases. The profiles of specific growth rate (c), glucose uptake rate (d), flux through ATP synthase (e) and oxygen uptake rate (f) computed by TFA

5.3.2a Modeling and TFA of increased ATP demand

We preconfigured the model for this study (Methods 5.2.2) and used it to simulate the impact of increased ATP demand on the *P. putida* metabolism by gradually increasing the minimally required flux through ATP hydrolysis in increments of 1 mmol/gDCW/h (Fig. 5.7). We set the upper bound of the specific growth rate to 0.73 1/h, as reported in Ebert *et al.* [24] for the DNP concentration of 0 mg/l. Based on the performed sensitivity

analysis of model responses to upper constraints on the oxygen uptake rate and ATP synthase (Methods 5.2.2), we set the upper bounds on the oxygen uptake rate and ATP synthase to 40 mmol/gDCW/h and 70 mmol/gDCW/h, respectively. The glucose uptake rate was left unconstrained.

In agreement with the experiments, the model predicted that the minimal glucose uptake of 7.51 mmol/gDCW/h is required to attain the specific growth rate of 0.73 1/h when the lower bound of the flux through ATP hydrolysis is set to 0 mmol/gDCW/h (Fig. 5.7c and 5.7d). Also consistent with the experiments, with the increase of the minimally required ATP hydrolysis flux, the required minimal glucose uptake was increasing (Fig. 5.7d) simultaneously with an increase of the ATP synthesis flux and minimal oxygen uptake (Fig. 5.7e and 5.7f), while the specific growth rate remained stable (Fig. 5.7c). For the ATP hydrolysis flux of 37 mmol/gDCW/h, the minimal glucose uptake was 9.56 mmol/gDCW/h and the slope of the minimal glucose and oxygen uptake became steeper (Fig. 5.7d and 5.7f). When the ATP hydrolysis flux reached 44 mmol/gDCW/h, the oxygen uptake rate and ATP synthase flux simultaneously attained their upper bounds (Fig. 5e and 5f). The corresponding minimal glucose uptake was 11.89 mmol/gDCW/h, which was consistent with Ebert *et al.* [24] (11.6 ± 1.2 mmol/gDCW/h). After this point, the required minimal glucose uptake started to decline (Fig. 5.7d) together with a decline in the specific growth rate (Fig. 5c). For the ATP hydrolysis flux of 73 mmol/gDCW/h, the model predicted the specific growth rate of 0.25 1/h and the minimal glucose uptake rate of 8.54 mmol/gDCW/h, which was slightly more than what was reported in the Ebert *et al.* [24] (7.5 ± 0.8 mmol/gDCW/h).

The thermodynamically-curated core stoichiometric model described well the qualitative behavior of *P. putida* in the stress condition of increased ATP demand. However, the model failed to capture a decrease of the specific growth rate for DNP concentrations in the range of 300-700 mg/l (Fig. 5.7c). A possible explanation for this discrepancy is that the decrease of specific growth rate in this region might be due to kinetic effects that cannot be captured by stoichiometric models. It is also important to observe that in Ebert *et al.* [24] the increased ATP demand was indirectly induced by titrating different levels of DNP, whereas we simulated that effect by increasing the ATP hydrolysis flux. Since *P. putida* does not necessarily respond to a linear increase in the

DNP levels by linearly increasing the ATP hydrolysis, the exact correspondence of the data points in the graphs obtained through experiments and computational simulation was not expected.

5.3.2b Improving the robustness of *P. putida* under stress conditions

We then undertook to devise a metabolic engineering strategy that will allow *P. putida* to maintain the specific growth rate for more severe stress conditions. To this end, we computed the steady-state metabolic flux and metabolite concentration vectors for the ATP hydrolysis flux of 44 mmol/gDCW/h. We then built a population of 50'000 kinetic models around the computed steady-state, and we computed the control coefficients for all fluxes and concentrations in the metabolic network.

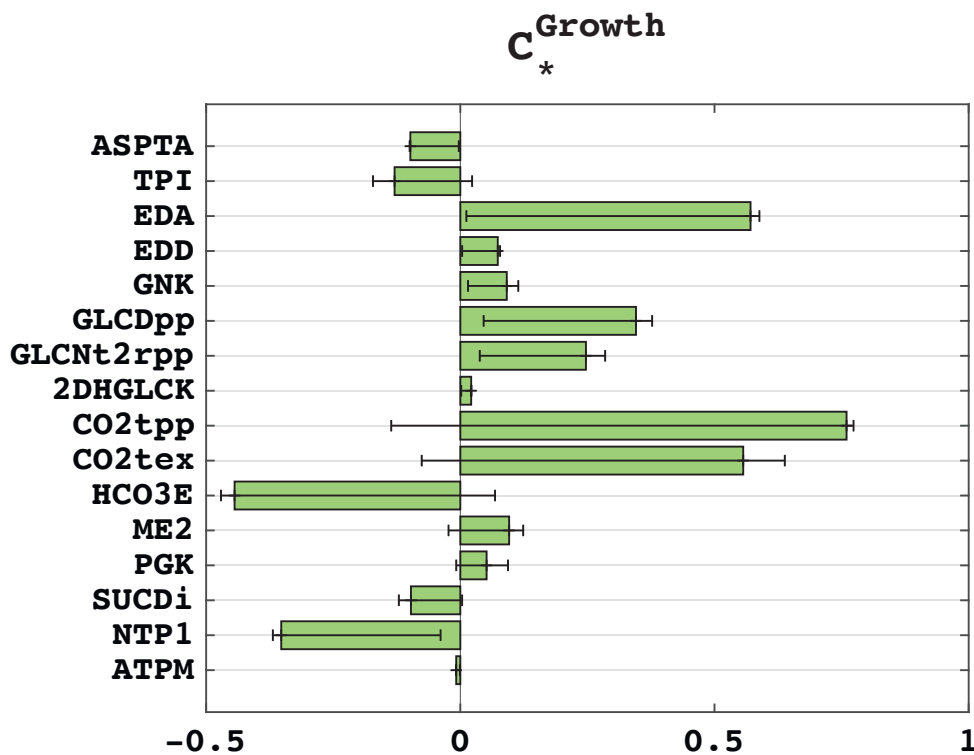


Figure 5.8 Control coefficients of the growth in the stress conditions

The green bars are the mean values of the control coefficients, whereas the errorbars correspond to the 25 and 75 percentiles of the distributions.

Analysis of the control coefficients for the specific growth rate revealed several strategies for maintaining high growth in the presence of stress agent 2,4-dinitrophenol which increases ATP demand (Fig. 5.8). The major positive control over the specific growth at this stress condition have the key enzymes from the Entner-Doudoroff pathway (EDA, EDD and GNK), e.g., the two-fold increase in activity of EDA would improve the specific growth by more than 50%. This control is tightly connected with the ability of ED pathway to generate additional NADPH, necessary to fuel proton-motive-force-driven efflux pumps, the major mechanism of solvent tolerance in *P. putida* [280] or to reduce stress through antioxidant systems that utilize NADPH [281].

Similarly, our analysis suggests that an increase in the activity of GLCDpp that catalyzes the conversion of glucose to periplasmic gluconate would increase the specific growth, i.e., the two-fold increase in GLCDpp activity would result in improved specific growth by ~40% (Fig. 7). Furthermore, reduced activity of aspartate transaminase (ASPTA) or succinate dehydrogenase (SUCDi) would also increase the specific growth.

It is rather intriguing that CO₂ transports (CO_{2tp} and CO_{2tex}) have a positive and bicarbonate equilibration reaction (HCO_{3E}) a negative control over the specific growth at this steady-state if we consider that all three reactions consume CO₂. Indeed, there was a strong coupling between the control coefficients $C_{CO_{2tp}}^{Growth}$ and $C_{HCO_{3E}}^{Growth}$ with the Pearson correlation coefficient of -0.77 (Appendix Chapter 5 Table S3). We also sampled metabolic fluxes from the space characterized by reactions that operate only in one direction determined by the analyzed steady state [53, 282]. Interestingly, the correlation analysis of the samples showed that there was no stoichiometric coupling of the fluxes through CO_{2tp} and HCO_{3E}, i.e., the Pearson correlation coefficient for these two fluxes was 0.06 (Appendix Chapter 5 Table 4)

We observed the similar effect of the ED enzymes on the glucose uptake (Appendix Chapter 5 Table S5).

5.4 Conclusions

We presented here the first large-scale kinetic model of *P. putida*. To our best knowledge, this is the largest kinetic model of any organism that could be found in the literature. For comparison, a recently published genome-scale kinetic model of *Escherichia coli* [219] consists of 457 reactions and 337 metabolites, while the model presented in this study has 775 reactions and 245 metabolites. We used the developed model in two studies of *P. putida* metabolism.

In the first study, we used the middle-size core reduced stoichiometric model of *P. putida* as a scaffold to generate a population of 50'000 kinetic models for predicting the behavior of six mutants of *P. putida* described in the literature [111]. A comparison between the models responses and the experimentally observed phenotypes revealed that only about 10% percent of our models could correctly predict experimentally reported behavior of these mutants due to the large uncertainty in the assigned values of the kinetic parameters, which remains one the major difficulties towards reliable predictions of the kinetic models [180]. To address this issue, we employed the iSCHRUNK method [275], and we have shown that constraining the values of only three kinetic parameters has a huge effect on the model response, i.e., the novel population of models with constrained three parameters related to glucose dehydrogenase was consistent with the experimental observations. This further implies urgency for accurate experimental determination of kinetic parameters, not only *in vitro*, where different kinetic effects such as crowding are disregarded [283-285], but also *in vivo*.

In the second study, we identified new strategies to maintain the robust behavior of *P. putida* under stress conditions. Our analysis suggests that a metabolic engineering strategy that targets enzymes from the ED pathway would have the positive control on growth and glucose uptake in the presence of stress agent 2,4-dinitrophenol which increases ATP hydrolysis. This control is tightly connected with the ability of this pathway to generate additional NADPH, necessary to fuel proton-motive-force-driven efflux pumps, the major mechanism of solvent tolerance in *P. putida* [280] or to reduce stress through antioxidant systems that utilize NADPH [281].

Conclusions and Perspectives

The production of second-generation biofuels by conversion of inexpensive feedstocks via microbial fermentation is a basis for future bio-sustainable economy. Compared with the other renewable sources of energy like solar electricity, liquid biofuels are immediately compatible with the existing infrastructure and with the higher energy density [286]. Fundamental to this problem is the selection of the host organism that can produce biofuels at high volumetric productivity and yield from cheap and abundantly available renewable energy sources [10]. Usually, when we talk about the heterologous expression of non-native pathways, the first choice for the host is *Escherichia coli* due to its ability for high growth on a large number of different carbon sources in both aerobic and anaerobic conditions. The use of *E. coli* as a microbial cell factory for biofuel production and its comparison against other frequently used hosts like *Saccharomyces cerevisiae* is reviewed elsewhere [287].

Recently, *P. putida* drew a lot of attention as a valid alternative to *E. coli*, in particular due to its ability to tolerate high concentrations of many organic solvents, thus representing a potentially superior host for biofuel production. In this thesis, we investigated the use of *Pseudomonas putida* as an alternative host for biofuel production, assessed its metabolic capacities and compared them to the ones of *E. coli*.

We undertook to assess *P. putida* prospects as a biofuel production host by analyzing its metabolism and more specifically its bioenergy requirements, redox metabolism and the supply of biofuel precursors. For these analyses we needed to develop large-scale kinetic models of this organism capable of capturing its system-wide properties. To this end, we started with the most recent genome-scale stoichiometric models of *P. putida*, iJN1411, and we used thermodynamic-based flux balance analysis to obtain the first

thermodynamically curated GEM of *P. putida*. The curated model will allow to a broader scientific community to perform more reliable computational analyses and predictions, e.g., without thermodynamics and the gap-filling we performed in this thesis, iJN1411 could not capture experimentally measured concentrations of ATP. With TFA, we could impose thermodynamic information for 62.3 % metabolites and 59.3 % reactions. In comparison, for *E. coli* GEM iJO1366 the coverage is higher, i.e., 84.1 % for metabolites and 81 % for reactions. The difference stems from a relatively low thermodynamic coverage in the peripheric metabolic subsystems of *P. putida* such as polyhydroxyalkanoates metabolism. We then used such curated GEM to derive three reduced models of iJN1411 using the redGEM and lumGEM algorithms. These models are faithful replicas of their GEM counterpart in terms of growth, gene essentiality and ranges of allowable fluxes and metabolite concentrations, and as such they can be used as a scaffold for development of large-scale kinetic models.

We took five precursors of Methyl Ethyl Ketone (MEK), one of the most prominent fuel candidates, as a test case for the assessment of *P. putida* metabolic capabilities and the comparison with *E. coli*. These five compounds can serve as platform chemicals for other biofuels. We discovered 18'662 thermodynamically feasible pathways from the central carbon metabolism subsystems of *E. coli* toward these five compounds, and for each novel reaction in these pathways we proposed the candidate enzymes that could catalyze them. This is a valuable resource for studies in industry and academia for the production of chemicals that can be derived from these five MEK precursors. Beside the delivered pathways, we extended the already existing computational pipeline for discovery, evaluation and analysis of biosynthetic pathways, BNICE.ch, with the new set of analysis tools that will allow to metabolic engineers to evaluate the pathways, their common routes and differences and assess the supply of the required metabolic precursors, and analyze the alternative routes for production of a target molecule, which would lead towards improved production and higher yields. This pipeline is applicable to any compound of interest and any host.

For comparison purposes, we evaluated the pathways from the metabolites that appear in the central carbon metabolism subsystems of both *P. putida* and *E. coli* toward the five MEK precursors. We have shown that same pathway can have different properties

in two host organisms. In some cases, we identified pathways with a higher yield in *P. putida* than in *E. coli*. To unlock *P. putida*'s full potential, one should explore the space of metabolites that belong uniquely to *P. putida* metabolism, and which can serve as starting compounds in the pathways design. There were 54 metabolites in *P. putida* carbon metabolism that do not appear in *E. coli*, and that could be used as starting metabolites in the pathway discovery of 5 MEK precursors.

It might be expected that upon experimental implementation of a pathway, the cell will produce a target molecule with the lower yield than what was predicted as the theoretical maximum. Identification of enzymes from the metabolic network whose overexpression, downregulation or deletion would improve the production, require a reliable kinetic model. As a final delivery of this thesis, we developed the first large-scale kinetic model of *P. putida* consistent with the metabolic responses of several mutants.

The results obtained in this thesis highlight the importance of the choice of constraints in FBA and TFA as they can influence our conclusions on reaction directionalities, and the reaction directionalities have a critical impact on network properties such as gene essentiality or yields [55]. Our results further suggest that particular caution should be exercised when using "off-the-shelf" models because some of them have *ad hoc* pre-assigned directionalities [53, 55, 56]. Additionally, this indicates that there is a need for revisiting assumptions on reaction directionalities in the current genome-scale reconstructions. This task can be performed by integrating thermodynamic constraints in metabolic networks and thus allowing for systematical assigning of reaction directionalities [55, 56]. However, for an accurate estimation of the reaction directionalities using thermodynamics, it is crucial to consider the contribution of metabolite activities to the Gibbs free energy of reactions instead of using only the standard values [53, 56]. Since metabolite activities are proportional to metabolite concentrations [288], this further emphasizes the importance of integrating metabolomics data.

The ultimate goal of kinetic modeling is to procure the models of such a scope and level of details that metabolic engineering and synthetic biology designs and hypotheses can reliably be tested *in silico* before implementing them into the host organisms. The

principal challenges in development of large- to genome-scale kinetic models remain the uncertainties and the complexity that increases with the size of the networks. There is a clear need for computational tools that will model the uncertainty and that will analyze and reduce the uncertainty propagation in the system. Recognizing this, recently a method has been proposed that makes use of machine learning techniques and the Monte Carlo sampling based methods for generating kinetic parameters such as ORACLE [275]. With this method, we have shown that constraining just three parameters related with the *P. putida*'s periplasmic enzyme glucose dehydrogenase is enough to capture the experimentally reported behavior of 6 mutants, while the values of kinetic parameters in most of the enzymes can vary in a broad range. This is an important finding as it will allow reducing the space of kinetic parameters significantly, and therefore it will enable more comprehensive analyses of the metabolic networks. During this work, we demonstrated the potential and usefulness of kinetic models in rational metabolic engineering strategies for (i) understanding the physiology of production hosts, (ii) optimizing production pathways, and (iii) improving the metabolic responses of organisms to environmental stresses.

Mathematical models and different computational techniques described in this thesis are the essential parts of a design-build-test-analyse strain engineering workflow [67] (Figure 6.1). This flow includes following steps: (i) design (pathway design, identification of a suitable host, yield estimation, identification of the precursors and alternative routes for the production of the target molecule, building a kinetic model), (ii) synthesis (pathway expression in a host, design of the novel enzymes, gene expression etc), (iii) testing (lab-scale production, titers) and (iv) analysis (strain characterization). The results from the analysis in step (iv) will be used to improve the model in step (i). Comparison of the performance with the original design will lead towards novel strategies for the production enhancement. This circle should be performed in an iterative manner, until the desired strain performance is reached.

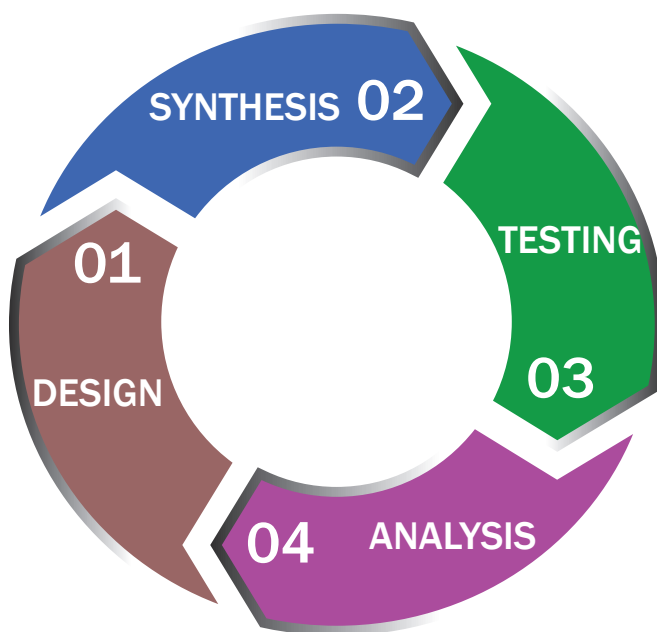


Figure 6.1 Design-based strain engineering workflow

Bibliography

1. Shafiee, S. and E. Topal, *When will fossil fuel reserves be diminished?* Energy Policy, 2009. **37**(1): p. 181-189.
2. <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/oil/oil-reserves.html>.
3. Caspeta, L., N.A.A. Buijs, and J. Nielsen, *The role of biofuels in the future energy supply*. Energy & Environmental Science, 2013. **6**(4): p. 1077-1082.
4. Stephanopoulos, G., *Challenges in engineering microbes for biofuels production*. Science, 2007. **315**(5813): p. 801-804.
5. <http://www.ethanolrfa.org/>.
6. Atsumi, S., T. Hanai, and J.C. Liao, *Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels*. Nature, 2008. **451**(7174): p. 86-9.
7. Schirmer, A., et al., *Microbial biosynthesis of alkanes*. Science (New York, N.Y.), 2010. **329**(5991): p. 559-62.
8. Steen, E.J., et al., *Microbial production of fatty-acid-derived fuels and chemicals from plant biomass*. Nature, 2010. **463**(7280): p. 559-62.
9. Ghiaci, P., J. Norbeck, and C. Larsson, *2-Butanol and Butanone Production in Saccharomyces cerevisiae through Combination of a B12 Dependent Dehydratase and a Secondary Alcohol Dehydrogenase Using a TEV-Based Expression System*. PLOS ONE, 2014. **9**(7): p. e102774.
10. Clomburg, J.M. and R. Gonzalez, *Biofuel production in Escherichia coli: the role of metabolic engineering and synthetic biology*. Applied Microbiology and Biotechnology, 2010. **86**(2): p. 419-434.
11. Cho, A., et al., *Prediction of novel synthetic pathways for the production of desired chemicals*. BMC Systems Biology, 2010. **4**.

Bibliography

12. Hadadi, N. and V. Hatzimanikatis, *Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways*. Current Opinion in Chemical Biology, 2015. **28**: p. 99-104.
13. Nielsen, L.K., *METABOLIC ENGINEERING From retrofitting to green field*. Nature Chemical Biology, 2011. **7**(7): p. 407-408.
14. Alper, H. and G. Stephanopoulos, *Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential?* Nature Reviews Microbiology, 2009. **7**(10): p. 715-723.
15. Trinh, C.T., et al., *Redesigning Escherichia coli Metabolism for Anaerobic Production of Isobutanol*. Applied and Environmental Microbiology, 2011. **77**(14): p. 4894-4904.
16. Atsumi, S., T. Hanai, and J.C. Liao, *Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels*. Nature, 2008. **451**(7174): p. 86-U13.
17. Steen, E.J., et al., *Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol*. Microbial Cell Factories, 2008. **7**.
18. Jarboe, L.R., et al., *Optimization of enzyme parameters for fermentative production of biorenewable fuels and chemicals*. Computational and Structural Biotechnology Journal, 2012. **3**: p. e201210005.
19. Nikel, P., *A Brief Guide to Pseudomonas putida as a microbial cell factory*. BioEssays, 2012.
20. Rojo, F., *Carbon catabolite repression in Pseudomonas: optimizing metabolic versatility and interactions with the environment*. Fems Microbiology Reviews, 2010. **34**(5): p. 658-684.
21. Isken, S., et al., *Effect of organic solvents on the yield of solvent-tolerant Pseudomonas putida S12*. Applied and Environmental Microbiology, 1999. **65**(6): p. 2631-2635.
22. Inoue, A., M. Yamamoto, and K. Horikoshi, *Pseudomonas-Putida Which Can Grow in the Presence of Toluene*. Applied and Environmental Microbiology, 1991. **57**(5): p. 1560-1562.
23. Ruhl, J., A. Schmid, and L.M. Blank, *Selected Pseudomonas putida Strains Able To Grow in the Presence of High Butanol Concentrations*. Applied and Environmental Microbiology, 2009. **75**(13): p. 4653-4656.

24. Ebert, B.E., et al., *Response of Pseudomonas putida KT2440 to Increased NADH and ATP Demand*. Applied and Environmental Microbiology, 2011. **77**(18): p. 6597-6605.
25. Nikel, P., *Systems and Synthetic Biology Approaches for Metabolic Engineering of Pseudomonas putida*, in *Microbial Models: From Environmental to Industrial Sustainability*. 2016.
26. Thiele, I., et al., *A community-driven global reconstruction of human metabolism*. Nature Biotechnology, 2013. **31**(5): p. 419-+.
27. Osterlund, T., et al., *Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling*. BMC Systems Biology, 2013. **7**.
28. Henry, C.S., et al., *High-throughput generation, optimization and analysis of genome-scale metabolic models*. Nature Biotechnology, 2010. **28**(9): p. 977-U22.
29. Orth, J.D., et al., *A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011*. Molecular Systems Biology, 2011. **7**.
30. Asadollahi, M.A., et al., *Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering*. Metab Eng, 2009.
31. Borodina, I., et al., *Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in Saccharomyces cerevisiae via beta-alanine*. Metabolic Engineering, 2015. **27**: p. 57-64.
32. Dash, S., et al., *Capturing the response of Clostridium acetobutylicum to chemical stressors using a regulated genome-scale metabolic model*. Biotechnology for Biofuels, 2014. **7**.
33. Noguchi, R., et al., *The selective control of glycolysis, gluconeogenesis and glycogenesis by temporal insulin patterns*. Molecular systems biology, 2013. **9**(1): p. 664.
34. Stanford, N.J., et al., *Systematic construction of kinetic models from genome-scale metabolic networks*. PloS ONE, 2013.
35. Di Maggio, J., et al., *Parameter estimation in kinetic models for large scale biotechnological systems with advanced mathematical programming techniques*. Biochemical Engineering Journal, 2014. **83**: p. 104-115.
36. Levering, J., et al., *Glycolytic oscillations in a model of a lactic acid bacterium metabolism*. Biophysical chemistry, 2013. **172**: p. 53-60.

Bibliography

37. du Preez, F.B., et al., *From steady-state to synchronized yeast glycolytic oscillations I: model construction*. Febs Journal, 2012. **279**(16): p. 2810-2822.
38. Nikel, P.I. and V. de Lorenzo, *Robustness of Pseudomonas putida KT2440 as a host for ethanol biosynthesis*. New Biotechnology, 2014. **31**(6): p. 562-571.
39. Schellenberger, J., et al., *Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0*. Nature Protocols, 2011. **6**(9): p. 1290-1307.
40. Klipp, E., et al., *Integrative model of the response of yeast to osmotic shock (vol 23, pg 975, 2005)*. Nature Biotechnology, 2006. **24**(10): p. 1293-1293.
41. Raskevicius, V., et al., *Genome scale metabolic models as tools for drug design and personalized medicine*. Plos One, 2018. **13**(1).
42. Tokuyama, K., et al., *Increased 3-hydroxypropionic acid production from glycerol, by modification of central metabolism in Escherichia coli*. Microbial Cell Factories, 2014. **13**.
43. Lin, Z.Q., et al., *Metabolic engineering of Escherichia coli for poly(3-hydroxybutyrate) production via threonine bypass*. Microbial Cell Factories, 2015. **14**.
44. Liu, J., et al., *Model-driven intracellular redox status modulation for increasing isobutanol production in Escherichia coli*. Biotechnology for Biofuels, 2015. **8**.
45. Gold, N.D., et al., *Metabolic engineering of a tyrosine-overproducing yeast platform using targeted metabolomics*. Microbial Cell Factories, 2015. **14**.
46. Fondi, M., et al., *Genome-scale metabolic reconstruction and constraint-based modelling of the Antarctic bacterium Pseudoalteromonas haloplanktis TAC125*. Environmental Microbiology, 2015. **17**(3): p. 751-766.
47. Segre, D., D. Vitkup, and G.M. Church, *Analysis of optimality in natural and perturbed metabolic networks*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(23): p. 15112-15117.
48. O'Brien, E.J., J.M. Monk, and B.O. Palsson, *Using Genome-scale Models to Predict Biological Capabilities*. Cell, 2015. **161**(5): p. 971-987.
49. Orth, J.D., I. Thiele, and B.O. Palsson, *What is flux balance analysis?* Nature Biotechnology, 2010. **28**(3): p. 245-248.
50. Kauffman, K.J., P. Prakash, and J.S. Edwards, *Advances in flux balance analysis*. Current Opinion in Biotechnology, 2003. **14**(5): p. 491-496.

51. Henry, C.S., L.J. Broadbelt, and V. Hatzimanikatis, *Thermodynamics-based metabolic flux analysis*. Biophysical Journal, 2007. **92**(5): p. 1792-1805.
52. Henry, C.S., et al., *Genome-scale thermodynamic analysis of Escherichia coli metabolism*. Biophysical Journal, 2006. **90**(4): p. 1453-1461.
53. Soh, K.S. and V. Hatzimanikatis, *Constraining the flux space using thermodynamics and integration of metabolomics data*. Methods in Molecular Biology, 2014. **1191**: p. 49-63.
54. Soh, K.C., L. Miskovic, and V. Hatzimanikatis, *From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks*. Fems Yeast Research, 2012. **12**(2): p. 129-143.
55. Ataman, M. and V. Hatzimanikatis, *Heading in the right direction: thermodynamics-based network analysis and pathway engineering*. Curr Opin Biotechnol, 2015. **36**: p. 176-182.
56. Soh, K.C. and V. Hatzimanikatis, *Network thermodynamics in the post-genomic era*. Curr Opin Microbiol, 2010. **13**(3): p. 350-7.
57. Salvy, P., et al., *pyTFA and matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis*. Bioinformatics, 2018: p. bty499-bty499.
58. Jankowski, M.D., et al., *Group contribution method for thermodynamic analysis of complex metabolic networks*. Biophysical Journal, 2008. **95**(3): p. 1487-1499.
59. Mavrovouniotis, M.L., *Estimation of Standard Gibbs Energy Changes of Biotransformations*. Journal of Biological Chemistry, 1991. **266**(22): p. 14440-14445.
60. Covert, M.W. and B.O. Palsson, *Transcriptional regulation in constraints-based metabolic models of Escherichia coli*. Journal of Biological Chemistry, 2002. **277**(31): p. 28058-28064.
61. Shlomi, T., et al., *A genome-scale computational study of the interplay between transcriptional regulation and metabolism*. Molecular Systems Biology, 2007. **3**.
62. Chandrasekaran, S. and N.D. Price, *Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis*. Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(41): p. 17845-17850.
63. Covert, M.W., et al., *Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli*. Bioinformatics, 2008. **24**(18): p. 2044-2050.

Bibliography

64. Mahadevan, R., J.S. Edwards, and F.J. Doyle, *Dynamic flux balance analysis of diauxic growth in Escherichia coli*. Biophysical Journal, 2002. **83**(3): p. 1331-1340.
65. Mori, M., et al., *Constrained Allocation Flux Balance Analysis*. Plos Computational Biology, 2016. **12**(6).
66. Waldherr, S., D.A. Oyarzun, and A. Bockmayr, *Dynamic optimization of metabolic networks coupled with gene expression*. Journal of Theoretical Biology, 2015. **365**: p. 469-485.
67. Miskovic, L., et al., *A design-build-test cycle using modeling and experiments reveals interdependencies between upper glycolysis and xylose uptake in recombinant S. cerevisiae and improves predictive capabilities of large-scale kinetic models*. Biotechnology for Biofuels, 2017. **10**(1): p. 166.
68. Andreozzi, S., et al., *Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant E. coli using large-scale kinetic models*. Metabolic Engineering, 2016. **35**: p. 148-159.
69. Miskovic, L. and V. Hatzimanikatis, *Production of biofuels and biochemicals: in need of an ORACLE*. Trends in Biotechnology, 2010. **28**(8): p. 391-397.
70. Rutkis, R., et al., *Kinetic modelling of the Zymomonas mobilis Entner-Doudoroff pathway: insights into control and functionality*. Microbiology-Sgm, 2013. **159**: p. 2674-2689.
71. Sen, P., H.J. Vial, and O. Radulescu, *Kinetic modelling of phospholipid synthesis in Plasmodium knowlesi unravels crucial steps and relative importance of multiple pathways*. BMC Systems Biology, 2013. **7**.
72. Weaver, L.J., et al., *A Kinetic-Based Approach to Understanding Heterologous Mevalonate Pathway Function in E. Coli*. Biotechnology and Bioengineering, 2015. **112**(1): p. 111-119.
73. Kerkhoven, E.J., P.J. Lahtvee, and J. Nielsen, *Applications of computational modeling in metabolic engineering of yeast*. FEMS Yeast Research, 2015. **15**(1).
74. Penkler, G., et al., *Construction and validation of a detailed kinetic model of glycolysis in Plasmodium falciparum*. FEBS Journal, 2015. **282**(8): p. 1481-1511.
75. Snoep, J., G. Penkler, and M. Rautenbach, *Construction and validation of a detailed kinetic model for glycolysis in the asexual stage of Plasmodium falciparum; application for drug target identification*. New Biotechnology, 2009. **25**: p. S354-S354.

76. Ataman, M., et al., *redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models*. PLOS Computational Biology, 2017. **13**(7): p. e1005444.
77. Erdrich, P., R. Steuer, and S. Klamt, *An algorithm for the reduction of genome-scale metabolic network models to meaningful core models*. BMC Systems Biology, 2015. **9**.
78. Wang, L., I. Birol, and V. Hatzimanikatis, *Metabolic Control Analysis under Uncertainty: Framework Development and Case Studies*. Biophysical Journal, 2004. **87**(6): p. 3750-3763.
79. Wang, L. and V. Hatzimanikatis, *Metabolic engineering under uncertainty—II: Analysis of yeast metabolism*. Metabolic Engineering, 2006. **8**(2): p. 142-159.
80. Mišković, L. and V. Hatzimanikatis, *Modeling of uncertainties in biochemical reactions*. Biotechnology and Bioengineering, 2011. **108**(2): p. 413-423.
81. Almquist, J., et al., *Kinetic models in industrial biotechnology - Improving cell factory performance*. Metabolic Engineering, 2014. **24**: p. 38-60.
82. Rühl, J., A. Schmid, and L.M. Blank, *Selected Pseudomonas putida Strains Able To Grow in the Presence of High Butanol Concentrations*. Applied and Environmental Microbiology, 2009. **75**(13): p. 4653-4656.
83. Nogales, J., B.Ø. Palsson, and I. Thiele, *A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory*. BMC Systems Biology, 2008. **2**(1): p. 79.
84. Puchałka, J., et al., *Genome-Scale Reconstruction and Analysis of the Pseudomonas putida KT2440 Metabolic Network Facilitates Applications in Biotechnology*. PLOS Computational Biology, 2008. **4**(10): p. e1000210.
85. Sohn, S.B., et al., *In silico genome-scale metabolic analysis of Pseudomonas putida KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival*. Biotechnology Journal, 2010. **5**(7): p. 739-750.
86. Oberhardt, M.A., et al., *Reconciliation of Genome-Scale Metabolic Reconstructions for Comparative Systems Analysis*. PLOS Computational Biology, 2011. **7**(3): p. e1001116.
87. Poblete-Castro, I., et al., *In-silico-driven metabolic engineering of Pseudomonas putida for enhanced production of poly-hydroxyalkanoates*. Metabolic Engineering, 2013. **15**: p. 113-123.

Bibliography

88. van Duuren, J.B., et al., *Reconciling in vivo and in silico key biological parameters of Pseudomonas putida KT2440 during growth on glucose under carbon-limited condition*. BMC Biotechnology, 2013. **13**(1): p. 93.
89. Molina-Henares, M.A., et al., *Identification of conditionally essential genes for growth of Pseudomonas putida KT2440 on minimal medium through the screening of a genome-wide mutant library*. Environmental Microbiology, 2010. **12**(6): p. 1468-1485.
90. Yuan, Q., et al., *Pathway-Consensus Approach to Metabolic Network Reconstruction for Pseudomonas putida KT2440 by Systematic Comparison of Published Models*. PLOS ONE, 2017. **12**(1): p. e0169437.
91. Nogales, J., et al., *Expanding The Computable Reactome In Pseudomonas putida Reveals Metabolic Cycles Providing Robustness*. bioRxiv, 2017.
92. Tokic, M., et al., *Discovery and Evaluation of Biosynthetic Pathways for the Production of Five Methyl Ethyl Ketone Precursors*. bioRxiv, 2018.
93. Asplund-Samuelsson, J., M. Janasch, and E.P. Hudson, *Thermodynamic analysis of computed pathways integrated into the metabolic networks of E. coli and Synechocystis reveals contrasting expansion potential*. Metabolic Engineering, 2018. **45**: p. 223-236.
94. Bar-Even, A., et al., *Design and analysis of synthetic carbon fixation pathways*. Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(19): p. 8889-8894.
95. De Martino, D., et al., *Counting and Correcting Thermodynamically Infeasible Flux Cycles in Genome-Scale Metabolic Networks*. Metabolites, 2013. **3**(4): p. 946-966.
96. Desouki, A.A., et al., *CycleFreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions*. Bioinformatics, 2015. **31**(13): p. 2159-2165.
97. Schellenberger, J., N.E. Lewis, and B.O. Palsson, *Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models*. Biophysical Journal, 2011. **100**(3): p. 544-553.
98. Birkenmeier, M., M. Mack, and T. Roder, *A coupled thermodynamic and metabolic control analysis methodology and its evaluation on glycerol biosynthesis in Saccharomyces cerevisiae (vol 37, pg 307, 2015)*. Biotechnology Letters, 2015. **37**(2): p. 317-326.
99. Feist, A.M., et al., *A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information*. Molecular Systems Biology, 2007. **3**.

100. Hamilton, J.J., V. Dwivedi, and J.L. Reed, *Quantitative Assessment of Thermodynamic Constraints on the Solution Space of Genome-Scale Metabolic Models*. Biophysical Journal, 2013. **105**(2): p. 512-522.
101. Martiez, V.S., L.E. Quek, and L.K. Nielsen, *Network Thermodynamic Curation of Human and Yeast Genome-Scale Metabolic Models*. Biophysical Journal, 2014. **107**(2): p. 493-503.
102. Chiappino-Pepe, A., et al., *Bioenergetics-based modeling of Plasmodium falciparum metabolism reveals its essential genes, nutritional requirements, and thermodynamic bottlenecks*. Plos Computational Biology, 2017. **13**(3).
103. Trinh, C.T., A. Wlaschin, and F. Sreenc, *Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism*. Applied Microbiology and Biotechnology, 2009. **81**(5): p. 813-826.
104. Acuna, V., et al., *A note on the complexity of finding and enumerating elementary modes*. Biosystems, 2010. **99**(3): p. 210-214.
105. David, L. and A. Bockmayr, *Computing Elementary Flux Modes Involving a Set of Target Reactions*. Ieee-Acm Transactions on Computational Biology and Bioinformatics, 2014. **11**(6): p. 1099-1107.
106. Flynn, C.M., et al., *Construction and elementary mode analysis of a metabolic model for Shewanella oneidensis MR-1*. Biosystems, 2012. **107**(2): p. 120-128.
107. Jensen, J.L.W.V., *Sur les fonctions convexes et les inegalites entre les valeurs moyennes*. Acta Math., 1906. **30**: p. 175-193.
108. <http://bionumbers.hms.harvard.edu//bionumber.aspx?id=100010&ver=11>.
109. Chavarría, M., et al., *The Entner–Doudoroff pathway empowers Pseudomonas putida KT2440 with a high tolerance to oxidative stress*. Environmental Microbiology, 2013. **15**(6): p. 1772-1785.
110. Ataman, M. and V. Hatzimanikatis, *lumpGEM: Systematic generation of subnetworks and elementally balanced lumped reactions for the biosynthesis of target metabolites*. PLOS Computational Biology, 2017. **13**(7): p. e1005513.
111. del Castillo, T., et al., *Convergent Peripheral Pathways Catalyze Initial Glucose Catabolism in Pseudomonas putida: Genomic and Flux Analysis*. Journal of Bacteriology, 2007. **189**(14): p. 5142-5152.
112. Wordofa, G.G., et al., *Quantifying the metabolome of Pseudomonas taiwanensis VLB120: Evaluation of hot and cold combined quenching/extraction approaches*. Analytical Chemistry, 2017.

Bibliography

113. Blank, L.M., et al., *Redox Biocatalysis and Metabolism: Molecular Mechanisms and Metabolic Network Analysis*. Antioxidants & Redox Signaling, 2010. **13**(3): p. 349-394.
114. Neuwald, A.F., et al., *Cysq, a Gene Needed for Cysteine Synthesis in Escherichia-Coli K-12 Only during Aerobic Growth*. Journal of Bacteriology, 1992. **174**(2): p. 415-425.
115. Asadollahi, M.A., et al., *Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering*. Metabolic Engineering, 2009. **11**(6): p. 328-334.
116. Nikel, P.I., et al., *Pseudomonas putida KT2440 Metabolizes Glucose Through a Cycle Formed by Enzymes of the Entner-Doudoroff, Embden-Meyerhof-Parnas, and Pentose Phosphate Pathways*. Journal of Biological Chemistry, 2015: p. jbc.M115.687749-jbc.M115.687749.
117. Vicente, M. and J.L. Canovas, *Glucolysis in Pseudomonas-Putida - Physiological Role of Alternative Routes from Analysis of Defective Mutants*. Journal of Bacteriology, 1973. **116**(2): p. 908-914.
118. Beard, D.A., S.C. Liang, and H. Qian, *Energy balance for analysis of complex metabolic networks*. Biophysical Journal, 2002. **83**(1): p. 79-86.
119. Kummel, A., S. Panke, and M. Heinemann, *Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data*. Molecular Systems Biology, 2006. **2**.
120. Lee, S.K., et al., *Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels*. Current Opinion in Biotechnology, 2008. **19**(6): p. 556-563.
121. Yim, H., et al., *Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol*. Nature Chemical Biology, 2011. **7**(7): p. 445-452.
122. Jones, D.T. and D.R. Woods, *Acetone-Butanol Fermentation Revisited*. Microbiological Reviews, 1986. **50**(4): p. 484-524.
123. Atsumi, S., et al., *Metabolic engineering of Escherichia coli for 1-butanol production*. Metabolic Engineering, 2008. **10**(6): p. 305-311.
124. Hanai, T., S. Atsumi, and J.C. Liao, *Engineered synthetic pathway for isopropanol production in Escherichia coli*. Applied and Environmental Microbiology, 2007. **73**(24): p. 7814-7818.

125. Zhang, K.C., et al., *Expanding metabolism for biosynthesis of nonnatural alcohols*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(52): p. 20653-20658.
126. Connor, M.R., A.F. Cann, and J.C. Liao, *3-Methyl-1-butanol production in Escherichia coli: random mutagenesis and two-phase fermentation*. Applied Microbiology and Biotechnology, 2010. **86**(4): p. 1155-1164.
127. Nielsen, L.K., *From retrofitting to green field*. Nature Chemical Biology, 2011. **7**(7): p. 407-408.
128. Bogorad, I.W., T.S. Lin, and J.C. Liao, *Synthetic non-oxidative glycolysis enables complete carbon conservation*. Nature, 2013. **502**(7473): p. 693-+.
129. Henry, C.S., L.J. Broadbelt, and V. Hatzimanikatis, *Discovery and Analysis of Novel Metabolic Pathways for the Biosynthesis of Industrial Chemicals: 3-Hydroxypropionate*. Biotechnology and Bioengineering, 2010. **106**(3): p. 462-473.
130. Moriya, Y., et al., *PathPred: an enzyme-catalyzed metabolic pathway prediction server*. Nucleic Acids Research, 2010. **38**(suppl_2): p. W138-W143.
131. Hou, B.K., L.B.M. Ellis, and L.P. Wackett, *Encoding microbial metabolic logic: predicting biodegradation*. Journal of Industrial Microbiology and Biotechnology, 2004. **31**(6): p. 261-272.
132. Ellis, L.B.M., et al., *The University of Minnesota pathway prediction system: predicting metabolic logic*. Nucleic Acids Research, 2008. **36**(Web Server issue): p. W427-W432.
133. Campodonico, M.A., et al., *Generation of an atlas for commodity chemical production in Escherichia coli and a novel pathway prediction algorithm, GEM-Path*. Metabolic Engineering, 2014. **25**: p. 140-158.
134. Rodrigo, G., et al., *DESHARKY: automatic design of metabolic pathways for optimal cell growth*. Bioinformatics, 2008. **24**(21): p. 2554-2556.
135. Carbonell, P., et al., *XTMS: pathway design in an eXTended metabolic space*. Nucleic Acids Research, 2014. **42**(W1): p. W389-W394.
136. Heath, A.P., G.N. Bennett, and L.E. Kavraki, *Finding metabolic pathways using atom tracking*. Bioinformatics, 2010. **26**(12): p. 1548-1555.
137. Dale, J.M., L. Popescu, and P.D. Karp, *Machine learning methods for metabolic pathway prediction*. BMC Bioinformatics, 2010. **11**.

Bibliography

138. Prather, K.L.J. and C.H. Martin, *De novo biosynthetic pathways: rational design of microbial chemical factories*. Current Opinion in Biotechnology, 2008. **19**(5): p. 468-474.
139. Hatzimanikatis, V., et al., *Exploring the diversity of complex metabolic networks*. Bioinformatics, 2005. **21**(8): p. 1603-1609.
140. Hatzimanikatis, V., et al., *Metabolic networks: enzyme function and metabolite structure*. Current Opinion in Structural Biology, 2004. **14**(3): p. 300-306.
141. Hadadi, N., et al., *A computational framework for integration of lipidomics data into metabolic pathways*. Metabolic Engineering, 2014. **23**: p. 1-8.
142. Hadadi, N., et al., *ATLAS of Biochemistry: A Repository of All Possible Biochemical Reactions for Synthetic Biology and Metabolic Engineering Studies*. ACS Synthetic Biology, 2016: p. 1155–1166.
143. Soh, K.C. and V. Hatzimanikatis, *Dreams of Metabolism*. Trends in Biotechnology, 2010. **28**(10): p. 501-508.
144. Brunk, E., et al., *Integrating computational methods to retrofit enzymes to synthetic pathways*. Biotechnology and Bioengineering, 2012. **109**(2): p. 572-582.
145. Hoell, D., et al., *2-Butanone*, in *Ullmann's Encyclopedia of Industrial Chemistry*. 2009, Wiley-VCH Verlag GmbH & Co. KGaA.
146. Hoppe, F., et al., *Tailor-Made Fuels from Biomass: Potentials of 2-butanone and 2-methylfuran in direct injection spark ignition engines*. Fuel, 2016. **167**: p. 106-117.
147. Yoneda, H., D.J. Tantillo, and S. Atsumi, *Biological Production of 2-Butanone in Escherichia coli*. ChemSusChem, 2014. **7**(1): p. 92-95.
148. Srirangan, K., et al., *Engineering Escherichia coli for Microbial Production of Butanone*. Applied and Environmental Microbiology, 2016. **82**(9): p. 2574-2584.
149. Multer, A., et al., *Production of Methyl Ethyl Ketone from Biomass Using a Hybrid Biochemical/Catalytic Approach*. Industrial & Engineering Chemistry Research, 2013. **52**(1): p. 56-60.
150. Drabo, P., et al., *Anionic Extraction for Efficient Recovery of Biobased 2,3-Butanediol—A Platform for Bulk and Fine Chemicals*. ChemSusChem, 2017. **10**(16): p. 3252-3259.
151. Kanehisa, M., et al., *KEGG as a reference resource for gene and protein annotation*. Nucleic Acids Research, 2016. **44**(D1): p. D457-D462.

152. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Nucleic Acids Research, 2000. **28**(1): p. 27-30.
153. Engel, T., *The structural- and bioassay database PubChem*. Nachrichten Aus Der Chemie, 2007. **55**(5): p. 521-524.
154. Kim, S., et al., *PubChem Substance and Compound databases*. Nucleic Acids Research, 2016. **44**(Database issue): p. D1202-D1213.
155. Krumpfer, J.W., et al., *Poly(Methyl Vinyl Ketone) as a Potential Carbon Fiber Precursor*. Chemistry of Materials, 2017. **29**(2): p. 780-788.
156. Siegel, H. and M. Eggersdorfer, *Ketones*, in *Ullmann's Encyclopedia of Industrial Chemistry*. 2000, Wiley-VCH Verlag GmbH & Co. KGaA.
157. Eller, K., et al., *Amines, Aliphatic*, in *Ullmann's Encyclopedia of Industrial Chemistry*. 2000, Wiley-VCH Verlag GmbH & Co. KGaA.
158. Chaparro-Riggers, J.F., et al., *Comparison of three enoate reductases and their potential use for biotransformations*. Advanced Synthesis & Catalysis, 2007. **349**(8-9): p. 1521-1531.
159. Islam, M.A., et al., *Exploring biochemical pathways for mono-ethylene glycol (MEG) synthesis from synthesis gas*. Metabolic Engineering, 2017. **41**: p. 173-181.
160. Li, C.H., et al., *Computational discovery of biochemical routes to specialty chemicals*. Chemical Engineering Science, 2004. **59**(22-23): p. 5051-5060.
161. Corey, E.J., *The Logic of Chemical Synthesis - Multistep Synthesis of Complex Carbogenic Molecules*. Angewandte Chemie-International Edition in English, 1991. **30**(5): p. 455-465.
162. Orth, J.D., I. Thiele, and B.Ø. Palsson, *What is flux balance analysis?* Nature biotechnology, 2010. **28**(3): p. 245-248.
163. Hadadi, N., et al., *Knowledge of the Neighborhood of the Reactive Site up to Three Atoms Can Predict Biochemistry and Protein Sequences*. bioRxiv, 2017: p. <https://doi.org/10.1101/210039>.
164. Frainay, C. and F. Jourdan, *Computational methods to identify metabolic sub-networks based on metabolomic profiles*. Briefings in Bioinformatics, 2017. **18**(1): p. 43-56.
165. Ataman, M. and V. Hatzimanikatis, *Heading in the right direction: thermodynamics-based network analysis and pathway engineering*. Current Opinion in Biotechnology, 2015. **36**: p. 176-182.

Bibliography

166. James, C.A. and D. Weininger, *Daylight Theory Manual*. Daylight Chemical Information Systems, Inc.: Irvine, CA.
167. Bajusz, D., A. Racz, and K. Heberger, *Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?* Journal of Cheminformatics, 2015. **7**.
168. Giri, V., et al., *RxnSim: a tool to compare biochemical reactions*. Bioinformatics, 2015. **31**(22): p. 3712-3714.
169. Petzold, C.J., et al., *Analytics for Metabolic Engineering*. Frontiers in Bioengineering and Biotechnology, 2015. **3**: p. 135.
170. Campbell, K., J. Xia, and J. Nielsen, *The Impact of Systems Biology on Bioprocessing*. Trends in Biotechnology. **35**(12): p. 1156-1168.
171. Miskovic, L., et al., *A design-build-test cycle using modeling and experiments reveals interdependencies between upper glycolysis and xylose uptake in recombinant S. cerevisiae and improves predictive capabilities of large-scale kinetic models*. Biotechnology for Biofuels, 2017. **10**.
172. Lilley, D.M.J., et al., *A nomenclature of junctions and branchpoints in nucleic acids*. Nucleic Acids Research, 1995. **23**(17): p. 3363-3364.
173. Neidhardt, F.C., J.L. Ingraham, and M. Schaechter, *Physiology of the bacterial cell : a molecular approach*. 1990, Sunderland, Mass.: Sinauer Associates. xii, 506 p.
174. Sudarsan, S., et al., *The Functional Structure of Central Carbon Metabolism in Pseudomonas putida KT2440*. Applied and Environmental Microbiology, 2014. **80**(17): p. 5292-5303.
175. Chen, Y., et al., *Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism*. Metab Eng, 2013. **15**: p. 48-54.
176. Haller, T., et al., *Discovering New Enzymes and Metabolic Pathways: Conversion of Succinate to Propionate by Escherichia coli*. Biochemistry, 2000. **39**(16): p. 4622-4629.
177. Bordbar, A., et al., *Constraint-based models predict metabolic and associated cellular functions*. Nature Reviews Genetics, 2014. **15**(2): p. 107-120.
178. Maarleveld, T.R., et al., *Basic concepts and principles of stoichiometric modeling of metabolic networks*. Biotechnology Journal, 2013. **8**(9): p. 997-U52.
179. Garcia-Albornoz, M.A. and J. Nielsen, *Application of Genome-Scale Metabolic Models in Metabolic Engineering*. Industrial Biotechnology, 2013. **9**(4): p. 203-214.

180. Miskovic, L., et al., *Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes*. Current Opinion in Biotechnology, 2015. **36**: p. 146-153.
181. Chakrabarti, A., et al., *Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints*. Biotechnology journal, 2013. **8**(9): p. 1043-1057.
182. Savoglidis, G., et al., *A method for analysis and design of metabolism using metabolomics data and kinetic models: Application on lipidomics using a novel kinetic model of sphingolipid metabolism*. Metabolic Engineering, 2016. **37**: p. 46-62.
183. Stanford, N.J., et al., *Systematic Construction of Kinetic Models from Genome-Scale Metabolic Networks*. PLOS One, 2013. **8**(11).
184. Dash, S., et al., *Development of a core Clostridium thermocellum kinetic metabolic model consistent with multiple genetic perturbations*. Biotechnology for Biofuels, 2017. **10**.
185. Lee, Y., J.G. Lafontaine Rivera, and J.C. Liao, *Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways*. Metabolic Engineering, 2014. **25**(Supplement C): p. 63-71.
186. Esvelt, K.M. and H.H. Wang, *Genome-scale engineering for systems and synthetic biology*. Molecular Systems Biology, 2013. **9**.
187. Barrangou, R. and J.A. Doudna, *Applications of CRISPR technologies in research and beyond*. Nature Biotechnology, 2016. **34**(9): p. 933-941.
188. Bochner, B.R., *Global phenotypic characterization of bacteria*. Fems Microbiology Reviews, 2009. **33**(1): p. 191-205.
189. Blount, Z.D., *The unexhausted potential of E. coli*. Elife, 2015. **4**.
190. Liu, T.G. and C. Khosla, *Genetic Engineering of Escherichia coli for Biofuel Production*. Annual Review of Genetics, Vol 44, 2010. **44**: p. 53-69.
191. Wen, M., B.B. Bond-Watts, and M.C.Y. Chang, *Production of advanced biofuels in engineered E. coli*. Current Opinion in Chemical Biology, 2013. **17**(3): p. 472-479.
192. Brynildsen, M.P. and J.C. Liao, *An integrated network approach identifies the isobutanol response network of Escherichia coli*. Molecular Systems Biology, 2009. **5**.
193. Sheridan, C., *Making green*. Nature Biotechnology, 2009. **27**(12): p. 1074-1076.

Bibliography

194. Jin, H., et al., *Engineering biofuel tolerance in non-native producing microorganisms*. Biotechnology Advances, 2014. **32**(2): p. 541-548.
195. Nielsen, D.R., et al., *Engineering alternative butanol production platforms in heterologous bacteria*. Metabolic Engineering, 2009. **11**(4-5): p. 262-273.
196. Vicente, M. and J.L. Cánovas, *Glucolysis in *Pseudomonas putida*: Physiological Role of Alternative Routes from the Analysis of Defective Mutants*. Journal of Bacteriology, 1973. **116**(2): p. 908-914.
197. Hollinshead, W.D., et al., *Examining Escherichia coli glycolytic pathways, catabolite repression, and metabolite channeling using Delta pfk mutants*. Biotechnology for Biofuels, 2016. **9**.
198. Flamholz, A., et al., *Glycolytic strategy as a tradeoff between energy yield and protein cost*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(24): p. 10039-10044.
199. Noor, E., et al., *Pathway Thermodynamics Highlights Kinetic Obstacles in Central Metabolism*. Plos Computational Biology, 2014. **10**(2).
200. Duan, Y.K., et al., *De novo Biosynthesis of Biodiesel by Escherichia coli in Optimized Fed-Batch Cultivation*. Plos One, 2011. **6**(5).
201. Marella, E.R., et al., *Engineering microbial fatty acid metabolism for biofuels and biochemicals*. Current Opinion in Biotechnology, 2018. **50**: p. 39-46.
202. Sun, T., et al., *Re-direction of carbon flux to key precursor malonyl-CoA via artificial small RNAs in photosynthetic Synechocystis sp PCC 6803*. Biotechnology for Biofuels, 2018. **11**.
203. Kuepper, J., et al., *Metabolic Engineering of Pseudomonas putida KT2440 to Produce Anthranilate from Glucose*. Frontiers in Microbiology, 2015. **6**.
204. Yu, S., et al., *Metabolic Engineering of Pseudomonas putida KT2440 for the Production of para-Hydroxy Benzoic Acid*. Frontiers in Bioengineering and Biotechnology, 2016. **4**: p. 90.
205. Romeo, T. and J. Snoep, *Glycolysis and Flux Control*. EcoSal Plus, 2005.
206. Kerkhoven, E.J., P.-J. Lahtvee, and J. Nielsen, *Applications of computational modeling in metabolic engineering of yeast*. FEMS yeast research, 2014.
207. Schomburg, I., et al., *BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA*. Nucleic Acids Research, 2013. **41**(Database issue): p. D764-72.

208. Wittig, U., et al., *SABIO-RK-database for biochemical reaction kinetics*. Nucleic Acids Research, 2012. **40**(D1): p. D790-D796.
209. van Eunen, K., et al., *Testing biochemistry revisited: how in vivo metabolism can be understood from in vitro enzyme kinetics*. PLoS computational biology, 2012. **8**(4).
210. Smallbone, K., et al., *A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes*. FEBS letters, 2013. **587**(17): p. 2832-2841.
211. Almquist, J., et al., *Kinetic models in industrial biotechnology--improving cell factory performance*. Metabolic engineering, 2014. **24**: p. 38-60.
212. Murabito, E., et al., *Monte-Carlo modeling of the central carbon metabolism of Lactococcus lactis: insights into metabolic regulation*. PLoS One, 2014. **9**(9): p. e106453.
213. Lee, Y., J.G.L. Rivera, and J.C. Liao, *Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways*. Metabolic engineering, 2014. **25**: p. 63-71.
214. dos Santos, F.B., W.M. de Vos, and B. Teusink, *Towards metagenome-scale models for industrial applications—the case of Lactic Acid Bacteria*. Current opinion in biotechnology, 2013. **24**(2): p. 200-206.
215. Chowdhury, A., A.R. Zomorodi, and C.D. Maranas, *k-OptForce: Integrating Kinetics with Flux Balance Analysis for Strain Design*. Plos Computational Biology, 2014. **10**(2).
216. Jamshidi, N. and B.O. Palsson, *Mass Action Stoichiometric Simulation Models: Incorporating Kinetics and Regulation into Stoichiometric Models*. Biophysical Journal, 2010. **98**(2): p. 175-185.
217. Khodayari, A., et al., *A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data*. Metabolic Engineering, 2014. **25**: p. 50-62.
218. Smallbone, K., et al., *Towards a genome-scale kinetic model of cellular metabolism*. BMC Systems Biology, 2010. **4**(1): p. 6.
219. Khodayari, A. and C.D. Maranas, *A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains*. Nature Communications, 2016. **7**.
220. Chassagnole, C., et al., *Dynamic modeling of the central carbon metabolism of Escherichia coli*. Biotechnology and Bioengineering, 2002. **79**(1): p. 53-73.

Bibliography

221. Ishii, N., et al., *Dynamic simulation of an in vitro multi-enzyme system*. Febs Letters, 2007. **581**(3): p. 413-420.
222. Gutenkunst, R.N., et al., *Universally sloppy parameter sensitivities in systems biology models*. Plos Computational Biology, 2007. **3**(10): p. 1871-1878.
223. Costa, R.S., et al., *An extended dynamic model of Lactococcus lactis metabolism for mannitol and 2, 3-butanediol production*. Molecular BioSystems, 2014. **10**(3): p. 628-639.
224. Guillen-Gosalbez, G., et al., *Identification of regulatory structure and kinetic parameters of biochemical networks via mixed-integer dynamic optimization*. BMC systems biology, 2013. **7**(1): p. 113.
225. Khodayari, A., A. Chowdhury, and C.D. Maranas, *Succinate overproduction: a case study of computational strain design using a comprehensive Escherichia coli kinetic model*. Frontiers in bioengineering and biotechnology, 2014. **2**.
226. Chowdhury, A., A.R. Zomorodi, and C.D. Maranas, *k-OptForce: integrating kinetics with flux balance analysis for strain design*. PLoS Comput. Biol, 2014. **10**(2): p. e1003487.
227. Smallbone, K. and P. Mendes, *Large-scale metabolic models: From reconstruction to differential equations*. Industrial Biotechnology, 2013. **9**(4): p. 179-184.
228. Weaver, L.J., et al., *A kinetic-based approach to understanding heterologous mevalonate pathway function in E. coli*. Biotechnology and bioengineering, 2015. **112**(1): p. 111-119.
229. Levering, J., et al., *Role of phosphate in the central metabolism of two lactic acid bacteria--a comparative systems biology approach*. FEBS Journal, 2012. **279**(7): p. 1274-1290.
230. Kerkhoven, E.J., et al., *Handling Uncertainty in Dynamic Models: The Pentose Phosphate Pathway in Trypanosoma brucei*. Plos Computational Biology, 2013. **9**(12).
231. Link, H., D. Christodoulou, and U. Sauer, *Advancing metabolic models with kinetic information*. Current opinion in biotechnology, 2014. **29**: p. 8-14.
232. Rutkis, R., et al., *Kinetic modeling of Zymomonas mobilis Entner-Doudoroff pathway: insights into control and functionality*. Microbiology, 2013: p. mic-0.
233. Sen, P., H.J. Vial, and O. Radulescu, *Kinetic modelling of phospholipid synthesis in Plasmodium knowlesi unravels crucial steps and relative importance of multiple pathways*. BMC systems biology, 2013. **7**(1): p. 123.

234. Birkenmeier, M., S. Neumann, and T. Roeder, *Kinetic modeling of riboflavin biosynthesis in Bacillus subtilis under production conditions*. Biotechnology letters, 2014. **36**(5): p. 919-928.
235. Palsson, B.O. and I.D. Lee, *Model Complexity Has a Significant Effect on the Numerical Value and Interpretation of Metabolic Sensitivity Coefficients*. Journal of Theoretical Biology, 1993. **161**(3): p. 299-315.
236. Bakker, B.M., et al., *Systems biology from micro-organisms to human metabolic diseases: the role of detailed kinetic models*. Biochemical Society Transactions, 2010. **38**: p. 1294-1301.
237. Chen, Y., et al., *Ach1 is involved in shuttling mitochondrial acetyl units for cytosolic C2 provision in Saccharomyces cerevisiae lacking pyruvate decarboxylase*. Fems Yeast Research, 2015. **15**(3).
238. Wahrheit, J., A. Nicolae, and E. Heinzle, *Eukaryotic metabolism: Measuring compartment fluxes*. Biotechnology Journal, 2011. **6**(9): p. 1071-1085.
239. Klitgord, N. and D. Segré, *The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles*. Genome Inform, 2010. **22**: p. 41-55.
240. Chou, I.-C. and E.O. Voit, *Estimation of dynamic flux profiles from metabolic time series data*. BMC systems biology, 2012. **6**(1): p. 84.
241. Chen, N., et al., *Genome-based kinetic modeling of cytosolic glucose metabolism in industrially relevant cell lines: Saccharomyces cerevisiae and Chinese hamster ovary cells*. Bioprocess and biosystems engineering, 2012. **35**(6): p. 1023-1033.
242. Pontes Freitas Alberton, K., et al., *Simultaneous Parameters Identifiability and Estimation of an E. coli Metabolic Network Model*. BioMed research international, 2015. **2015**.
243. Link, H., K. Kochanowski, and U. Sauer, *Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo*. Nature biotechnology, 2013. **31**(4): p. 357-361.
244. Matsuoka, Y. and K. Shimizu, *Catabolite regulation analysis of Escherichia coli for acetate overflow mechanism and co-consumption of multiple sugars based on systems biology approach using computer simulation*. Journal of biotechnology, 2013. **168**(2): p. 155-173.
245. Cintolesi, A., et al., *Quantitative analysis of the fermentative metabolism of glycerol in Escherichia coli*. Biotechnology and Bioengineering, 2012. **109**(1): p. 187-198.

Bibliography

246. Peskov, K., E. Mogilevskaya, and O. Demin, *Kinetic modelling of central carbon metabolism in Escherichia coli*. FEBS Journal, 2012. **279**(18): p. 3374-3385.
247. Tohsato, Y., et al., *Parameter optimization and sensitivity analysis for large kinetic models using a real-coded genetic algorithm*. Gene, 2013. **518**(1): p. 84-90.
248. Andreozzi, S., et al., *Exploring enhancement of 1,4-butanediol production in recombinant E. coli using large-scale kinetic models*. Abstracts of Papers of the American Chemical Society, 2016. **251**.
249. Raganati, F., et al., *Kinetic study of butanol production from various sugars by Clostridium acetobutylicum using a dynamic model*. Biochemical Engineering Journal, 2015. **99**: p. 156-166.
250. Saitua, F., et al., *Dynamic genome-scale metabolic modeling of the yeast Pichia pastoris*. BMC Systems Biology, 2017. **11**.
251. Koenig, M., S. Bulik, and H.-G. Holzhutter, *Quantifying the contribution of the liver to glucose homeostasis: a detailed kinetic model of human hepatic glucose metabolism*. PLoS computational biology, 2012. **8**(6): p. e1002577.
252. Khazaei, T., A. McGuigan, and R. Mahadevan, *Ensemble modeling of cancer metabolism*. Frontiers in physiology, 2012. **3**.
253. Abuhamed, T., et al., *Kinetics model for growth of Pseudomonas putida F1 during benzene, toluene and phenol biodegradation*. Process Biochemistry, 2004. **39**(8): p. 983-988.
254. Annuar, M.S.M., et al., *A kinetic model for growth and biosynthesis of medium-chain-length poly-(3-hydroxyalkanoates) in Pseudomonas putida*. Brazilian Journal of Chemical Engineering, 2008. **25**(2): p. 217-228.
255. Bandyopadhyay, K., D. Das, and B.R. Maiti, *Kinetics of phenol degradation using Pseudomonas putida MTCC 1194*. Bioprocess Engineering, 1998. **18**(5): p. 373-377.
256. Calzada, J., et al., *Extended kinetic model for DBT desulfurization using Pseudomonas Putida CECT5279 in resting cells*. Biochemical Engineering Journal, 2012. **66**: p. 52-60.
257. Choi, N.C., et al., *Modeling of growth kinetics for Pseudomonas putida during toluene degradation*. Applied Microbiology and Biotechnology, 2008. **81**(1): p. 135-141.
258. Hasan, S.A. and S. Jabeen, *Degradation kinetics and pathway of phenol by Pseudomonas and Bacillus species*. Biotechnology & Biotechnological Equipment, 2015. **29**(1): p. 45-53.

259. Kumar, A., S. Kumar, and S. Kumar, *Biodegradation kinetics of phenol and catechol using Pseudomonas putida MTCC 1194*. Biochemical Engineering Journal, 2005. **22**(2): p. 151-159.
260. Martin, A.B., et al., *Production of a biocatalyst of Pseudomonas putida CECT5279 for dibenzothiophene (DBT) biodesulfurization for different media compositions*. Energy & Fuels, 2004. **18**(3): p. 851-857.
261. Seker, S., et al., *Multi-substrate growth kinetics of Pseudomonas putida for phenol removal*. Applied Microbiology and Biotechnology, 1997. **47**(5): p. 610-614.
262. Wang, S.J. and K.C. Loh, *Biotransformation kinetics of Pseudomonas putida for cometabolism of phenol and 4-chlorophenol in the presence of sodium glutamate*. Biodegradation, 2001. **12**(3): p. 189-199.
263. Sudarsan, S., et al., *Dynamics of benzoate metabolism in Pseudomonas putida KT2440*. Metabolic Engineering Communications, 2016. **3**: p. 97-110.
264. Chavarria, M., et al., *A Metabolic Widget Adjusts the Phosphoenolpyruvate-Dependent Fructose Influx in Pseudomonas putida*. Msystems, 2016. **1**(6).
265. del Castillo, T., et al., *Convergent peripheral pathways catalyze initial glucose catabolism in Pseudomonas putida: Genomic and flux analysis*. Journal of Bacteriology, 2007. **189**(14): p. 5142-5152.
266. Ambrus, A., et al., *Stimulation of reactive oxygen species generation by disease-causing mutations of lipoamide dehydrogenase*. Human Molecular Genetics, 2011. **20**(15): p. 2984-2995.
267. Burgard, A.P., et al., *Flux coupling analysis of genome-scale metabolic network reconstructions*. Genome Research, 2004. **14**(2): p. 301-312.
268. David, L., et al., *FFCA: a feasibility-based method for flux coupling analysis of metabolic networks*. BMC Bioinformatics, 2011. **12**.
269. Placzek, S., et al., *BRENDA in 2017: new perspectives and new tools in BRENDA*. Nucleic Acids Research, 2017. **45**(Database issue): p. D380-D388.
270. Schomburg, I., et al., *BRENDA: a resource for enzyme data and metabolic information*. Trends in Biochemical Sciences, 2002. **27**(1): p. 54-56.
271. Schomburg, I., et al., *BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA*. Nucleic Acids Research, 2013. **41**(Database issue): p. D764-D772.

Bibliography

272. Schomburg, I., A. Chang, and D. Schomburg, *BRENDA, enzyme data and metabolic information*. Nucleic Acids Research, 2002. **30**(1): p. 47-49.
273. Hatzimanikatis, V. and J.E. Bailey, *MCA has more to say*. Journal of Theoretical Biology, 1996. **182**(3): p. 233-242.
274. Kacser, H., J.A. Burns, and D.A. Fell, *The Control of Flux*. Biochemical Society Transactions, 1995. **23**(2): p. 341-366.
275. Andreozzi, S., L. Miskovic, and V. Hatzimanikatis, *iSCHRUNK - In Silico Approach to Characterization and Reduction of Uncertainty in the Kinetic Models of Genome-scale Metabolic Networks*. Metabolic Engineering, 2016. **33**: p. 158-168.
276. Weilandt, D.R. and V. Hatzimanikatis, *Particle-based simulation reveals macromolecular crowding effects on the Michaelis-Menten mechanism*. bioRxiv 2018. **429316**.
277. Albe, K.R., M.H. Butler, and B.E. Wright, *Cellular Concentrations of Enzymes and Their Substrates*. Journal of Theoretical Biology, 1990. **143**(2): p. 163-195.
278. Gabor, A. and J.R. Banga, *Robust and efficient parameter estimation in dynamic models of biological systems*. BMC Systems Biology, 2015. **9**.
279. Gage, D.J. and F.C. Neidhardt, *Adaptation of Escherichia-Coli to the Uncoupler of Oxidative-Phosphorylation 2,4-Dinitrophenol*. Journal of Bacteriology, 1993. **175**(21): p. 7105-7108.
280. Blank, L.M., et al., *Metabolic response of Pseudomonas putida during redox biocatalysis in the presence of a second octanol phase*. Febs Journal, 2008. **275**(20): p. 5173-5190.
281. Christodoulou, D., et al., *Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress*. Cell Systems, 2018. **6**(5): p. 569-578.e7.
282. Hameri, T., et al., *Kinetic models of metabolism that consider alternative steady-state solutions of intracellular fluxes and concentrations*. Metabolic Engineering, 2019. **52**: p. 29-41.
283. Beg, Q.K., et al., *Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(31): p. 12663-12668.

- 284. Schnell, S. and T.E. Turner, *Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws*. Progress in Biophysics & Molecular Biology, 2004. **85**(2-3): p. 235-260.
- 285. ten Wolde, P.R. and A. Mugler, *Importance of Crowding in Signaling, Genetic, and Metabolic Networks*. New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals, 2014. **307**: p. 419-442.
- 286. Liao, J.C., et al., *Fuelling the future: microbial engineering for the production of sustainable biofuels*. Nature Reviews Microbiology, 2016. **14**(5): p. 288-304.
- 287. Wang, C.L., B.F. Pfleger, and S.W. Kim, *Reassessing Escherichia coli as a cell factory for biofuel production*. Current Opinion in Biotechnology, 2017. **45**: p. 92-103.
- 288. Demirel, Y.s., *Nonequilibrium thermodynamics transport and rate processes in physical, chemical and biological systems*. 3rd ed. 2014, Amsterdam: Elsevier. 758 S.

Milenko TOKIC, PhD

Avenue de Sévelin 13 C | 1004 Lausanne

+41 76 816 12 80 | milenko.tokic@gmail.com

Born on June 24th, 1984 | Serb, in possession of Swiss work permit | Married, one child

<https://www.linkedin.com/in/milenkotokic/>



CURRICULUM VITAE

Accomplished professional with interest in mathematical modelling, chemical engineering, systems biology, data analysis, modeling of large-scale metabolic networks and bacteria metabolism. Ambitious and always looking to expand knowledge and skills. Demonstrates good problem identifying skills and ability to use various approaches for their quick resolution.

PROFESSIONAL EXPERIENCE

- 2014 – ... **Doctoral assistant | Scientist** | Laboratory of Computational Systems Biotechnology (LCSB), Institute of Chemical Sciences and Engineering, EPFL, Switzerland
- Analysis of bacteria metabolism
 - Analysis of different types of omics data and their implementation into mathematical models
 - Evaluation and feasibility analysis of novel biosynthetic pathways
 - Classification of the pathways and identification of the enzymes and precursors that are fundamental for the production of the target molecule
 - Development of kinetic models and identification of metabolic engineering targets for the enhancement of production of desired molecule
 - Teaching assistant for the following courses: Introduction to chemical engineering, Numerical methods in chemistry, Computational biotechnology lab, Biochemical engineering, Bioreactor modeling and simulation
- 2009 – 2014 **Research assistant** | Department of Chemical Engineering, University of Novi Sad, Serbia
- Development and optimization of novel types of airlift reactors
 - Experimental investigation of the hydrodynamics and gas-liquid mass transfer coefficients
 - Determination of flow regime transitions
 - Prediction of hydrodynamic parameters
 - Supervision of students in their practical classes
 - Preparation of class notes
 - Teaching assistant for the following courses: Unit operations, Projecting of chemical industry devices

LANGUAGES

English	Fluent	Serbian	Native
French	Intermediate level, A2/B1		

COMPUTER SKILLS

Writing/Presenting	MS office (Word, Excel, PowerPoint)
Scientific computing	MATLAB
Programing languages	Limited proficiency with R and Python

EDUCATION

- 2014 – 2018 **PhD** | Laboratory of Computational Systems Biotechnology (LCSB), EPFL, Switzerland | supervised by prof. Vassily Hatzimanikatis and Dr Ljubisa Miskovic
- 2003 – 2008 **BSc and MSc in Chemical engineering** | Faculty of Technology, University of Novi Sad, Serbia
Average success: 9.03 (out of 10.0)

PUBLICATIONS AND PRESENTATIONS

Publications in peer reviewed journals:

- **Milenko Tokic**, Noushin Hadadi, Meric Ataman, Dario Neves, Birgitta E. Ebert, Lars M. Blank, Ljubisa Miskovic and Vassily Hatzimanikatis (2018) Discovery and evaluation of biosynthetic pathways for the production of five methyl ethyl ketone precursors, *ACS Synth. Biol.* 7, 1858–1873
- Ljubisa Miskovic, **Milenko Tokic**, Georgios Fengos and Vassily Hatzimanikatis (2015) Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes, *Curr Opin Biotechnol.* 36, 146-153.
- Predrag Kojic, **Milenko Tokic**, Ivana Sijacki, Natasa Lukic, Dragan Petrovic, Dragica Jovicevic and Svetlana Popovic (2015) Influence of the sparger type and added alcohol on the gas holdup of an external-loop airlift reactor, *Chem. Eng. Technol.* 38, No. 4, 701–708
- Ivana Sijacki, **Milenko Tokic**, Predrag Kojic, Dragan Lj. Petrovic, Miodrag N. Tekic and Natasa Lukic (2013) Sparger type influence on the volumetric mass transfer coefficient in the draft tube airlift reactor with diluted alcohol solutions, *Ind. Eng. Chem. Res.*, 52, 6812–6821
- Ivana Sijacki, **Milenko Tokic**, Predrag Kojic, Dragan Petrovic, Miodrag Tekic, Mirjana Djuric and Slobodan Milovancev (2011) Sparger type influence on the hydrodynamics of the draft tube airlift reactor with diluted alcohol solutions, *Ind. Eng. Chem. Res.*, 50, 3580–3591

Publications in preparations for peer reviewed journals:

- **Milenko Tokic**, Ljubisa Miskovic and Vassily Hatzimanikatis (2018) Large-scale kinetic metabolic models of *Pseudomonas putida* for consistent design of metabolic engineering strategies

Conference presentations:

- Work presented on numerous conferences such as: Metabolic engineering 2016 and 2018, Constraint-Based Reconstruction and Analysis (COBRA) 2015 and 2018 and Biochemical and Molecular Engineering 2017

AWARDS

Best poster award at 4th Conference on Constraint-Based Reconstruction and Analysis (COBRA 2015), Heidelberg, Germany

Serbian chemical society Special prize gainer 2009

ADDITIONAL INTERESTS

Traveling | Skiing | Football | Classical and Japanese samurai cinema
Driving license B

REFERENCES

Prof. Vassily Hatzimanikatis PhD advisor | Laboratory of Computational Systems Biotechnology EPFL, Switzerland
email: vassily.hatzimanikatis@epfl.ch | Tel: +41 (0)21 693 98 70

Dr Ljubisa Miskovic PhD co-advisor | Laboratory of Computational Systems Biotechnology EPFL, Switzerland
email: ljubisa.miskovic@epfl.ch | Tel: +41 (0)21 69 39892

Prof. Radovan Omorjan Professor | Faculty of Technology, Department of Chemical Engineering, Novi Sad, Serbia
email: omorr@uns.ac.rs | Tel: +381 (0)21 485 3671

