UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA QUÍMICA Y
BIOTECNOLOGÍA

SYSTEMS BIOLOGY AND CHEMOINFORMATICS METHODS FOR BIOMINING AND SYSTEMS METABOLIC ENGINEERING APPLICATIONS

TESIS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS DE LA
INGENIERÍA MENCIÓN QUÍMICA

MIGUEL ANGEL CAMPODONICO ALT

PROFESOR GUÍA:
JUAN A. ASENJO DE LEUZE
PROFESOR CO-GUIA:
BARBARA A. ANDREWS

MIEMBROS DE LA COMISIÓN:
EDUARDO AGOSIN TRUMPER
MA. ORIANA SALAZAR AGUIRRE
IVAN RAPAPORT

SANTIAGO DE CHILE
2014

SYSTEMS BIOLOGY AND CHEMOINFORMATICS METHODS FOR BIOMINING AND SYSTEMS METABOLIC ENGINEERING APPLICATIONS

In the first chapter, this thesis aims to demonstrate the great potential of Constraint-Based Reconstruction and Analysis (COBRA) methods for studying and predicting specific phenotypes in the bacterium *Acidithiobacillus ferrooxidans*. A genome-scale metabolic reconstruction of *Acidithiobacillus ferrooxidans ATCC 23270* (iMC507) is presented and characterized. *i*MC507 is validated for aerobic chemolithoautotrophic conditions by fixating carbon dioxide and using three different electron donors: ferrous ion, tetrathionate and thiosulfate. Furthermore, the model is utilized for (i) quantitatively studying and analyzing key reactions and pathways involved in the electron transfer metabolism, (ii) describing the central carbon metabolism and (iii) for evaluating the potential to couple the production of extracellular polymeric substances through knock-outs. The second chapter work outlines the effort towards advancing the field of systems metabolic engineering by using COBRA methods in conjunction with chemoinformatic approaches to metabolically engineer the bacterium *Escherichia coli*. A complete strain design workflow integrating synthetic pathway prediction with growth-coupled designs for the production of non-native compounds in a target organism of interest is outlined. The generated enabling technology is a computational pipeline including chemoinformatics, bioinformatics, constraint-based modeling, and GEMs to aid in the process of metabolic engineering of microbes for industrial bioprocessing purposes. A retrosynthetic based pathway predictor algorithm containing a novel integration with GEMs and reaction promiscuity analysis is developed and demonstrated. Specifically, the production potential of 20 industrially-relevant chemicals in *E. coli* and feasible designs for production strains generation is outlined. A comprehensive mapping from *E. coli's* native metabolome to commodity chemicals that are 4 reactions or less away from a natural metabolite is performed. Sets of metabolic interventions, specifically knock-outs and knock-ins that coupled the target chemical production to growth rate were determined. In the third chapter, in order to aid the field of cancer metabolism, potential biomarkers were determined through gain of function oncometabolites predictions. Based on a chemoinformatic approach in conjunction with the global human metabolic network Recon 2, a workflow for predicting potential oncometabolites is constructed. Starting from a list of mutated enzymes genes, described as GoF mutations, a range of promiscuous catalytic activities are inferred. In total 24 chemical substructures of oncometabolites resulting from the GoF analysis are predicted.

MÉTODOS DE BIOLOGÍA DE SISTEMAS Y QUIMIOINFORMÁTICOS APLICADOS EN EL ÁREA DE BIOMINERÍA E INGENIERÍA METABÓLICA

En el primer capítulo de esta tesis, se demostró el potencial que tienen los método COBRA, "Constraint-Based Reconstruction and Analysis" en ingles, para el estudio y la predicción de fenotipos experimentales en la bacteria *Acidithiobacillus ferrooxidans ATCC 23270*. Para esto se reconstruyó el primer modelo a escala genómica de *A. ferrooxidans* (iMC507). El modelo se validó utilizando dióxido de carbono como única fuente de carbono bajo tres distintas condiciones de crecimiento aeróbico-quimiolitoautótrofico dependiendo el dador final electrones (estos son: ion ferroso, tetrationato y tiosulfato). Junto con lo anterior se estudió de forma cuantitativa y analizaron las reacciones y vías metabólicas asociadas a la cadena de transporte de electrones y el metabolismo central. Además, se evaluó el potencial de producción de sustancias poliméricas extracelulares acopladas al crecimiento a través de knock-outs. En el segundo capítulo se esboza el trabajo enfocado al desarrollo del campo de "Systems Metabolic Engineering" a través de la utilización de métodos COBRA junto con herramientas quimioinformáticas para diseñar de forma insilico la bacteria *Escherichia coli*. Se desarrolló un flujo de trabajo, integrando la predicción de vías metabólicas sintéticas junto con el diseño computacional de cepas capaces de asociar la producción de un determinado compuesto al crecimiento. La tecnología desarrollada involucra una serie de herramientas computacionales asociadas a los campos de investigación en quimioinformática, bioinformática, métodos COBRA y modelos a escala genómica. Específicamente, se desarrolló y valido un algoritmo de predicción de vías metabólicas basado en reglas de reacción restrosintéticas junto con un análisis de promiscuidad de reacciones. El flujo de trabajo se aplicó de forma exhaustiva en *E. coli* para estudiar la factibilidad de producir 20 compuestos químicos (en su mayoría no nativos a la cepa) relevantes para la industria. De acuerdo a lo anterior se determinó el conjunto de intervenciones metabólicas, tales como knock-ins y knock-outs, necesarios para la producción asociada al crecimiento de los compuestos químicos seleccionados. En el tercer capítulo, con el fin de expandir el campo de estudio asociado al metabolismo del cáncer, se determinaron una serie de biomarcadores a través de la predicción de oncometabolitos debido a mutaciones enzimáticas GoF ("Gain-of-Function" en inglés). El flujo de trabajo utilizado se basa principalmente en la utilización de herramientas quimioinformáticas enfocadas al modelo a escala genómica humano Recon 2. Una serie de actividades catalíticas promiscuas se infirieron a partir de un conjunto de genes mutados asociados a enzimas previamente catalogadas como mutaciones GoF. Basados en el análisis anterior, se determinaron 24 potenciales subestructuras de oncometabolitos capaces de funcionar como biomarcadores.

Para Manuela

Hope lies in dreams, in imagination and in the courage of those

who dare to make dreams into reality

Jonas Salk

Acknowledgements

During this almost five years of graduate work, I would like to thank to everyone that I have encounter during this period, from which I have gained support, knowledge, inspiration and motivation that made possible the culmination of this thesis. Although there are too many to name, I would like to first thank my advisor, Juan A. Asenjo, for the support and guidance he has provided during my Ph.D research. I will be forever grateful for all the opportunities to develop my career and enlightened myself in the scientific research that I received under his hospice. I also want to also thank the members of my thesis committee for their insights and guidance in the work presented here

I would like to specially thank Adam A. Feist, from the Systems Biology Research Group (UCSD, CA), for the invaluable guidance and mentoring he has provided through my scientific development. His particular vision and way of managing scientific research will deeply influence my future scientific career.

Also I would like to thank all my fellow graduate colleagues and researches, from both the Centre for Biochemical Engineering and Biotechnology (CeBiB) and to those who I collaborated with in the Systems Biology Research Group (UCSD, CA) during an almost two year-long visit, for their insights, collaboration and friendship.

Nothing of this would have happened without the unbreakable support of my family and friends. I thank my parents, Julio and Ana, who thought me to live for my dreams and work hard to make them happen. And lastly and most importantly I am immensely grateful to my everlasting companion, Manuela C. Gajardo, who I share this thesis with. None of this would have been possible without their continuous support.

Chapter 1, in full, is adapted from Genome-scale reconstruction of the acidophilic chemolithoautotrophic bacterium Acidithiobacillus ferrooxidans and a model-driven analysis of carbon fixation, electron transfer metabolism and extracellular polymeric substances production, that is in preparation. The dissertation author was the primary author of this paper, which was co-authored by Daniela Vaisman, Jean F. Castro,

Table of Content

List of Tables

List of Figures

List of Publications

Miguel A. Campodonico, Barbara A. Andrews, B.A., Juan A. Asenjo, Bernhard O. Palsson, Adam M. Feist, 2014. Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path, Metabolic Engineering, 25, 140-158.

Hojung Nam, Miguel A. Campodonico, Aarash Bordbar, Daniel R. Hyduke, Sangwoo Kim, Daniel Zielinski, Bernhard O. Palsson, 2014. A systems approach to predict oncometabolites via context-specific genome-scale metabolic networks. PLoS Computational Biology, 10(9), e1003837.

Miguel A. Campodonico, Daniela Vaisman, Jean F. Castro, Valeria Razmlich, Francesca Mercado, Barbara A. Andrews, Adam M. Feist, Juan A. Asenjo. Genome-scale reconstruction of the acidophilic chemolithoautotrophic bacterium Acidithiobacillus ferrooxidans and a model-driven analysis of carbon fixation, electron transfer metabolism and extracellular polymeric substances production. In preparation.

Introduction

Constraint-Based Reconstruction and Analysis (COBRA) modeling, a systems biology emerging field of research, enables to utilize high-throughput experimental data and quantitatively computational analysis to study and predict metabolic phenotypes. The "cornerstone" of this method is represented by the genome-scale model (GEM) that is built systematically using genome annotation, "omics" data sets and legacy knowledge. Genes, proteins, reactions and metabolites that participate in the metabolism are identified, categorized and systematically interconnected, enabling a mechanistic description of metabolic physiology. A GEM combined with constraints-based methods can be used to formulate mechanistic predictions of metabolic physiology that can be used in a prospective manner to elucidate new biological knowledge and understanding, as well as design and engineer the cellular metabolism (Feist et al. 2009; Orth et al. 2011). Several workflows have been implemented to predict metabolic phenotypes, by integrating high-throughput data sets with COBRA methods. These recent successes have considerable implications in the fields of microbial evolution, interaction networks, genetic engineering and drug discovery (Bordbar et al. 2014b). In this work three different aspects, in the context of COBRA methods and approaches, have been explored.

First, in Chapter 1 attention to the gram-negative, highly acidophilic, chemolithoautotrophic γ-proteobacterium *Acidithiobacillus ferrooxidans* was drawn. This organism typically grows at an external pH 2 or lower using the oxidation of ferrous ions by oxygen, producing ferric ions and water, while fixing carbon dioxide from the environment. It can also obtain energy by the oxidation of reduced inorganic sulfur compounds (RISCs), hydrogen and formate. *A. ferrooxidans* is of great interest for biomining and environmental applications, as they can process mineral ores by alleviating the negative environmental consequences derived from the mining processes. Based on the complete genome sequence of *A. ferrooxidans* (Valdes et al. 2008), several studies have provided insights into its physiological properties, including the most relevant metabolisms (Esparza et al. 2010; Ferguson and Ingledew 2008;

1

Osorio et al. 2013; Valdes, Pedroso, Quatrini, Dodson, Tettelin, Blake, Eisen and Holmes 2008). However, from this knowledge it is not possible to neither predict nor quantitatively describe the physiological outcome from the annotated sequence alone. Nevertheless, only small scale metabolic models for *A. ferrooxidans* (Hold et al. 2009; Sepúlveda et al. 2011), have been proven successful on describing the main aspects of his metabolism. However these models lack the ability of capturing the complex physiological characteristic, behavior and metabolic capabilities of the cell as a whole integrated system. In order to overcome this difficulty and fully explore metabolic genotype-phenotype relationships, implementation of constraint-based reconstruction and analysis (COBRA) methods for *A. ferrooxidans* are needed. In this work, a GEM is reconstructed, validated and analyzed to better understand key metabolic capabilities of *A. ferrooxidans* ATCC 23270. A systematic workflow for the elucidation of proton translocation stoichiometry of key enzymes based on physiological data is implemented under three different aerobic chemolithoautotrophic conditions (i.e., by using ferrous ion, tetrathionate and thiosulfate as electron donors). Transfer electron and central carbon metabolism is characterized and studied. Furthermore, in order to determine the potential production of extracellular polymeric substances (EPS) in *A. ferrooxidans*, a retrospective model drive analysis was performed.

Second, in Chapter 2, a systematic workflow to evaluate the production potential of 20 industrially relevant chemicals (Assary and Broadbelt 2011; Curran and Alper 2012; Fischer et al. 2008; Lee et al. 2012; Paster et al. 2003; Werpy et al. 2004; Zeng and Sabra 2011) in *E. coli*, by integrating a combination of computational methods and developing a new pathway prediction algorithm, GEM-Path (Genome-scale Model Pathway Predictor) is implemented. Computational approaches for the prediction of non-native pathways exist, but are limited in their design and scope. Different approaches have been implemented for pathway prediction (Arita 2000; Carbonell et al. 2011; Cho et al. 2010; Dale et al. 2010; Greene et al. 1999; Hatzimanikatis et al. 2005; Heath et al. 2010; Hou et al. 2003; McShan et al. 2003; Pharkya et al. 2004), where increasing attention has been focused mainly on retrosynthetic algorithms (Carbonell, Planson, Fichera and Faulon 2011; Cho, Yun, Park, Lee and Park 2010; Henry et al. 2010; Yim et al. 2011) based on Biochemical Reaction Operators (BROs). In these analyses, BROs

are used to go from a target compound to a predefined set of metabolites in an iterative backward search. In summary, all of these methods shared basically the same workflow, first calculating all structurally possible pathways and then scoring them using different kinds of metrics. During the synthetic pathway calculation, these algorithms unnecessarily expand the reaction space, generating all possible pathways that link a specific metabolite to a final specific product without performing pathway integration with content known to exist in a given production host. Furthermore, previous algorithms do not integrate the bioprocessing condition-specific cofactor usage/generation, substrate usage, strain/oxygenation conditions, and related energy balances during the computation of pathways. In order to address these problems, we developed GEM-Path, by integrating retrosynthetic algorithms based on BROs and filtering procedures with GEMs at each iteration step. Furthermore, a novel reaction promiscuity analysis is introduced, which is based on known reaction substrate similarities. These two features distinguish GEM-Path from other computational approaches. Once a synthetic pathway is successfully established, additional approaches can be taken to further engineer the host strain and synthetic pathways for enhanced production of a desired chemical. Adaptive laboratory evolution together with COBRA methods and organism-specific models have proven successful for the calculation of wild type *E. coli* optimal growth rates (Ibarra et al. 2002), native *E. coli* metabolite production through knock-outs (Fong et al. 2005), and for non-native *E. coli* metabolite production through heterologous pathway incorporation and knock-outs implementations (Yim, Haselbeck, Niu, Pujol-Baxley, Burgard, Boldt, Khandurina, Trawick, Osterhout, Stephen, Estadilla, Teisan, Schreyer, Andrae, Yang, Lee, Burk and Van Dien 2011). Furthermore, the use of adaptive laboratory evolution together with growth-coupled knock-outs design, allows to select for strains with higher target compound production rates by coupling them to the selection for faster growth (Portnoy et al. 2011). Here, we integrate each of the predicted pathways under several different substrates/strain/oxygenation conditions with growth-coupled designs generated through reaction knock-outs by utilizing the RobustKnock (Tepper and Shlomi 2010) and GDLS algorithms (Lun et al. 2009). Finally, in order to characterize *E. coli*'s potential production landscape for the studied compounds and for designs implementation purposes, a productivity analysis for maximum theoretical yield and maximum theoretical growth-coupled yield was performed.

And third, in Chapter 3, in order to aid the development of cancer diagnosis, prognosis, and biomarkers identification a chemoinformatic approach is implemented for oncometabolite predictions. The metabolic state coined as "aerobic glycolysis" (Warburg et al. 1924), which has been described as a passive response causing a malignant transformation (Hanahan and Weinberg 2011), is being challenged by recent discoveries. Studies have been proved that altered metabolism by itself can be a driver for oncogenesis (Letouze et al. 2013; Lu et al. 2012a; Xiao et al. 2012; Xu et al. 2011; Yang et al. 2012b). Specifically, characterized isocitrate dehydrogenase (IDH1, IDH2) mutations have established a new paradigm in cancer development in that the heterozygous point mutations confer a new metabolic enzymatic activity that produce an oncometabolite (e.g. 2-hydroxyglutarate (2-HG), from α-ketoglutarate(α-KG)). In fact, 2-HG shows a 100-fold increase concentration in glioma and acute myeloid leukemia's (AML) patients with IDH1 or IDH2 missense mutations. This increased concentration of 2-HG competitively inhibits α-ketoglutarate binding to histone demethylases, thus blocking differentiation of cells (Lu, Ward, Kapoor, Rohle, Turcan, Abdel-Wahab, Edwards, Khanin, Figueroa, Melnick, Wellen, O'Rourke, Berger, Chan, Levine, Mellinghoff and Thompson 2012a; Xu, Yang, Liu, Yang, Wang, Kim, Ito, Yang, Xiao, Liu, Jiang, Liu, Zhang, Wang, Frye, Zhang, Xu, Lei, Guan, Zhao and Xiong 2011). Recently, the COBRA approach has been proven successful in addressing heterogeneity in cancer by integrating experimental data with the GEMs to tailor the models to the unique gene expression profiles of general cancer tissue, and even individual cell lines and tumors. This enables to study the cancer metabolism and simulate tumor phenotypes from a genome wide perspective (Lewis and Abdel-Haleem 2013). Based on genetic mutation information on a massive scale collected from more than 1,700 cancer genomes into context-specific GEMs of metabolism for nine cancer types, Nam et al., 2014 (Nam et al. 2014) determined 6 different mutated metabolic enzymatic genes for GoF analysis, which presented missense mutations similar to those presented in the IDH case. According to the latter, in this work, a chemoinformatic based workflow is generated for predicting potential oncometabolites chemical structures due to GoF mutation.

Chapter 1

Genome-scale reconstruction of the acidophilic chemolithoautotrophic bacterium *Acidithiobacillus ferrooxidans* and a model-driven analysis of carbon fixation, electron transfer metabolism and extracellular polymeric substances production.

1.1 Abstract

Acidithiobacillus ferrooxidans is a gram-negative, highly acidophilic, chemolithoautotrophic γ-proteobacterium. It typically grows at an external pH of 2 or lower using the oxidation of ferrous ions by oxygen, producing ferric ions and water, while fixing carbon dioxide from the environment. It can also obtain energy by the oxidation of reduced inorganic sulfur compounds (RISCs), hydrogen and formate. *A. ferrooxidans* is of great interest for biomining and environmental applications, as it can process mineral ores and alleviate the negative environmental consequences derived from the mining processes. In this study, the first genome-scale metabolic reconstruction of *A. ferrooxidans* ATCC 23270 was generated (*i*MC507). Based on the annotated genome and available biochemical data, a total of 587 metabolic and transport/exchange reactions, 507 genes and 573 unique metabolites organized in over 42 subsystems were incorporated into the model. Based on a new genetic algorithm, an approach that integrates flux balance analysis (FBA), chemiosmotic theory, and physiological data was used to estimate proton translocation stoichiometry and maintenance parameters under aerobic chemolithoautotrophic conditions using thee

different electron donors (specifically for ferrous ion, tetrathionate and thiosulfate). The robustness of the metabolism with respect to reaction deletions using FBA was evaluated. Approximately 68 % of all possible single reaction deletions would be lethal for the organism. Detailed electron transfer flux distributions during chemolithoautotrophic growth using ferrous ion, tetrathionate and thiosulfate were determined and reported. Furthermore, a metabolic carbon dioxide transport outside the cell was suggested to alleviate the network rigidity in the carbon metabolism. Finally, 134 growth-coupled designs were calculated that enables EPS production. Two enzymes (i.e. MDH and FUM in the TCA cycle) were found to be predominantly involved in the metabolism disruption for EPS growth–coupled production. The reconstruction serves as a knowledgebase for summarizing and categorizing the information currently available for *A. ferrooxidans* and enables the understanding and engineering of *Acidithiobacillus* and similar species from a comprehensive model-driven perspective.

1.2 Introduction

The mining industry is a major force in the world economy, occupying a primary position at the start of the resource supply chain, supporting 14.4 % of the world's total economy, while using less than 1 % of the global surface area (CIA 2011). Production patterns are driven by consumption, which continues to rise in middle- to high- income countries, and is reaching unprecedented levels in low-income countries, whose appetite for the world's minerals reflects their rapid development (Fischer-Kowalski and Swilling 2011). However, extraction and processing are associated with a number of sustainable development challenges, including economic, environmental and social issues. For example, poor waste management practice, one of the most conspicuous features of the global mineral industry, can result in severe and long-term environmental and social consequences. Furthermore, it can also impose costs on mining companies by eroding share value, increasing the risks of temporary or permanent shut down, exposure to compensation, fines and litigation costs, lost future opportunities and increased remediation and monitoring (Franks et al. 2011).

A way to alleviate the negative consequences of mining is through the application of microbial processes, referred generically as "biomining". Almost without exception,

microbial extraction procedures are more environmentally friendly. They do not require the high amounts of energy used during roasting or smelting and do not produce sulfur dioxide or other environmentally harmful gaseous emissions. Furthermore, mine tailings and wastes produced from physicochemical processes when exposed to rain and air may be biologically leached, producing unwanted acid and metal pollution. Tailings and wastes from biomining operations are less chemically active, and the biological activity they can support is reduced by at least the extent to which they have already been bioleached. From an economical point of view, biomining has a clear advantage in the extraction and recovery of precious and base metals from low-grade ores, where many metals are not economically recoverable by non-biological methods (Rawlings 2002) (ores of copper, nickel, cobalt, zinc and uranium). At least 20 % of the copper produced worldwide comes from bioleaching (Rawlings and Johnson 2007).

There exist two major microbial mediated processes in biomining. The first is bioleaching, basically a strategy for metal recovery, whose underlying mechanism is the oxidation of metallic and/or sulfuric compounds by either enzymatic or mediated chemical oxidation caused by the catabolism of microorganisms. Depending on the mineral, chemical attack is by a combination of ferric iron and acid (protons), whereas the role of the microorganisms is to generate the ferric iron and acid. The second process is called biooxidation. This strategy applies mainly to the recovery of gold from difficult-to-treat arsenopyrites ores and concentrates. The aim is to use biooxidation to decompose the mineral matrix and expose entrapped gold (Rawlings et al. 2003). These processes are mediated by a consortium of Gram-negative bacteria (*Acidithiobacillus, Leptospirillum, Sulfobacillus, Acidimicrobium*) and archaeal genus (*Ferroplasma, Sulfolobulus and Metallosphaers*). There are many factors that affect the microbial composition of ores, such as, the type of mineral to be treated, the temperature, and the type of reactor used. Furthermore, industrial applications use both mixed populations (Ishigaki et al. 2005) as well as isolated cultures (Falco et al. 2003; Sand et al. 1992).

One of the most important, and by far the best characterized member of the biomining microbial consortia is *Acidithiobacillus ferrooxidans* (Brandl 2008; Edwards 1990; Ingledew 1982; Rawlings 2002). Formerly known as *Thiobacillus ferrooxidans,* it is a gram-negative, highly acidophilic, chemolithoautotrophic γ-proteobacterium

(Rohwerder et al. 2003). Beyond its biomining capabilities, *A. ferrooxidans* offers exceptional opportunities to study life under extremely challenging conditions. It typically grows at an external pH of 2 or lower using the oxidation of ferrous ions (fe2) by oxygen (o2), producing ferric ions (fe3) and water (h2o), while fixing carbon dioxide (co2) from the environment. It can also obtain energy by the oxidation of reduced inorganic sulfur compounds (RISCs), hydrogen (Drobner et al. 1990), and formate (Pronk et al. 1991). Furthermore, *A. ferrooxidans* has the potential to drive respiration by directly transferring electrons from electrodes to the microorganism (Carbajosa et al. 2010; Li et al. 2010). Due to the lack of well-developed systems for genetic manipulations, the study and exploration of the molecular biology and physiology of *A. ferrooxidans* has proven to be deficient. In terms on how the whole system behaves, different aspects of metabolism, such as, iron oxidation, co2 uptake and fixation, and the anaerobic metabolism of sulfur-coupled iron reduction remain little described. Furthermore, this organism has often proved to be source of much confusion, because it requires understanding of the consequences of both growing at very acidic external pH and of using a relatively weak reductant (ferrous iron) as the sole source of electrons for respiration (Ferguson and Ingledew 2008). Several aspects regarding its energetic metabolism remain scarcely described in quantitative terms, such as, how it balances the use of iron as both a micronutrient and as a required energy source. Also, how proton-driven membrane transport and energy processes function in the face of a proton motive force across the inner membrane that is several orders of magnitude higher, and how the large pH gradient is maintained across the cytoplasmic membrane of *A. ferrooxidans* (Ferguson and Ingledew 2008).

Based on the complete genome sequence of *A. ferrooxidans* (Valdes, Pedroso, Quatrini, Dodson, Tettelin, Blake, Eisen and Holmes 2008), several studies have provided insights into its physiological properties, including the most relevant metabolisms (Esparza, Cardenas, Bowien, Jedlicki and Holmes 2010; Ferguson and Ingledew 2008; Osorio, Mangold, Denis, Nancucheo, Esparza, Johnson, Bonnefoy, Dopson and Holmes 2013; Valdes, Pedroso, Quatrini, Dodson, Tettelin, Blake, Eisen and Holmes 2008). However, from this knowledge it is not possible to either predict or quantitatively describe the physiological outcome from the annotated sequence alone.

Nevertheless, only small scale metabolic models for *A. ferrooxidans* (Hold, Andrews and Asenjo 2009; Sepúlveda, Cortéz, Abarca, Valdecantos, Iglesias and Roa 2011), have been proven successful on describing the main aspects of its metabolism. However, these models lack the ability to capture the complex physiological characteristics, behavior and metabolic capabilities of the cell as a whole integrated system. In order to overcome this difficulty and fully explore metabolic genotype-phenotype relationships, implementation of constraint-based reconstruction and analysis (COBRA) methods for *A. ferrooxidans* was employed. The "cornerstone" of this method is represented by the genome-scale network reconstruction (GENRE) (Thiele and Palsson 2010) that is built systematically using genome annotation, "omics" data sets and legacy knowledge. Genes, proteins, reactions and metabolites that participate in the metabolism are identified, categorized and systematically interconnected, enabling a mechanistic description of metabolic physiology.
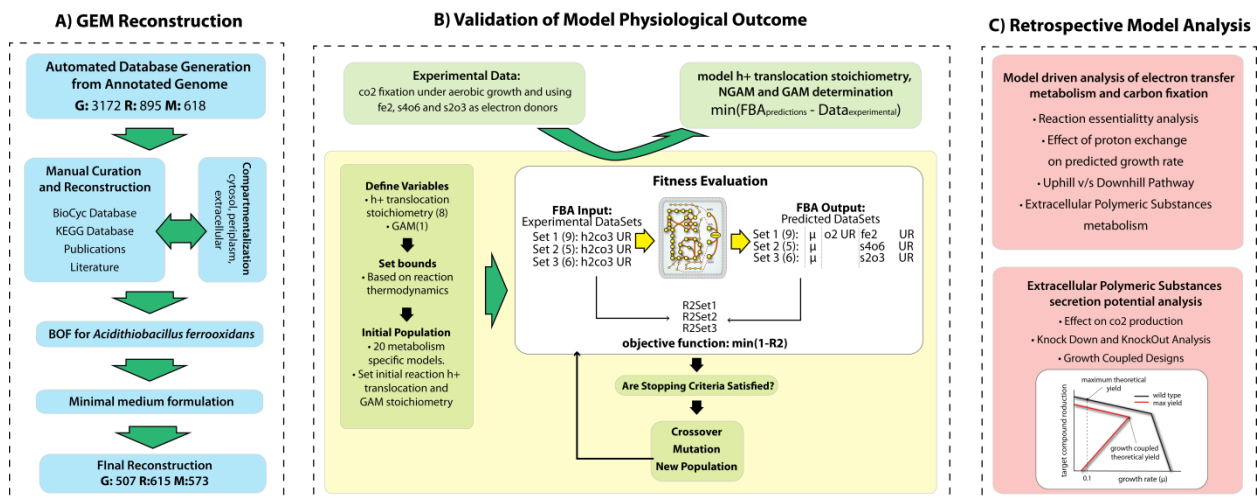
Figure 1-1: General workflow to generate, validate and further retrospective analysis of *i*MC507. (A) the workflow detailing the iterative model building procedure used to generate *i*MC507. The reconstruction process was initiated based on the annotated genome generated by PathoLogic. Manual curation and reconstruction was performed by using the Insilico Biotechnology software aided by the KEGG and MetaCyc databases. Publications and literature sources were used to refine the network content, assigning a specific confidence score to each reaction. The reconstructed network in conjunction with the BOF, were used to formulate a minimum medium for the three different metabolism studied. B) Proton translocation for 8 different membrane reaction and GAM stoichiometry were estimated using a genetic algorithm. Experimental data for growth under three different electron donors and FBA was used to decipher the model parameters that best represent cell behavior. Based on these results the *i*MC507 network was validated with physiological data for growth under ferrous ion, tetrathionate and thiosulfate. C) To study the most relevant aspects of the electron transfer metabolism and carbon fixation, a retrospective model analysis was performed by using *i*MC507. Further analysis on EPS secretion potential for growth-coupled production through knock-outs was performed.

A GENRE combined with constraints-based methods can be used to formulate mechanistic predictions of metabolic physiology that can be used in a prospective manner to elucidate new biological knowledge and understanding, as well as design and engineer the cellular metabolism (McCloskey et al. 2013). Several workflows have been

implemented to predict metabolic phenotypes, by integrating high-throughput data sets with COBRA methods (Bordbar, Monk, King and Palsson 2014b). In this work, a GENRE was reconstructed, validated and analyzed to better understand key metabolic capabilities of *A. ferrooxidans* ATCC 23270 (see Figure 1-1). A systematic workflow for the elucidation of proton translocation stoichiometry of key enzymes based on physiological data was implemented under three different aerobic chemolithoautotrophic conditions (i.e., by using ferrous ion, tetrathionate and thiosulfate as electron donors). The electron transport system and central carbon metabolism was characterized and studied. Furthermore, in order to determine the potential production of extracellular polymeric substances (EPS) in *A. ferrooxidans*, a retrospective model-driven analysis was performed.

1.3 Materials and Methods

1.3.1 Network reconstruction process

The reconstruction software Insilico Discovery™, version 3.3 (Insilico Biotechnology AG, Stuttgart, Germany), was used to build and curate the first draft of the *A. ferrooxidans* genome-scale network. First, the specific Pathway/Genome Database (PGDB) for *A. ferrooxidans* ATCC 23270, version 14.1, was downloaded from BioCyc (Caspi et al. 2012), where based on the annotated genome sequence the PathoLogic program (Dale, Popescu and Karp 2010; Karp et al. 2010) automatically infer metabolic pathways, assign enzymes to reactions they catalyzed, and infer transport reactions among others capabilities. The draft genome accounts for 3217 protein-coding genes, of which 64% were assigned a putative function(Valdes, Pedroso, Quatrini, Dodson, Tettelin, Blake, Eisen and Holmes 2008). Next, in order to connect the database with the reconstruction platform, four different files were generated by Insilico Biotechnology AG (Stuttgart, Germany) from the original *A. ferrooxidans* ATCC 23270 PGDB. The generated files specify different kinds of metabolic interaction levels, such as, metabolites, reactions, genes and operons. At the beginning, most gene-protein-reaction (GPR) association assignments were made from the annotated genome and the model was reconstructed on a pathway basis manually. In general, the reconstruction process was implemented to minimize the number of grouped, or lumped,

reactions in the network reconstruction. During the manual curation process online database, e. g., KEGG (Kanehisa et al. 2006), MetaCyc (Caspi, Altman, Dreher, Fulcher, Subhraveti, Keseler, Kothari, Krummenacker, Latendresse, Mueller, Ong, Paley, Pujar, Shearer, Travers, Weerasinghe, Zhang and Karp 2012) and BRENDA (Schomburg et al. 2002), were extensively used. The ORFs that encode the included proteins were integrated into the GPRs associations for the reactions in which they participate. GPR associations were also determined directly from biochemical evidence presented in journal publications and reviews. Transport reactions were added to the network from the genome annotation or from physiological data. All reactions added into the model were both elementally and charged balanced. Reaction reversibility was determined from thermodynamic considerations. Confidence scores were assigned for each reaction, based on the available evidence for its presence in the model (Thiele and Palsson 2010). Biochemically characterized enzymes received a confidence score of 4. If genetic knockout information or physiological evidence was available, a score of 3 was assigned. A score of 2 was assigned to reactions for which indirect evidence or sequence homology information was available. During gap-filling and evaluation of the network functionality some reactions were added with a confidence score of 1.

1.3.2 Generation of the Biomass Objective Function (BOF)

In order to simulate a cell that strives to maximize biomass production from available media substrates, a detailed and precise biomass reaction is needed for realistic metabolic network analysis. The BOF is a linear equation consisting of the fractional molar amounts of metabolites that constitute the dry weight content of the cell along with a growth associated maintenance (GAM) reaction to account for nonmetabolic growth activity (e.g., energy required for macromolecular synthesis), represented in the BOF as ATP hydrolysis reaction. Aside from the BOF, a non-growth associated maintenance (NGAM) reaction (i.e. an independent ATP hydrolysis reaction) was used as an energy "drain" on the system during the linear programming calculations, and accounts for nongrowth cellular activities (e.g., turgor pressure). The *A. ferrooxidans* BOF was formulated according to the procedures described elsewhere (Thiele and Palsson 2010). Biosynthetic macromolecule fractional content was obtained from experimentally reported values when possible, if not, the corresponding fractional

content was estimated as reported in *i*AF1260 (Feist et al. 2007). Experimental values for proteins (60 %) (Kuenen 1979) and inorganic ions (0.3 %) (Sublette 1988) were found in literature. Assuming a linear behavior between genome size and DNA cell fractional content in different organisms, the corresponding DNA fractional content for *A. ferrooxidans* was estimated from a correlation specified in Table 1-1. RNA fractional content was estimated according to the genome composition (see Table A1-2). Glycogen, murein, LPS, phospholipids and the soluble pool fractional content was estimated according to *i*AF1260.

Table 1-1: DNA fractional content and A. ferrooxidans estimation from different species

Specie	Genome size[bp]	Genome size[Mbp]	mg DNA/mg DW
Escherichia coli K-12[3]	4639221	4.6	0.03
Mannheimia succiniciproducens[4]	2314078	2.3	0.03
Salmonella typhimurium[5]	4951371	5.0	0.04
Shewanella oneidensis[6]	5102455	5.1	0.05
Vibrio vulnificus[7]	3281945	3.3	0.03
Corynebacterium glutamicum ATCC 13032[8]	3282708	3.3	0.01
Acidithiobacillus ferrooxidans ATCC 23270[9]	2982397	2.982397	0.023

The relative fraction of amino acids was obtained from previous studies (Sublette 1988). The nucleotide content for DNA and RNA was estimated based on the genome composition. The relative fraction of fatty acids (Mykytczuk et al. 2010), phospholipids (Shively and Benson 1967), LPS (Mayer et al. 1989) and Inorganic Ions (Sublette 1988) was taken from experimental data reported for *Acidithiobacillus*. EPS production was modeled as a biomass independent reaction. Fractional content of precursor was obtained from experimental results (Harneit et al. 2006). The GAM reaction stoichiometry and the NGAM flux were estimated by using a genetic algorithm and previously reported experimental results. A detailed description of the *i*MC507's biomass reactions is depicted in Table A1-1.

1.3.3 Modeling Simulations

In order to mathematically represent the reconstructed metabolic network, a stoichiometric matrix (S_{mxn}) was generated, where m is the number of metabolites and n is the number of reactions. Each entry S_{ij}, represents the stoichiometric coefficient for the corresponding *i*th metabolite in the *j*th reaction. More details can be found elsewhere (Thiele and Palsson 2010). Flux Balance Analysis (FBA) (Orth et al. 2010) was utilized for predicting growth and analyzing the reaction flux (v_j) through the metabolic network. Based on the stoichiometry matrix, a linear programming (LP) problem is solved by maximizing the BOF reaction (v_{BOF}) under steady-state criteria. Additionally, reactions are constrainted by setting an upper (**ub**) and lower bound (**lb**), which define the maximum and minimum allowable flux of the reactions. The general LP problem can be represented as:

$$\max \; v_{BOF}$$

$$s.t. \quad \sum_{j \in N} S_{ij} v_j = 0 \qquad \forall i \in M$$

$$lb_j \; \leq \; v_j \; \leq \; ub_j \qquad \forall j \in N$$

$$M := Set\ of\ metabolites\ in\ the\ network$$
$$N := Set\ of\ reactions\ in\ the\ network$$

For reversible reactions and for reactions containing metabolites present in the extracellular space the lb and ub constraints were set to -1000 and 1000, respectively. For irreversible reactions and reactions containing metabolites that are not in the medium, meaning that the metabolite could leave the cell but not enter the system, lb and ub constraints were set to 0 and 1000 respectively. Based on the inorganic electron donor, three different *A. ferrooxidans's* aerobic chemolithoautotrophic metabolisms were studied. Specifically for ferrous ion, tetrathionate and thiosulfate, by using as a unique carbon source carbonic acid (h2co3). Flux variability analysis (FVA) was used to find the minimum and maximum flux for reactions in the network, while maintaining a predefined

flux state of the network. More details about the FVA method can be found elsewhere (Mahadevan and Schilling 2003). In this work, FVA was used to find the minimum and maximum flux through each reaction while supporting 100 % of the maximal growth rate. Linear programing calculations were performed using Insilico Discovery™, version 3.3 (Insilico Biotechnology AG, Stuttgart, Germany) and the MATLAB® version 8.1.0.604 (The MathWorks Inc., Natick, MA) linked to the COBRA Toolbox 2.0 (Schellenberger et al. 2011). The linear programing package GUROBI version 5.5.0 (Gurobi Otimization Inc., Houston, TX) was used as a solver.

1.3.4 Minimal medium formulation

A synthetic minimal medium was determined based on the biomass composition specifically determined in the BOF. Biomass constituent were grouped in two different sets. The first group correspond to inorganic ions. By assuming that inorganic ions do not impose a growth restriction, the corresponding exchange reactions were allowed to freely enter and leave the network by setting lb and ub to -1000 and 1000, respectively. For carbon based biomass constituents, a manual gap filling procedure was performed. Biomass components were sequentially added to the BOF individually and further FBA was performed for BOF maximization. If the optimization leads to a positive flux through the biomass reaction, a subsequent component was added to the BOF and simulation was re-ran. For optimizations resulted in no flux through the biomass reaction, the network was updated by adding the needed reactions able to sustain growth. This process was repeated until all biomass constituents were added to the BOF. Simulations were performed under aerobic chemolithoautotrophic conditions (external oxygen reaction exchange lb and ub were set to -1000 and 1000) and using h2co3 as a unique carbon source for the three major electron donors: fe2, tetrathionate (ttton) and thiosulfate (tsul). In order to avoid operation of the rusticianin complex for tetrathionate and thiosulfate simulations, CYT2 reaction bounds were set to 0. GAM and NGAM were not considered for the gap filling procedure and determination of minimal media. Three different minimal synthetic media were generated according to the specific electron donor related to the corresponding metabolism. Metabolism specific simulations were performed by using the previously determined media conditions, which define *A. ferrooxidans* aerobic chemolithoautotrophic growth, using 3 three different electron

14

donors. The metabolisms studied were abbreviated based according to the specific electron donor: specifically for aerobic ferrous ion (FIM), tetrathionate (TTM) and thiosulfate (TSM) metabolism

1.3.5 Genetic Algorithm for proton translocation stoichiometry estimation and sensitivity analysis

The proton translocation stoichiometry for ATPS5rpp, NADHI, CYTAA32, CYTAA31, CYTRED, CYTBC1, CYTBO3, CYTBD (see Figure 1-8), the GAM reaction stoichiometry and NGAM flux were estimated based on the analysis outlined in Figure 1-1. A genetic algorithm was implemented in order to adjust the model parameters based on previously reported experimental results (Boon 1996; Eccleston and Kelly 1978; Hold, Andrews and Asenjo 2009) using three different electron donors (i.e. fe2, ttton, and tsul) under aerobic chemolithoautotrophic metabolic conditions. In total 20 different experimental points specifying the growth rate, h2co3 uptake rate, o2 uptake rate (only for FIM), and the corresponding final electron donor (fe2, ttton or tsul) were used for simulations (see Table 1-2, Table 1-3 and Table 1-4). For the proton translocation reactions, upper and lower bounds on the number of allowable protons able to cross the periplasmic membrane were estimated.

These bounds were estimated based on Mitchell's chemiosmotic theory and also obtained from experimental data (Cox et al. 1979) (see

Table 1-5, Table 1-6 and

Table 1-7 for more details and specific calculations) . Bounds on GAM were set according to reasonable reported values regarding similar analysis. Furthermore, to reduce the number of unknown variables , the NGAM was set as 2.5% of the GAM (Feist et al. 2006). In order to initialize the genetic algorithm, first, 20 different metabolism-specific models were generated by setting the corresponding metabolism constraints and the specific experimental h2co3 uptake rate. Second, the corresponding reaction to evaluate, and initial values for proton translocation and GAM reactions stoichiometry were set for each one of the 20 models. Initial parameter lower and upper bounds for genetic algorithm initialization were set randomly inbetween the

corresponding minimum lower and maximum upper allowable bound space. Third, genetic algorithm performed variables were defined, such as: crossover function, mutation function and stopping criteria. After models and genetic algorithm initialization, the algorithm started.

Table 1-2: ferrous ion experimental data obtain from (Boon 1996)

μ [h⁻¹]	fe2 uptake rate [mmol*gDW⁻¹*h⁻¹]	o2 uptake rate [mmol*gDW⁻¹*h⁻¹]	co2 consumption rate [mmol*gDW⁻¹*h⁻¹]
0.09	297	65.0	3.9
0.08	277	58.4	3.4
0.06	204	48.7	2.3
0.05	172	43.8	2.0
0.04	160	39.0	1.9
0.04	147	33.1	1.6
0.03	115	27.3	1.3
0.02	79	17.5	0.6
0.01	50	12.9	0.4

Table 1-3: Tetrathionate experimental data obtain from (Eccleston and Kelly 1978)

μ [h⁻¹]	s4o6 uptake rate [mmol*gDW⁻¹*h⁻¹]	co2 consumption rate [mmol*gDW⁻¹*h⁻¹]
0.02	2.6	0.720
0.029	3.5	1.044
0.037	4.2	1.332
0.041	4.45	1.476
0.063	6	2.268

Table 1-4: Thiosulfate experimental data obtain from (Eccleston and Kelly 1978)

μ [h⁻¹]	s2o3 uptake rate [mmol*gDW⁻¹*h⁻¹]	co2 consumption rate [mmol*gDW⁻¹*h⁻¹]
0.026	4.5	0.94
0.039	6.7	1.40
0.051	7.6	1.84
0.065	9.25	2.34
0.079	11	2.84
0.125	15.6	4.50

Table 1-5: Half reaction (reduction) potential value involved in A. ferrooxidans electron transfer membrane reactions

F [kj·mol⁻¹·V⁻¹] = 96.485		ΔG	
Half reaction (reduction)		[mV]	[kJ/mol]
$2H^+ + 0.5\ O_2 + 2\ e^-$	-> h2o	1120	108
$2\ fe3 + 2\ e^-$	-> 2 fe2	770	74
adp + pi + h	-> atp + h2o	466	45
etpcycAox + 2 e⁻	-> etpcycArd	285	27
q8 + 2 h + 2 e⁻	-> q8h2	110	11
$s_4o_6^{-2} + 2\ e^-$	-> $2\ s_2o_3^{-2}$	80	8
nad + 2 h + 2 e⁻	-> nadh + h	-320	-31

For each round, FBA was utilized for optimizing growth and quantitative phenotypic evaluation. Fitness was calculated by comparing FBA predictions with previously reported experimental results, specifically for o2, fe2, ttton and tsul uptake rate and growth rate, for each one of the 20 models. The coefficient of determination (R2) was evaluated for each dataset (R2Set1 (fe2), R2Set2 (ttton) and R2Set3 (tsul)), and the average (R2m) was used for fitness evaluation. The objective function for the genetic algorithm was the minimization of (1-R2m). MATLAB® version 8.1.0.604 (The MathWorks Inc., Natick, MA) genetic algorithm was used for all simulations. Once proton translocation, GAM and NGAM stoichiometry were estimated, a metabolism specific sensitivity analysis for each estimated variable was performed. For each metabolism, the estimated parameters were independently varied from the optimal values. According to the variation, FBA was used to calculate the corresponding R2Set1, R2Set2 and R2Set3 deviation, observed from the FBA predictions and experimental results.

Table 1-6: Thermodynamically calculated min and max bounds of potential proton translocation reactions involved in A. ferrooxidans electron transfer membrane reactions, based on (Cox, Nicholls and Ingledew 1979)

pH Gradient	External pH	Δμ_H+ [mV]	ATPS5rpp nH⁺/nATPmin	CYTAA31 nH+/2e-max	NADHI nH+/2e-min	CYTBC1 nH+/2e-min	CYTAA32 nH+/2e-max	CYTRED nH+/2e-max	CYTBD nH+/2e-max	CYTBO3 nH+/2e-max
-5.35	0.94	261	1.8	1.3	1.65	2.53	3.20	0.67	3.87	3.87
-4.41	1.97	258	1.8	1.4	1.67	2.56	3.24	0.68	3.92	3.92
-3.53	2.93	235	2.0	1.5	1.83	2.81	3.56	0.75	4.30	4.30
-2.59	3.95	170	2.7	2.1	2.53	3.89	4.92	1.03	5.95	5.95
-0.72	6.00	131	3.6	2.7	3.29	5.05	6.39	1.34	7.73	7.73
0.22	7.03	98	4.8	3.6	4.39	6.73	8.52	1.79	10.30	10.30
1.08	7.96	94	5.0	3.7	4.59	7.04	8.91	1.87	10.78	10.78

Table 1-7: Proton translocation lower and upper bound used for genetic algorithm parameter estimation.

reaction name	reaction formula	ΔEh [mV]	x_nTh bounds pH 2 lb	ub	x_0
ATPS5rpp[18]	pi[c] + adp[c] + x_1Th[p] -> h2o[c] + atp[c] + x_1Th[c]	466	3.3 [18]	5.0 [18]	4.15
CYTAA31	2 fe2[e] + 0.5 o2[c] + 2 h[c] + x_2Th[c] -> 2 fe3[e] + h2o[c] + x_2Th[p]	350	0.0	1.4 ‡	0.7
NADHI[†]	nad[c] + q8h2[c] + x_3Th[p] -> h[c] + nadh[c] + q8[c] + x_3Th[c]	-430	1.7 ‡	5.0	3.35
CYTBC1	2 fe2[e] + q8[c] + 2 h[c] + x_4Th[p] -> 2 fe3[e] + q8h2[c] + x_4Th[c]	-660	2.6 ‡	5.0	3.8
CYTAA32	0.5 o2[c] + etpcycArd[p] + x_5Th[c] -> h2o[c] + etpcycAox[p] + x_5Th[p]	835	0.0	3.2 ‡	1.6
CYTRED	q8h2[c] + etpcycAox[p] + x_6Th[c] -> q8[c] + etpcycArd[p] + x_6Th[p]	-175	0.0	0.7 ‡	2.85
CYTBD	q8h2[c] + 0.5 o2[c] + x_7Th[c] -> h2o[c] + q8[c] + x_7Th[p]	1010	0.0	3.9 ‡	1.95
CYTBO3	q8h2[c] + 0.5 o2[c] + x_8Th[c] -> h2o[c] + q8[c] + x_8Th[p]	1010	0.0	3.9 ‡	1.95
GAM	Afe_biomass_mc506_WT_139p0M		0.0	140.0	70
NGAM	h2o[c] + atp[c] -> h[c] + pi[c] + adp[c]		0.0	4.0	2

NGAM was estimated as 2.5% of GAM(Feist, Scholten, Palsson, Brockman and Ideker 2006). Translocated protons and thermodynamically determined bound were denoted as x_nTh and ‡, respectively. ATP synthase (ATPS5rpp) bounds were set according to previous work(Ferguson and Ingledew 2008).

1.3.6 Reaction essentiality analysis

Reaction essentiality analysis was performed for FIM, TTM and TSM. The analysis consists of the sequential independent reaction removal from the model, followed by FBA simulations for growth maximization. For each simulation the h2co3 uptake rate was set to 2.34 mmol/gDW/h. Reactions were removed by setting the corresponding lb and ub to 0. Simulations predicting growth rates higher than 0, were defined as non-lethal reaction knock-outs. Lethal knock-outs were clustered in terms of subsystems and further analysis was performed.

1.3.7 Reaction knock-down simulations

Reaction knock-down simulations were performed only for aerobic chemolithoautotrophic FIM. The h2co3 uptake rate for each simulation was set to 2 mmol/gDW/h. fe2 and o2 were freely allowed to enter and leave the cell. In order to analyze the growth-rate and EPS production due to metabolic co2 production and independent reaction flux constriction, 12 different simulations were calculated for each of the reactions in the model. Those conditions where defined by two parameters. First, each reaction was allowed to proceed by constraining the optimal flux at 0 %, 25 %, 50

%, 75 % and 100 %. When constraining reaction flux at 0%, all reactions were freely allowed to proceed according to the corresponding directionality. On the contrary, reaction fluxes constrained at 100 %, mean a knock-out. The others constraining percentages represent the knock-down ratio from the optimum. Additionally, a co2 transport was incorporated and constrained in the same manner. By constraining this transport reaction, metabolically produced co2 was allowed to leave the cell at 0 %, 50 % and 100 %. FBA was used to maximize BOF, and for each reaction and condition, the corresponding growth rate and EPS production rate were saved.

1.3.8 Reaction knock-out simulations

EPS production in *A. ferrooxidans* is crucial for the bioleaching process. It has been demonstrated that EPS activation in *A. ferrooxidans*, significantly increases the pyrite bioleaching capacity (Gehrke et al. 1998). EPS aids the process, by mediating the bacterial adhesion to the sulfide mineral surface. And by concentrating ferric ion in the mineral-microorganism interface by complexation with uronic acids or the EPS residues, allowing the oxidative attack on the sulfur to take place (Sand and Gehrke 2006). Based on the genome-scale reconstruction, model-driven growth-coupled designs through reactions knock-outs for EPS production were calculated. This growth-coupled designs could be difficult to achieve and may require multiple knock-outs. A number of algorithms, such as OptKnock (Burgard et al. 2003), OptGene (Patil et al. 2005), RobustKnock (Tepper and Shlomi 2010), and GDLS (Lun, Rockwell, Guido, Baym, Kelner, Berger, Galagan and Church 2009), have been proposed for designing production strains through gene knock-outs. Still, the search for knock-out phenotypes is computationally extensive, since the solution of one or more mixed-integer linear problems (MILP) are involved. This means that the time taken to solve MILPs arising from network reconstructions becomes prohibitive. Instead of formulating a MILP, an exhaustive search over all single, double, and triple knockout mutants was performed. A major benefit of this strategy is that it finds all growth-coupled designs instead of a single mutant returned by most of the MILP based algorithms. Reaction knock-out simulations were performed only for aerobic chemolithoautotrophic ferrous ion metabolism. The h2co3 uptake rate for each simulation was set to 2.34 mmol/gDW/h. fe2 and o2 were freely allowed to enter and leave the cell. Furthermore, a metabolic co2

19

transport reaction step was added to the model. Single, double and triple knock-outs were simulated. First, in order to decrease the number of simulations, a subset of reactions was determined. Specifically, all non-lethal reactions were taken into account for the simulations. In total 180 reactions were used for this analysis. Second, from this subset of reactions, all possible combination knock-outs for single, double, and triple deletions were determined. Third, for each deletion combination, FBA was used to optimize growth. Finally, in order to obtain EPS production associated to growth, knock-out combinations associated with phenotypes able to grow and simultaneously produced EPS were saved. Knock-out reactions were simulated by setting the ub and lb to 0.

1.4 Results and Discussions

The results are presented in according to Figure 1-1. First, a description on the content added to create the genome-scale *A. ferrooxidans* metabolic reconstruction and further conversion into a computational model is described. Second, by using a genetic algorithm in conjunction with experimental data, the GENRE was validated and proton translocation stoichiometry for key metabolic reactions was estimated. Third, a retrospective model-driven analysis describing *A. ferrooxidans* key metabolic capabilities and potential applications for EPS production was studied.

1.4.1 *Acidithiobacillus ferrooxidans* ATCC 23270 genome-scale metabolic network reconstruction and unique metabolic capabilities

A genome-scale metabolic reconstruction *of A. ferrooxidans* ATCC 23270, *i*MC507, was generated by performing a bottom-up reconstruction approach (Figure 1-1A). Based on the annotated genome, an automated database was generated for further curation and integration of the available genomic and biochemical data. The final reconstruction captures all major known metabolic pathways and contained 507 genes, 587 metabolic and transport reactions, and 573 nonunique metabolites, which were distributed over 42 subsytems and three different cellular compartments: extracellular, periplasm and cytoplasm. In general, the reconstruction accounts for 16 %, 69 %, and 92 % of the initial automated database genes, reactions, and metabolites respectively. Reactions were subdivided into 13 high-level functional categories based on the major

metabolic roles of the cell. Figure 1-2 shows the reaction, gene and metabolite reconstruction content in terms of the function categories. The largest number of reactions was involved in the biosynthesis of amino acids, cofactors and prosthetic groups (Figure 1-2A). In total, 79 % of the reactions have a GPR association (Figure 1-2A). The high number of transport reactions with no gene assignments, points to the fact that further work is needed to characterize the biochemical foundations involved in the transport of molecules in *A. ferrooxidans*. As shown in Figure 1-2B, almost all genes were specific for each functional category. In the case of metabolites, this is completely the opposite (Figure 1-2C), showing that almost all metabolites are shared among the different functional categories.
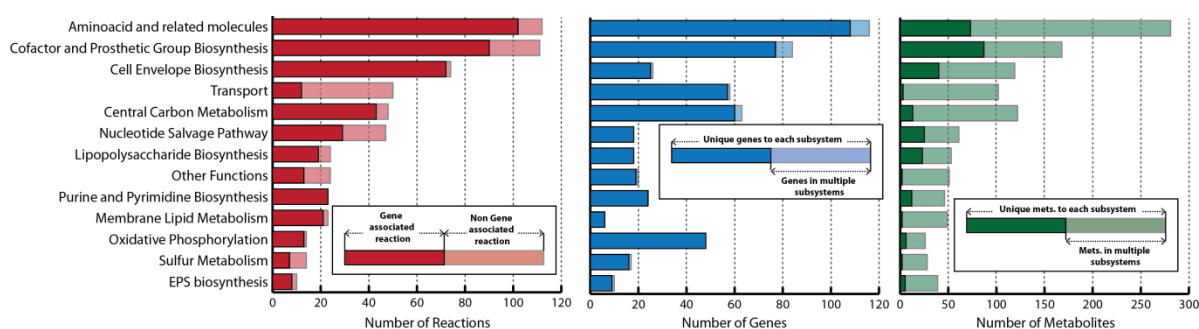
Figure 1-2: Properties of *i*MC507. (A) The number of reactions in each functional category. Non gene associated reactions are shown by the shaded portion of each bar. (B) The number of independent genes with associated reactions in each category are indicated by the lighter portion at each bar. The solid portions correspond to the number of genes unique to each category (i.e. associated only with reactions in one category). (C) The number of independent metabolites with associated reactions in each category are indicated by the lighter portion at each bar. The solid portions correspond to the number of metabolites unique to each category (i.e. associated only with reactions in one category).

Since *A. ferrooxidans* has the ability to grow under extremely low pH and on sources of electrons that yield scarcely sufficient energy for ATP synthesis and other endergonic processes vital for the cell (Ferguson and Ingledew 2008), a detailed characterization of the electron and hydrogen transfer pathways through all three cellular compartments was crucial for enabling the systems analysis of the energetic processes and understanding it's unique capabilities. An extensive effort was made to construct the most reliable model that exceeds previously published work in detail and coverage (Hold, Andrews and Asenjo 2009). Together with reconstructing the main carbon and precursor metabolism for biomass, the reconstruction process was focused on describing the energy metabolism regarding three different electron donors: ferrous ion (FIM), tetrathionate (TTM) and thiosulfate (TSM). All major metabolic subsystems

are outlined in Figure 1-3, specifying the reaction name and the corresponding gene association. For FIM, the reconstruction includes the flow of electrons from the oxidation of fe2 by Cyc2 (CYT) to reduce o2 via the aa3 complex (CYTAA31, downhill pathways), and to reduce nad via the bc1/quinone/nadh complex working in reverse (CYTBC1 and NADHI, uphill pathway). The switch point between the downhill and the uphill flow was represented at the level of the rusticyanin complex, where CyA1(CYTA2) and Cyc1(CYT1) diverge the electron flux through the uphill and downhill pathways, respectively (Quatrini et al. 2009). For the uphill pathway, individual reactions for the rusticyanin complex (CYT1, CYT2, and CYTA2), cytochrome oxidase bc1 (CYTBC1), cytochrome c oxidase (CYTAA31) and a ubiquinone nadh dehydrogenase (NADHI). Reactions CYTBC1 and NADHI, in the uphill pathway, were modeled as being energetically driven by the proton motive force (Bruscella et al. 2007; Levican et al. 2002). Nevertheless, no specific proton translocation stoichiometry was found. For reaction CYTAA31, in the downhill pathway, discrepancies arise when trying to determine whether this reaction translocates , protons from the cytoplasm into the periplasmic space (Ferguson and Ingledew 2008). For this, CYAA31 was modeled as if it were actually translocating protons through the membrane. In order to computationally estimate the missing proton translocation stoichiometry, a genetic algorithm was implemented (see Methods). Reduced inorganic sulfur compound (RISC) metabolism is of great importance for biohydrometallurgical technologies. Biologically relevant RISCs are sulfide, polysulfides, elemental sulfur, sulfite, thiosulfate and polythionates, such as tri-, tertra-, and pentathionate. All of these species are finally oxidized to sulfate, being responsible for lowering the pH value in their habitats (Rohwerder and Sand 2007). Since RICSs exist in multiple oxidation steps, from -2 to +6, many relevant enzymes, reactions steps and intermediates remain undiscovered and unconnected (Quatrini, Appia-Ayme, Denis, Jedlicki, Holmes and Bonnefoy 2009).   With the aim of characterizing already identified RISC oxidation pathways, individual reactions for elemental sulfur, tetrathionate and thiosulfate metabolism were incorporated into the model. Additionally, computational model validation for TTM and TSM was performed. RISC oxidation was modeled by incorporating the so-called Sox system. Reaction compartmentalization was based on previously reported work (Chi et al. 2007; Quatrini, Appia-Ayme, Denis, Jedlicki, Holmes and Bonnefoy 2009). Elemental sulfur was

proposed to be oxidized by a dioxygenase (SULDO) (Rohwerder and Sand 2003). The product of SULDO is sulfite, which is further oxidized to sulfate by a sulfite cytochrome c reductase (SCCR). Sulfate can be exported outside the cell by SO4tex.

The reduced cytochrome c (CycA2) is further oxidized by the cytochrome oxidase CYTAA32 (Chi, Valenzuela, Beard, Mackey, Shabanowitz, Hunt and Jerez 2007), which is different from the ferrous ion metabolism oxidase. Thiosulfate and tetrathionate metabolism were proposed to work as a cycle (Beard et al. 2011), where tetrathionate hydrolase converts tetrathionate to thiosulfate (4THASE1 or 4THASE2) (Kanao et al. 2007; Kikumoto et al. 2013). Then, by using a thiosulfate quinone oxidoreduxtase (TSQOC), thiosulfate is regenerated to tetrathionate and contributs to the quinone pool (Quatrini, Appia-Ayme, Denis, Jedlicki, Holmes and Bonnefoy 2009). TSQOC represents a switch point where electrons are diverged i) in the form of reduced ubiquinone (q8h2) to NADHI for nadh generation, described as the "RISC uphill pathway", and ii) for o2 reduction and generation of proton motif force via two different electron transfer systems through the "RISC downhill pathway". The first system is formed by a quinone dependent bc1 complex (CYTRED) catalyzing cytochrome c (CycA2) reduction. Further, CycA2 is oxidized by the cytochrome oxidase CYTAA32 which catalyzes direct o2 reduction into water. The second system able to directly reduce o2 into water might be accomplished by two different cytochrome oxidases, CYTBD or CYTBO2 (Quatrini, Appia-Ayme, Denis, Jedlicki, Holmes and Bonnefoy 2009). Using the implemented genetic algorithm, proton translocation reaction stoichiometry for CYTAA32, CYTRED and CYTBO3 was determined. Results are shown in the following section.

The first step of co2 consumption in *A. ferrooxidans* was modeled according to co2 aqueous equilibrium. When atmospheric co2 is dissolved in water, three different major molecular species are formed: carbonic acid (h2co3), bicarbonate (hco3) and carbonate (co3) (Valdés et al. 2010). Below pH 4 the only specie is h2co3, and at pH 6.5 the predominant specie is hco3. Since *A. ferrooxidans* periplasmic pH is equal to 2 and cytoplasmic pH equal to 6.5, the transport reaction (HCO3tpp) that moves dissolved co2 through the periplasmic membrane was modeled by simulating the chemical h2co3 conversion to hco3 due to the pH aqueous equilibrium. Once hco3 enters the cell, a carbonic anhydrase (HCO3E) is used to generate co2 which is fixed via the Calvin Cycle

(Gale and Beck 1967; Valdes et al. 2003) using energy and reducing power derived from the oxidation of RISCs and ferrous ions.

A key reaction in the Calvin Cycle was the ribulose-bisphosphate carboxylase/oxygenase (abbreviated RUBISCO in the reconstruction), which enables co2 incorporation into the chemolithoautotrophic metabolism (Esparza, Cardenas, Bowien, Jedlicki and Holmes 2010), connecting the Calvin Cycle with the Glycolysis/Gluconeogenesis pathway. In order to feed RUBISCO and ensure rapid conversion of cytosolic hco3 to co2 at concentrations that support optimal RUBISCO activity, a carbonic anhydrase (HCO3E in the model) that catalyzes the reversible hydration of co2 was added to the model. Fixed carbon through RUBISCO can be channeled for glycogen biosynthesis and to provide precursors for anabolic reaction. In total 90 % of all reactions in the central carbon metabolism (TCA, Calvin Cycle, Pentose Phosphate, Glycolysis/Gluconeogenesis, pyruvate, glyoxylate and anaplerotic metabolism) were associated with at least one GPR association. The TCA cycle was modeled as incomplete(Valdes, Pedroso, Quatrini, Dodson, Tettelin, Blake, Eisen and Holmes 2008), as has been previously described in a number of obligate autotrophic bacteria and archaea(Wood et al. 2004). Nevertheless, these pathways were incorporated based only on genome annotation (i.e. confidence score equal to 2), meaning that further biochemical characterization is needed. EPS production, for mediating bacterial attachment, was modeled in the same way the as BOF reaction. The EPS reaction is a linear equation consisting of the fractional molar amounts of metabolites that constitute the EPS, which have been proposed to be formed via the Leloir pathway (Barreto et al. 2005; Harneit, Göksel, Kock, Klock, Gehrke and Sand 2006). Specifically, eight out of ten reactions labeled as EPS biosynthesis, were linked with at least one GPR association.
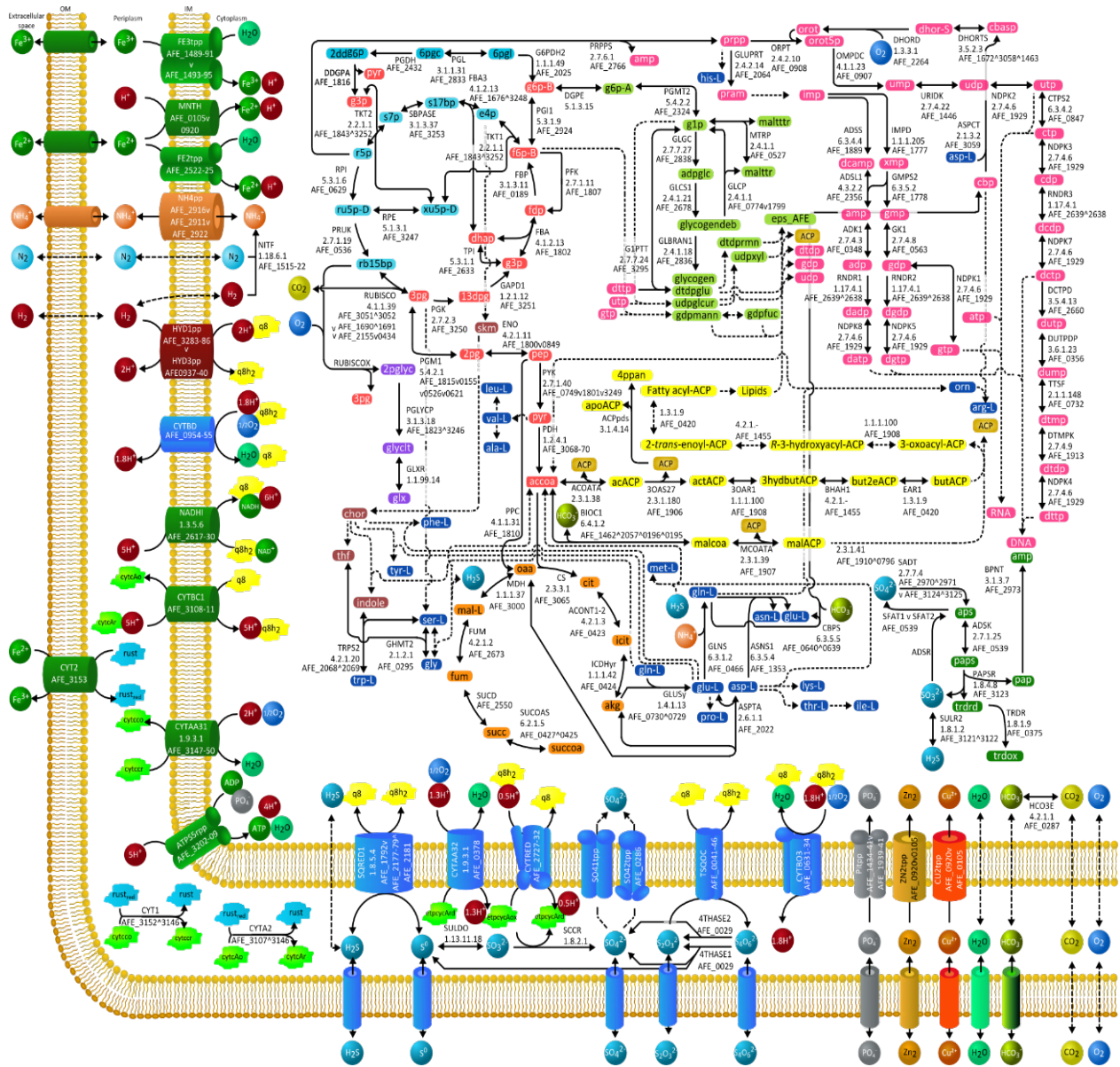
Figure 1-3: Schematic representation of the main metabolic pathways of Acidithiobacillus ferrooxidans ATCC 23270 metabolism. Continuous line depicts one reaction, which is associated with relevant information such as enzyme name abbreviation, gene ID and E. C. number; dashed line indicates that more than one reaction are involved in the pathway. Hyphen (-) indicates a consecutive series of genes (e.g. AFE_0001-3 means AFE_0001 and AFE_0002 and AFE_0003). Logical operators were used: ^ indicates "AND"; v indicates "OR". Each color represents one pathway. In the cytoplasm: Red = Glycolisis/Gluconeogenesis; ligth blue = Pentose Phosphate Pathway; orange = incomplete Citric Acid Cycle; blue = Aminoacid Biosynthesis; brown = Chorismate Biosynthesis; purple = Alternate Carbon Metabolism; yellow = Membrane Lipid Metabolism/Fatty acid biosynthesis; pink = Nucleotide Salvage Pathway/Purine and Pyrimidine Biosynthesis; light green = Glycogen Metabolism; green = Cysteine Metabolism/Sulfur Metabolism. In the membranes/periplasm: green = Oxidative Phosphorylation; dark red = Hydrogen and Formate Utilization; ligth blue = Sulfur Metabolism. Top-left cycle = Calvin-Benson-Bassham Cycle. OM = Outer membrane; IM = Internal membrane. Compounds 3-oxoacyl-ACP, R-3-hydroxyacyl-ACP, 2-trans-enoyl-ACP, Fatty acyl-ACP and Lipids are generic intermediates in the fatty acids biosynthesis pathway.

25

Table 1-8: Total amount of reactions and reactions with GPRs associations organized by subsystems

Sub Systems	Nr of Reactions	Reactions with Gene Associations	Reactions with No Gene Associations
Sulfur Metabolism	**14**	**7**	**7**
Arginine and Proline Metabolism	18	15	3
Threonine and Lysine Metabolism	13	13	0
Tyrosine, Tryptophan, and Phenylalanine Metabolism	18	18	0
Histidine Metabolism	11	11	0
Valine, Leucine, and Isoleucine Metabolism	16	14	2
Methionine Metabolism	17	13	4
Alanine and Aspartate Metabolism	3	3	0
Cysteine Metabolism	7	7	0
Glutamate Metabolism	2	2	0
Glycine and Serine Metabolism	5	5	0
Glutathione Metabolism	1	0	1
Nitrogen fixation	1	1	0
Aminoacid and related molecules	**112**	**102**	**10**
Folate Metabolism	7	3	4
Cofactor and Prosthetic Group Biosynthesis	**111**	**90**	**21**
Nucleotide Salvage Pathway	**47**	**29**	**18**
EPS biosynthesis	**10**	**8**	**2**
Glycerophospholipid Metabolism	**12**	**9**	**3**
Lipopolysaccharide Biosynthesis	**24**	**19**	**5**
Citric Acid Cycle	9	9	0
Calvin-Benson-Bassham Cycle	16	15	1
Pentose Phosphate Pathway	5	5	0
Glycolysis/Gluconeogenesis	11	10	1
Pyruvate Metabolism	2	2	0
Glyoxylate Metabolism	2	0	2
Anaplerotic Reactions	3	2	1
Central Carbon Metabolism	**48**	**43**	**5**
Membrane Lipid Metabolism	**23**	**21**	**2**
Cell Envelope Biosynthesis	**74**	**72**	**2**
Oxidative Phosphorylation	**14**	**13**	**1**
Transport, Inner Membrane	23	12	11
Polyphosphate Storage	2	0	2
Exchange	29	0	29
Transport, Outer Membrane Porin	25	0	25
Transport	**79**	**12**	**67**
Purine and Pyrimidine Biosynthesis	**23**	**23**	**0**
Glycogen Metabolism	8	5	3
Hydrogen and Formate Utilization	3	3	0
Alternate Carbon Metabolism	11	5	6
Unassigned	2	0	2
Other Functions	**24**	**13**	**11**
TOTAL	**615**	**461**	**154**

The number of essential reactions for growth under aerobic FIM, TSM and TTM was determined using FBA (see Methods). The results of such calculations are presented in Figure 1-4. These results have to be interpreted with caution, since calculations were based on the assumption that all enzymes are expressed. Transcriptomics or proteomic data would considerably improve the results by constraining the internal fluxes. Approximately 68% of all possible single reaction deletions would be lethal for the organism under aerobic FIM, TTM and TSM. In general,

when comparing the percentage of knock-out reactions non-essential for growth, for all studied subsystems, no considerable fluctuations were observed under FIM, TSM, and TTM.
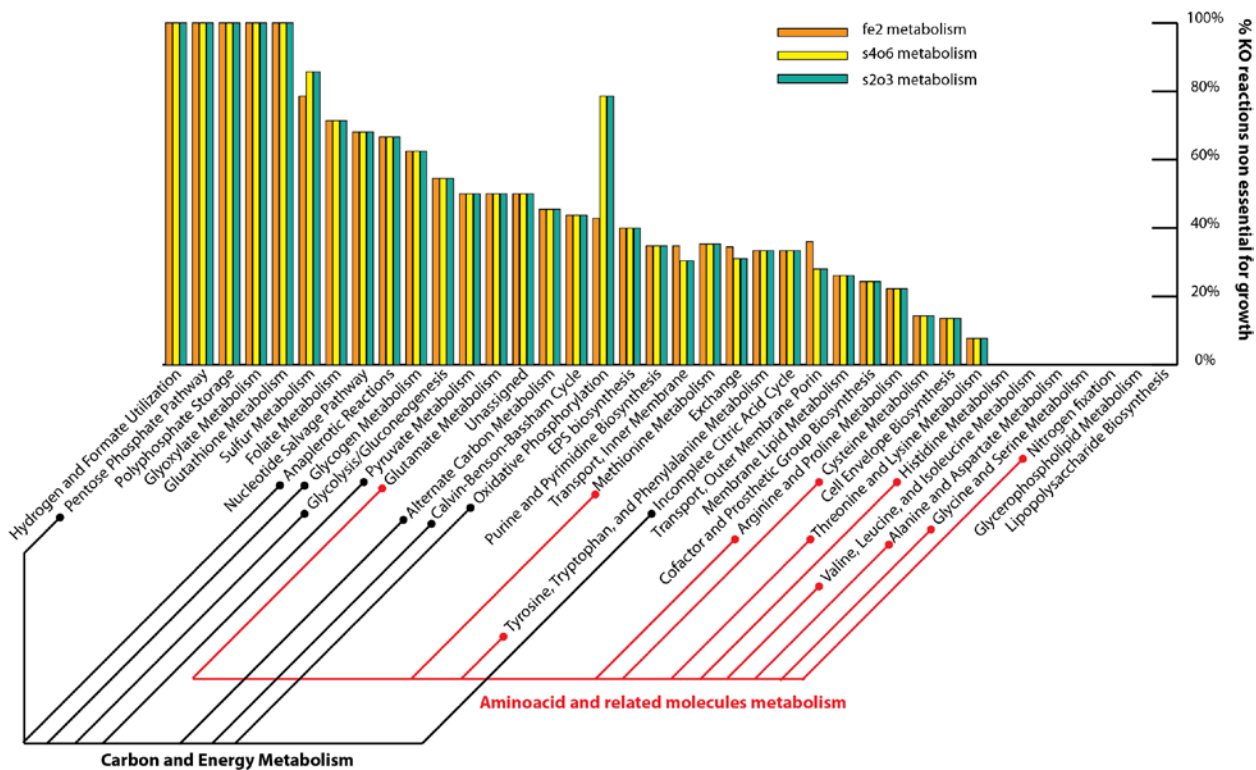
Figure 1-4: Reaction essentiality analysis. For each studied metabolism, ferrous ion (orange), tetrathionate (yellow), and thiosulfate (green) the percentage of knock-out reactions non-essential for growth in various subsystems were plotted. Specific associations with carbon and energy (black lines), and aminoacid related molecules (red lines) metabolism were specified.

The only substantial difference between these three metabolisms on reaction essentiality was observed for oxidative phosphorylation and sulfur metabolism, where in both cases the number of knock-out reactions being non-essential for growth was higher for TSM and TTM compared to FIM. Besides, higher percentage of knock-out reactions non-essential for growth are associated to Carbon and Energy metabolism compared to Amino acid and related molecules metabolism. This shows how flexible the Carbon and Energy metabolism and how rigid for Amino acid and related molecules metabolism are in *A. ferrooxidans* due to knock-out perturbations.

In general, *i*MC507 represents a database of current biochemical, genetic and genomic knowledge about *A. ferrooxidans.* Based on the available experimental evidence, a confidence score was assigned to each reaction in the metabolic network.

On completion, the GENRE had an overall average confidence of 2.02. In fact, 19% of all reactions in *A. ferrooxidans* included in *i*MC507 have been very well or well-studied, while 58% were fundamentally based on the genome annotation and 18% of all reactions reflects that no evidence is available, but the reaction is required for modeling. Future research efforts should be directed towards this latter group, which is described in Figure 1-5.

1.4.2 GEM proton translocation stoichiometry estimation and network validation

Genome-scale reconstructions need to be validated by assessing their ability to compute physiological states(Thiele and Palsson 2010). *i*MC507 was validated by comparing growth predictions with experimental values (see Supplementary Table 5). A genetic algorithm able to estimate several model parameters (such as proton translocation for 8 different periplasmic proteins and GAM stoichiometry), which minimizes the error between growth predictions and experimental values was implemented (see Methods). The genetic algorithm simultaneously fitted FBA predictions with experimental results for chemolithoautotrophic aerobic FIM, TTM and TSM conditions. Proton translocation stoichiometry for ATPS5rpp, CYTAA31, NADHI, CYTBC1, CYTAA32, CYTRED, CYTBO3 and CYTBD and, GAM stoichiometry, were determined by performing three different genetic algorithm based simulations. Since CYTBO3 and CYTBD are the same reaction (but with different GPR association), and in order to decrease prediction uncertainties, only CYTBO3 was used for all analysis and to further extrapolate the results to CYTBD.

First, for all studied reactions proton translocation stoichiometry variables were fed into the genetic algorithm for parameter estimation. A total of 100 simulations were performed and statistical results showing each parameter mean, standard deviation and violin plots are shown in Figure 1-6. According to the results, for reactions ATPS5rpp, CYTAA31, NADHI and CYTBC1, the smallest proton translocation stoichiometry standard deviation from the set was determined (approximately in the order of hundredths). A particular characteristic of this subset of reactions is that all reactions are involved in FIM. While for reactions involved in the RISC's, such as, CYTAA32, CYTRED and CYTBO3/CYTBD, standard deviations were in the order of tenths. An

R2m equal to 0.92 was calculated, meaning that each simulation was run until an acceptable numerical completion. A detailed analysis on the RISCs electron transfer metabolism showed that reactions CYTAA32 and CYTRED in conjunction performed the same function as CYTBO3.
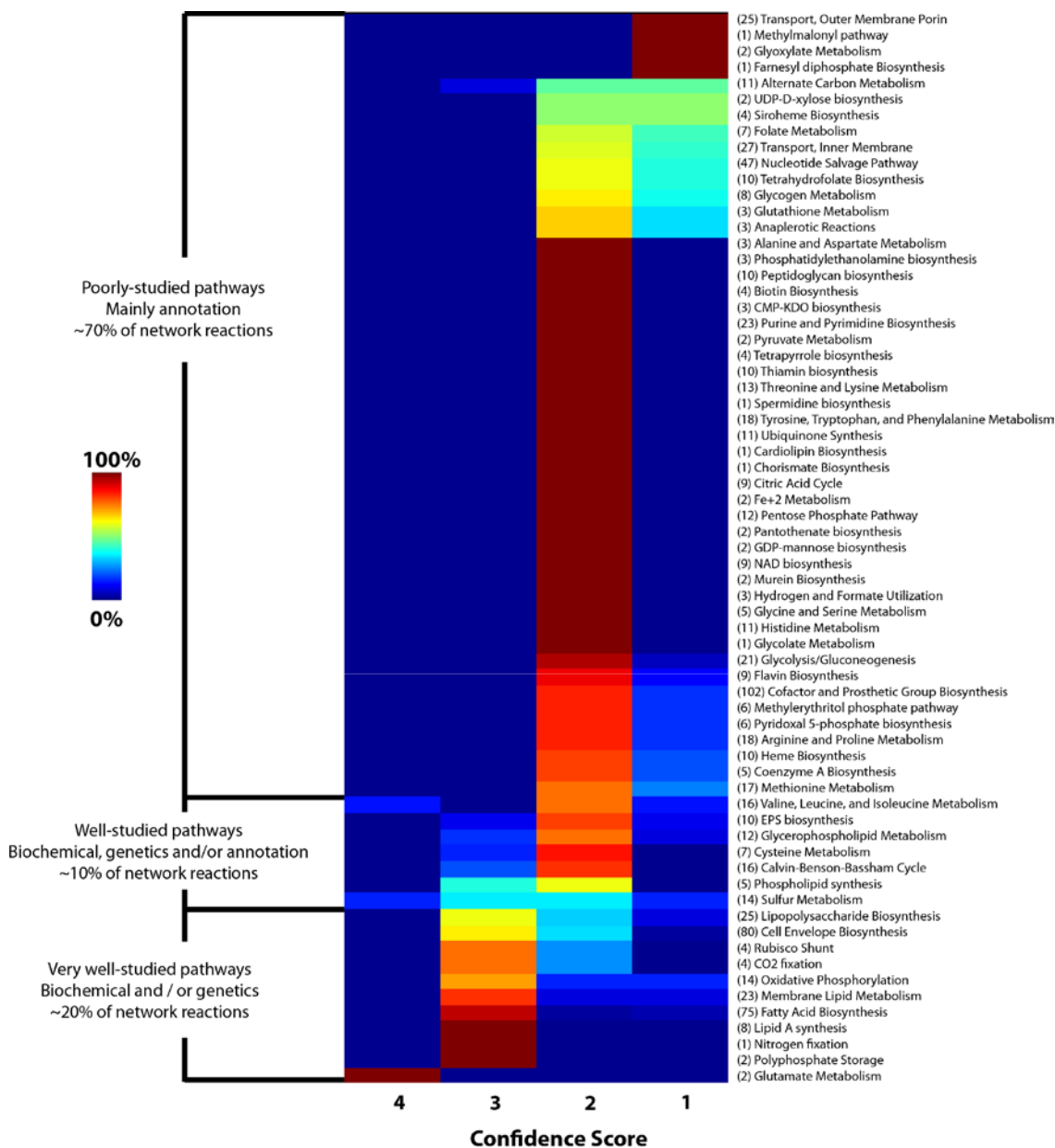
Figure 1-5: Comprehensive categorization of current available information for *A. ferrooxidans*. Subsystems and the number of reaction in each one of them are listed. Colors represent the percentage of reactions in the corresponding subsystem, organized by confidence score. The confidence score was assigned to each reaction in the model on scale from 1 to 4. A score of 4 was assigned when experimental, gene product function and biochemical reaction evidence are available; 3 represents physiological, genetic or proteomic evidence; 2 correspond only to genome annotation evidence for a gene product and its reaction(s) association; and a score of 1 shows that no evidence is available, but the reaction is required for modeling porpoises.

This network characteristic generates higher variability on the predictions when trying to estimate the parameters during the genetic algorithm procedure, by adding more degrees of freedom to the minimization problem. In order to accurately determine these proton translocation reaction stoichiometry, each system (i.e. CYTRED/CYTAA32 and CYTBO3) was independently studied. Based on the latter, the second genetic algorithm based simulation was performed similarly to the first one, excluding CYTBO3 and CYTBD from the analysis. Similarly, the third genetic algorithm based simulation was performed excluding CYTRED and CYTAA32 from the analysis and reincorporating the CYTBO3/CYTBD function. From these results, more accurate predictions for CYTRED, CYTAA32 and CYTBO3 were determined, by considerably decreasing the proton translocation stoichiometry standard deviation on the predictions. Furthermore, for the second and third simulations, no considerable changes in the parameters mean and standard deviation for reactions ATPS5rpp, CYTAA31, NADHI and CYTBC1 were observed.

Based on the previous analysis, proton translocation stoichiometry for ATPS5rpp, CYTAA31, NADHI, CYTBC1, CYTAA32, CYTRED, CYTBO3 and CYTBD was set to 5, 0, 5, 5, 1.3, 0.5, 1.8 and 1.8 respectively. GAM stoichiometry was set to 139 and NGAM to 3.48. Based on this solution, for all 7 adjusted curves (Figure 1-7, B and C), an R2m equal to 0.92 was calculated. Under the simulated conditions, iMC507 exhibited a growth rate, oxygen uptake rate and the corresponding electron donor uptake rate practically identical to the experimental results. This result shows the predictive potential and preciseness of COBRA methods for simulation of cellular outcomes by using additional constraints. In order to analyze the predicted proton translocation and GAM stoichiometry sensitivity on the specific R2Set (1:fe2, 2:tton and 3:tsul), independent perturbations were imposed for each parameter as a factor from the local optimal solution and a further R2Set was calculated by using FBA.

Figure 1-6: Genetic algorithm based analysis output summary. Proton translocation and GAM stoichiometry violin plots, mean (red cross), median (yellow square) and standard deviation for each fitted parameter in the corresponding reaction were plotted and tabulated. Furthermore, R2m calculations were incorporated. Results were given for three different genetic algorithm based simulations. 1) All parameters were fed into the model for computational estimation according to methods. Then 2) and 3) were executed similar to 1), but in these cases CYTBO3 and CYRED/CYTAA32 were excluded from the genetic algorithm based estimation, respectively. In total each simulation was run 100 times. Initial values for lower and upper bounds proton translocation and GAM reactions stoichiometry were set randomly inbetween the corresponding minimum lower and maximum upper allowable bound space.

		ATPS5rpp	CYTAA31	NADHI	CYTBC1	CYTAA32	CYTRED	CYTBO3	GAM	R2
1	mean	4.99	0.01	4.96	4.97	1.23	0.36	1.28	13.92	0.92
	stdev	0.0122	0.0069	0.0313	0.0273	0.3958	0.2361	0.6293	0.0911	0.0006
2	mean	4.983	0.009	4.967	4.974	1.344	0.460	----	13.908	0.921
	stdev	0.0179	0.0099	0.0313	0.0252	0.1824	0.1788	----	0.0784	0.0005
3	mean	4.98	0.01	4.96	4.98	----	----	1.82	13.92	0.92
	stdev	0.0155	0.0104	0.0367	0.0206	----	----	0.0452	0.0801	0.0008

R2Set variations were represented as a violin plot along the y-axe. The highest calculated R2Set for each studied parameter was reported and represented as red lines. Since CYTAA31 proton translocation stoichiometry was calculated as 0, variations were performed not as a factor from the optimal solution, but instead the direct proton

31

translocation value corresponding to the y-axis was set for overall R2Set sensitivity analysis (Figure 1-7D). The most sensitive and constrained reaction in the metabolism is ATPS5rpp. Under FIM (Figure 1-7D), when setting the proton translocation and GAM stoichiometry to 0 in the y-axis, no significant changes in the overall R2 were observed for NADHI, CYTBC1, CYTBD/CYTBO3 and GAM. Furthermore for these reactions, when increasing the parameters by approximately 100 % (factor equal to 2), no considerable decrease in the R2 was observed. Specifically for reactions NADHI and CYTBC1, when increasing the number of translocated protons, FBA predictions shows an increase in reaction CYTAA31 flux, thus calculating ferrous ion and oxygen uptake rates greater than the experimental results, displacing the R2Set ratio to 0 (see Figure 1-8 for flux visualization). For reaction CYTBD/CYTBO3, proton translocation was modeled leaving the cytoplasm. When the proton translocation stoichiometry is increased, FBA predicts that less oxygen and ferrous ions are needed to balance the cytoplasmic hydrogen in reaction CYTAA31. When GAM is increased, FBA predicted that ATP synthase (ATPS5rpp) needs to generate more ATP to fulfill the GAM and NGAM demand (see Figure 1-8 for flux visualization). Similarly, as the generation of ATP by ATPS5rpp is associated with a net flux of protons inside the cell, CYTAA31 needs to balance the cytoplasmic protons by increasing oxygen and ferrous ion uptake rates, deviating the R2Set ratio. Due to the existence of uphill and downhill pathways, electrons can be diverted for biomass production (through reaction CYTBC1) and for balancing cytoplasmic protons (through reaction CYTAA31). For TTM and TSM the same kind of behavior regarding R2Set deviations and cytoplasmic proton balance was observed. But, different reactions were involved in the process. Electrons for biomass production are taken by the NADHI reaction, and electrons for balancing the cytoplasmic protons are used in CYTBD/CYTBO3 or CYTAA32 reactions. It is worthwhile drawing attention to the fact that for TSM (Figure 1-7F), better solutions for proton translocation and GAM stoichiometry values, compared to TTM (Figure 1-7E) were found, specifically for ATPS5rpp, NADHI, CYTAA32, CYTRED, and GAM. This is shown by the displacement of the maximum R2Set value from the line equal to 1 (no changes from the local optimum). Furthermore, TSM showed more degrees of freedom in comparison to TTM, since R2Set solutions might be achieved at higher perturbations.
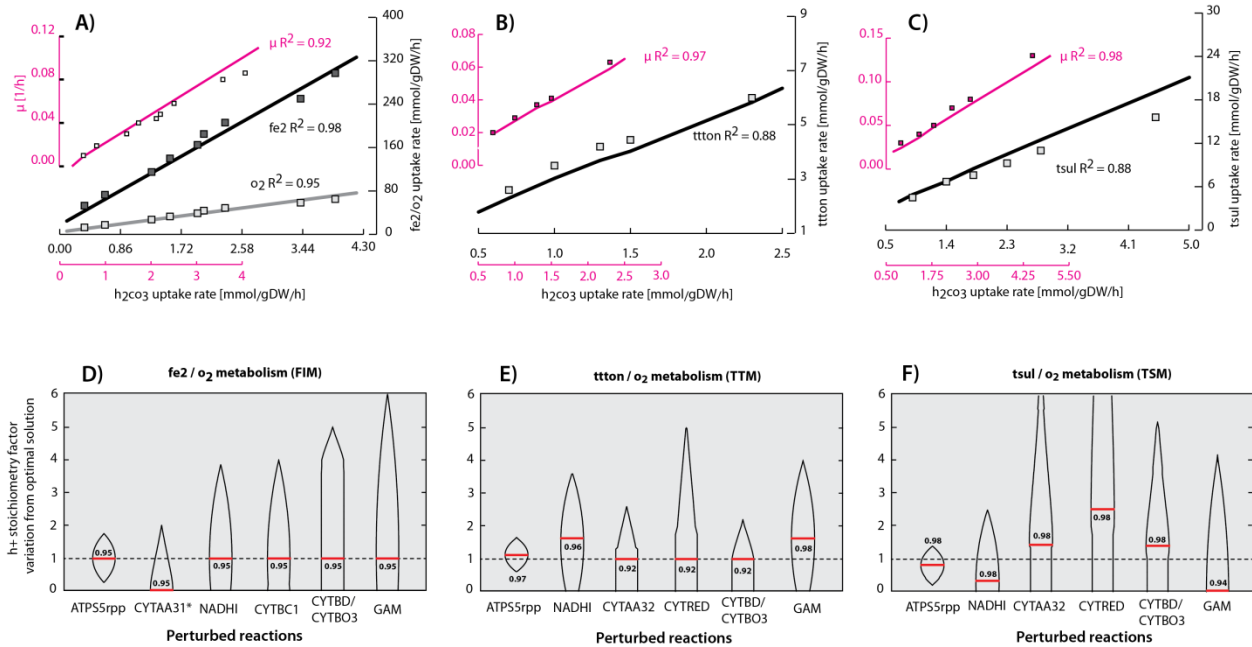
Figure 1-7: Parameter estimation and validation of the carbon fixation and electron donor pathways results. Based on the genetic algorithm model parameter estimation results, three graphs of phenotypic data that demonstrate the model-predicted growth conditions on ferrous ion (A), tetrathionate (B) and thiosulfate (C) as electron donors and using oxygen as an electron acceptor, are viable for growth conditions. The corresponding R2Set (1:fe2, 2:ttton and 3:tsul) for each predicted/experimental dataset pair are shown. A R2Set sensitivity analysis for ferrous ion (D), tetrathionate (E) and thiosulfate (F) metabolism was performed. Proton translocation predicted stoichiometry values were varied along the y axe and the corresponding R2Set variation from the optimum was calculated and plot for each one of the proton translocated stoichiometry predicted reactions as a violin graph. In this case R2Set represent the prediction error for each studied metabolism independently. The maximum calculated R2Set due to parameter variations was reported and represented as a red line.

1.4.3 *i*MC507 model driven analysis of electron transfer metabolism

The electron transfer metabolism in *A. ferrooxidans* is mostly distributed in the periplasmic membrane. Depending on the electron donor, different reactions might be involved in the electron transfer process. Three different electron transfer metabolisms were studied; FIM, TTM and TSM. In order to obtain the reaction flux distribution for each metabolism, FBA was performed by setting the h2co3 uptake rate at 2 mmol/gDW/h, and optimizing the BOF. The corresponding electron donor was allowed to freely enter the system. In the case of FIM, results indicated (Figure 1-8) that the flow of electrons is split at the rusticyanin branch point, which has actually been suggested as the balancing point for nadh and ATP requirements in the cell by adjusting the flow of electrons(Elbehti et al. 2000). According to Figure 1-8 most of the electron flow from ferrous ion goes to oxygen via cytochrome oxidase (CYTAA31) which generates proton motive force by consuming hydrogen for the oxygen reduction. Specifically 3 % of the electrons goes to the formation of nadh by the CYTBC1 and NADHI reactions. nadh is further utilized by GAPD1 (i.e. in the Calvin cycle) for carbon fixation. This result is

consistent with reality, where a previous report (Ferguson and Ingledew 2008) shows that less than 5 % of electrons go to NADHI. Furthermore, for simulations under oxygen limitation (see Figure A1-2), this behavior remains valid. It has been demonstrated that before ferric compounds start precipitating in the form of jarosite, growth on ferrous ion tends to increase the culture pH (Qiu et al. 2005; Yarzabal et al. 2004). This ferrous ion oxidation might be due to *A. ferrooxidans* metabolism. In fact, a net inward proton flux through the extracellular membrane was calculated. In more detail, predictions show that the majority of the inward proton flux into the cytoplasm goes through the ATP synthase (ATPS5rpp), specifically 67 % of all protons entering the periplasm. The remaining proton flux is mainly distributed in equal amounts between NADHI and CYTBC1. In the cytoplasm, approximately 50 % of all incoming protons are used for oxygen reduction through CYTAA31 and the rest mainly for anabolic processes. Furthermore, as shown in Figure 1-9, growth is only achievable under alkalization conditions. FVA analysis also supports this fact (see Figure A1-1, reaction Htex), where no proton outward production flux was calculated.

For TTM and TSM a completely different set of periplasmic membranes for transferring electrons into the Calvin cycle and for proton motive force generation were found to be active in the metabolic network. Furthermore, similar flux distribution results were obtained for both RISCs metabolism. TTM and TSM were outlined in Figure 1-8 describing the flux distributions in blue and light blue arrows, and discrepancies between fluxes solutions were highlighted specifically for TSM in orange. The rest of the flux solutions remain the same for both metabolisms. As shown in Figure 1-8 ttton is hydrolyzed by tetrathionate hydrolase (4THASE2) generating sulfate (so4) , which is excreted by the system, and tsul which is used by the thiosulfate-quinone oxireductase complex (TSQOC) to transfer electrons into the quinone pool. The 4THASE2/TSQOC cycle plays an important role at the beginning of the electron transfer system by first incorporating the electrons coming from tsul into the metabolism, and second transferring electrons from water to ttton, recycling the water generated by either CYTBO3/CYTBD or CYTRED in conjunction with CYTAA32. Interestingly, discrepancies showed that TSM produces less protons and sulfate through reactions Htex and SO4tex, specifically 40 % and 13 % less than TTM, respectively. Furthermore, based on

34

a sulfur molar basis, approximately 12 % less sulfur in the form of tsul is needed to sustain growth for TSM. This demonstrates that in terms of electron transfer efficiency TSM is more efficient than TTM. This is due to the amount of electrons that each electron donor is able to give on a sulfur molar basis, where tsul showed twice as many of electrons per sulfur. Furthermore, for TTM, water is needed to overcome the electron deficiency in order to sustain growth. Specifically for TTM, 70 % of all incoming electrons from water goes to TSQCO via thiosulfate, the rest are excreted in the form of sulfate. Basically the tetrathionate/thiosulfate/sulfate complex worked as a shuttle electron transfer system to deliver electrons from water to the quinone pool. At this point, the rest of the network behaves similarly for TSM and TTM, where electrons are diverged analogous to the FIM, to a "RISC uphill pathway" through NADHI, and to a "RISC downhill pathway" through CYTBO3/CYTBD or CYTRED in conjunction with CYTAA32. Due to the existence of a FBA alternate optimum, for the "RISC downhill pathway" different cytochrome oxidase combinations might worked as well as CYTBO3/CYTBD. By knocking-out the CYBO3/CYTBD reactions, the parallel solution regarding the CYTRED/CYTAA32 system was calculated and outlined in Figure 1-8 According to the flux distribution results both systems worked similarly. Furthermore, the total amount of proton translocation stoichiometry of both systems is the same (i.e. 1.8 protons translocating through the periplasmic membrane). Approximately 10 % of all incoming electrons from TSQCO go to NADHI. The rest, just like the FIM are used to balance the cytoplasmic protons and generate proton motive force through the cytochrome oxidases. It is well known that during growth *A. ferrooxidans* acidifies the media when using RISCs compounds as electron donors (Rohwerder and Sand 2007). This fact is represented by predicting a flux of protons outside the external membrane through the transport reaction Htex under TSM and TTM (Figure 1-8).
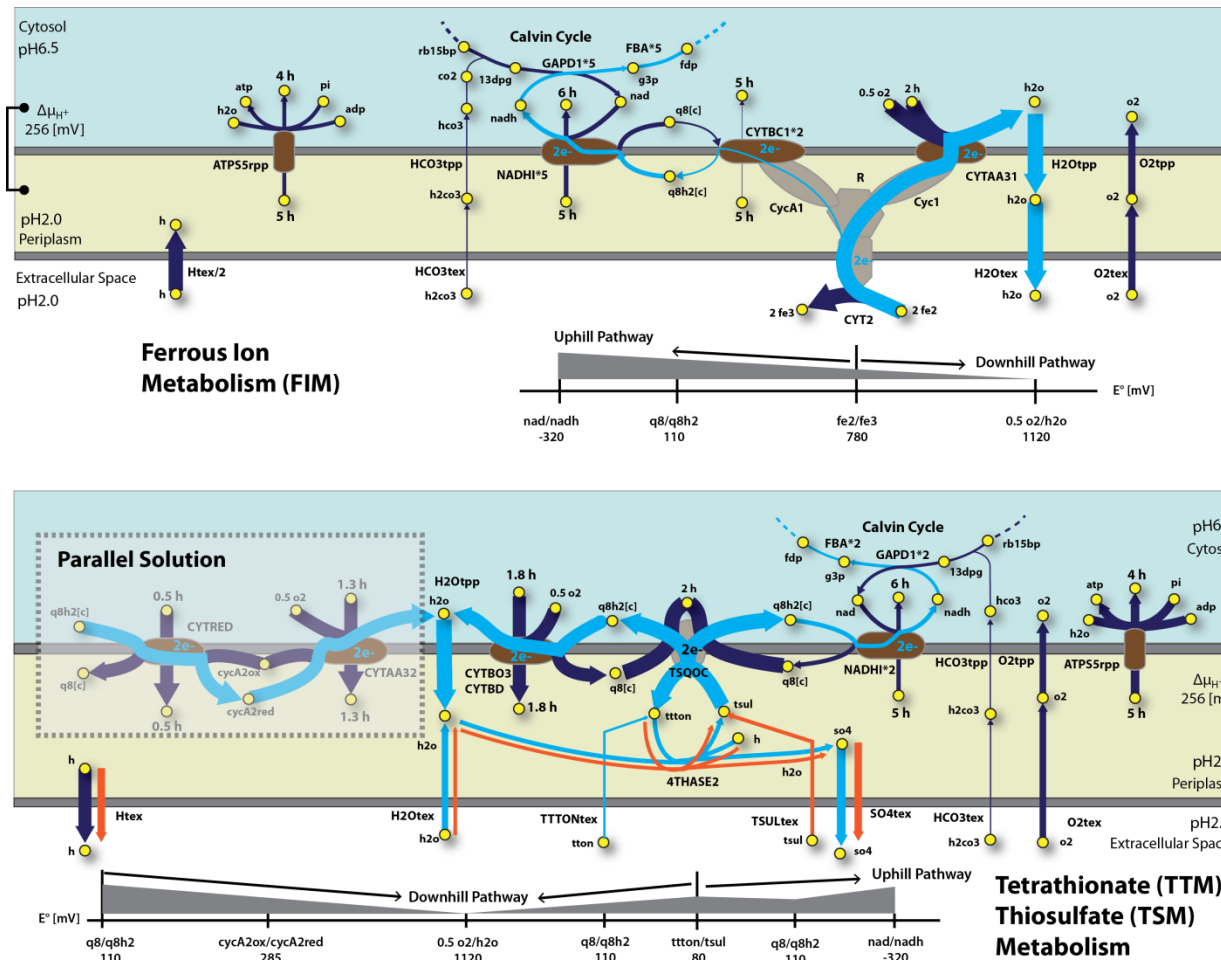
Figure 1-8: Electron transfer flux distribution predicted for ferrous ion, tetrathionate and thiosulfate metabolism. A flux map illustrating the electron transfer metabolism for ferrous ion (FIM), tetrathionate (TTM) and thiosulfate (TSM) metabolism. Reactions are specify in capital letters, while metabolites in lowercase letters and yellow circles. Arrows indicate the direction of enzymatic activity and the arrow thicknesses are proportional to the flux through each reaction (a thicker arrow has a larger flux). Light blue arrows represent the electron flow in the network. Visually undetectable fluxes for being so small were increased by a corresponding factor specified next to the corresponding reaction in the map. The diagram shows the energy-conserving ion translocating reactions, each labeled with the stoichiometry of the translocated ion. Proton translocation stoichiometry predicted reactions are shown in brown. The different colored regions correspond to the tree modeled spaces: cytoplasm (white), periplasm (yellow) and the extracellular space. For each network the corresponding standard reduction potential (in mV) associated with a specific transformation was specified. The uptake of h2co3 for each simulation was constraint at 2 mmol/gDW/h, and the corresponding electron donor was allowed to freely enter the system. FBA was performed for FIM, TTM and TSM. Due to flux similarities results, TTM and TSM were plotted together. For TTM, reactions, metabolites and fluxes we potted according to the previous description. In the case of TSM, reaction fluxes different from the ones obtain in TTM, were plotted in orange. For TTM and TSM, an additional flux distribution, called as parallel solution, was plotted. This solution was obtained when constraining CYTBO3 and CYTBD ub and lb to 0.

In a different manner, FVA analysis showed that Htex feasible bounds able to sustain grow are within the positive range of the scale (see Figure A1-1), demonstrating that in order to sustain growth, the system only has to produce protons, which are generated by 4THASE. The same kind of behavior was described in Figure 1-9, where growth was only possible under acidification media conditions.
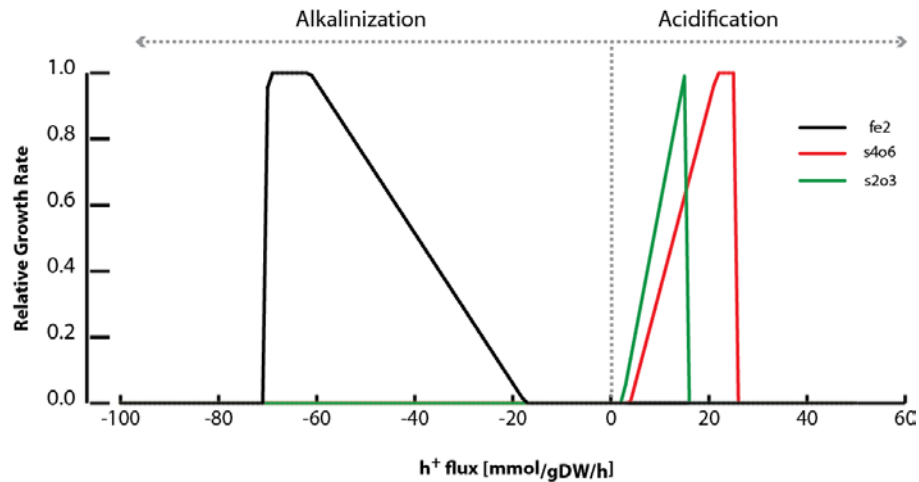
Figure 1-9: Effect of proton exchange on predicted growth rate: the proton exchange between the cell and the medium was varied from -100 to 60 mmol/gDW/h and the relative growth rate was computed under autotrophic conditions for ferrous ion (black line), tetrathionate (green line), and thiosulfate (red line) metabolism.

1.4.4 *i*MC507 model driven analysis of central carbon metabolism

Since FIM, TTM and TSM shared similar reactions for central carbon metabolism (carbon fixation, Calvin cycle, glycolysis and incomplete TCA), only FIM was studied in detail. For this, different FBA simulations were performed to quantitatively describe the fluxes involved in central carbon metabolism and study the optimum behavior of *i*M507 (Figure 1-10). For this, two different kind of analysis were performed. First, in order to decipher the internal flux distribution under different conditions, three different simulations were performed. For the first simulation the carbon uptake rate in the form of h2co3 was set to -2 mmol/gDW/h, oxygen and ferrous ion uptake rate were left unconstrained, and FBA was run to maximize the flux through BOF. Flux distributions are shown in Figure 1-10A. It is well known that *A. ferrooxidans* fixes co2 through the Calvin cycle. In order to feed RUBISCO and ensure rapid conversion of cytosolic hco3 to co2 at concentrations that support optimal RUBISCO activity, a carbonic anhydrase (HCO3E in the model) that catalyzes the reversible hydration of co2 converts almost all incoming hco3 to co2. Approximately 95 % of all hco3 goes through HCO3E to the RUBISCO reaction in the form of co2. The rest goes directly to fatty acid, arginine and proline metabolism. RUBISCO catalyzes the formation of two molecules of 3-phosphoglyceric acid (3pg) from ribulose biphosphate (rb15bp) and co2. 3pg represents the junction point between the Calvin cycle and glycolysis. Specifically, 83 % of all 3pg produced is taken by the phosphoglycerate kinase (PGK) to continue the Calvin cycle,

37

whereas the rest goes to glycolysis via phosphoglycerate mutase (PGM1), demonstrating that most of the activity is concentrated in the Calvin cycle. Through an ATP driven reaction, PGK phosphorylates 3pg to produced 13dpg. 98 % of all ATP is produced by ATP synthases and from all ATP produced approximately 16 % is used by PGK and 19 % and 33% is used by BOF and nitrogen fixation through NITF correspondingly. Next, using a nad dependent glyceraldehyde-3-phosphate dehydrogenase (GAPD1), electrons are incorporated to the metabolism by converting 13dpg to g3p. Approximately 93 % of all nadh is generated from NADHI, and GAPD1 uses 67% of all nadh. This reaction could be replaced by a nad dependent glyceraldehyde-3-phosphate dehydrogenase (GADP2), which might work as well as GAPD1, through the conversion of nadh to nadph by a nad transhydrogenase (NADTRHD). This alternate use can be observed from the FVA analysis (see Figure A1-1), where the GAPD1 range of performance without acting against biomass productions starts at 0. This means that, when GAPD1 flux is 0, GAPD2 takes over its function and allows the Calvin cycle to work. From this perspective, almost all of the reactions in the Calvin cycle need to be actively working in order to sustain growth, with the exception of FBA, FBA3, FBP, SBPASE and TALA, which can be taken over by other specific reactions or pathways. Same kind of trend can be found throughout glycolysis and the TCA cycle, where only MDH and FUM can be taken over by other specific reactions or pathways. In order to determine the effect on constraining the ferrous ion uptake rate, a second simulation was performed. In this case same the constraints as in the previous simulation were applied, but instead of leaving the ferrous ion uptake rate unconstrained, it was constrained to -140 mmol/gDW/h (only lower bound). Solutions from the second FBA simulation were divided by the solutions obtained in the first simulation. The corresponding ratio is outlined in Figure 1-10B. By constraining ferrous ion uptake rate, it was not possible for the system to generate the needed reducing power in the form of nadh, nor the sufficient ATP to use all the available h2co3, 83 % of the available carbon was used to generate biomass. This is shown in Figure 1-10B, where almost all reactions showed a decrease of approximately in 83 % compared to the first analysis. Discrepancies arose from alternate optimum, where reactions FBA and FBP are replaced by a cycle generated by the TALA reaction in conjunction with FBA3 and SBPASE. In order to analyze the carbon flux through the network, when forcing the

system to use all carbon under ferrous ion restriction, a third simulation was performed. In this simulation, the same constraints as in the second simulation were applied, but instead of setting only the lower bound for the h2co3 uptake rate, the upper bound was also constrained to -2 mmol/gDW/h. Solutions from the third FBA simulation were divided by the solutions obtained in the first FBA simulation. The corresponding ratio is shown in Figure 1-10C. For this simulation the system was not able to generate the needed reducing power in the form of nadh, nor the sufficient ATP for generating the optimum amount of biomass, instead, just like the previous result only 83 % of the available carbon was used to generate biomass. The rest of the carbon was used to generate EPS. All the changes observed are related to the production of EPS. In the case of PDH and SUCOAS, both reactions showed an increase compared to the first solution. This is due to the need of the system to generate precursors for fatty acid biosynthesis which enables the EPS production. Furthermore, a considerable increase was observed in the Calvin cycle, specifically in the TALA, FBA3 and SBAPSE loop. This is due to the need to produce f6p-B, which is also needed for producing the EPS precursor. This result shows how rigid the network is in terms of the carbon metabolism. Although evidence based on genome annotation indicates the presence of genes coding for the production of organic acids in *A. ferrooxidans*, specifically for a lactate/malate dehydrogenase (MDH), an acetyl-CoA synthetase (ACS) and a alcohol dehydrogenase (AFE_0697, not included in the model), no convincing experimental evidence regarding their production was found. Despite this, and in order to study the effect of carbon leaking throughout the cell due to metabolic co2 production, transport reactions allowing metabolic co2 transport through the periplasmic (CO2tpp) and extracellular (CO2tex) membrane were added.
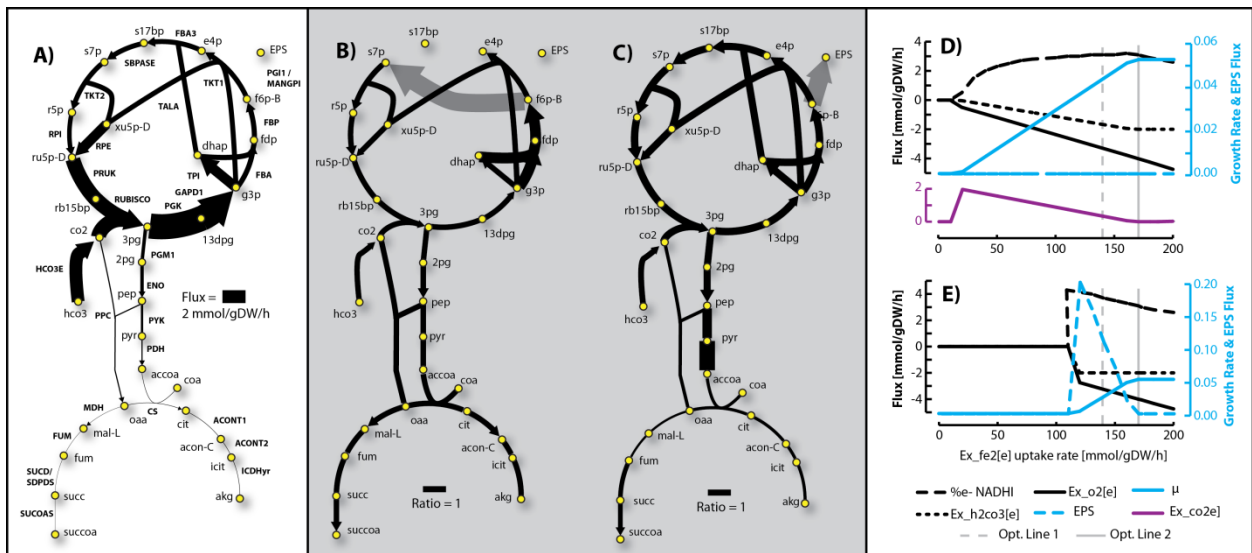
Figure 1-10: Carbon fixation and EPS flux distribution predicted for ferrous ion metabolism. A) This analysis was executed by setting the carbon uptake rate (lower bound) in the form of h2co3 to 2 mmol/gDW/h, ferrous ion was left unconstrained, and FBA was run to maximize the flux through BOF. Reactions are specify in capital letters, while metabolites in lowercase letters. Arrows indicate the direction of enzymatic activity and the arrow thicknesses are proportional to the flux through each reaction (a thicker arrow has a larger flux). B) A FBA flux comparison between flux distributions obtained in A and simulations ran when constraining the ferrous ion uptake rate (lower and upper bounds) to -140 mmol/gDW/h. Arrows indicate the direction of enzymatic activity and the arrow thicknesses are proportional to the ration between predictions (a thicker arrow shows that flux obtained in A is lower that the predicted when constraining ferrous ion uptake rate). C) A FBA flux comparison between flux distributions obtained in A and simulations ran when constraining the ferrous ion uptake rate (lower and upper bounds) to -140 mmol/gDW/h and carbon uptake rate (lower and upper bound) in the form of h2co3 to 2 mmol/gDW/h. Arrows indicate the direction of enzymatic activity and the arrow thicknesses are proportional to the ration between predictions (a thicker arrow shows that flux obtained in A is lower than the predicted when constraining ferrous ion and h2co3 uptake rate). D) The results of two different analyses were plotted. i) carbon uptake rate in the form of h2co3 was set to 2 mmol/gDW/h (lower bound), ferrous ion was varied between 0 and -140 mmol/gDW/h, and FBA was ran to maximize the flux through BOF. And ii), a metabolic co2 transport reaction was added, carbon uptake rate in the form of h2co3 was set to 2 mmol/gDW/h (lower and upper bound), ferrous ion was varied between 0 and -140 mmol/gDW/h, and FBA was ran to maximize the flux through BOF. Results for specific reactions were outlined. Lines in light blue represent reaction flux predictions for growth rate and EPS production. Reactions in black, represents the predicted flux in the oxygen and h2co3 uptake reactions. And the reaction in purple, represents the discrepancy between prediction i) and ii) (external metabolic co2 transport reaction flux). Optimality lines for predictions outline in A, B and C were plotted in gray. E) A flux distribution for simulations obtained by setting the carbon uptake rate in the form of h2co3 to 2 mmol/gDW/h (lower and upper bound), varying the ferrous ion uptake rate between 0 and -140 mmol/gDW/h, and running FBA to maximize the flux through BOF were outlined.

As mention previously, the second analysis was performed in order to study the system evolution in terms of carbon uptake, biomass, EPS and co2 production due to ferrous ion restriction. To study and compare different metabolic scenarios, three different simulations were performed (Figure 1-10D-E). The first simulation was performed by setting the carbon uptake rate in the form of h2co3 to -2 mmol/gDW/h (only lower bound), the co2 production rate was constraint to 0 (lower and upper bounds), oxygen uptake rate was left unconstraint, ferrous ion was varied between 0 and -200 mmol/gDW/h, and FBA was run to maximize the flux through BOF. The second simulation was performed in the same manner, but constraining the h2co3 uptake rate in both bounds, forcing the system to metabolize all of the carbon. The third simulation was performed identical to the second one, but this time metabolic co2 was allowed to leave the cell. Results from these simulations are shown in Figure 1-10D-E. For the first

simulation, results are specifically plotted in Figure 1-10D. Since only the h2co3 lower bound was set, the cell was freely allowed to metabolize the necessary amount of h2co3 to produce biomass, and not diverge fluxes towards EPS production. The specific amount of h2co3 consumed was highly constrained by the ferrous ion uptake rate, which delivers the electrons necessary for proton motive force, ATP and nadh generation. This is shown in Figure 1-10D, where the h2co3 uptake rate increases along the permissible ferrous ion uptake rate (x-axe). All carbon goes directly to biomass, reaching the optimum at approximately a ferrous ion uptake rate of -170 mmol/gDW/h (optimality line 2). This point represent the FBA flux distribution result used to plot Figure 1-10A, and the point described by the optimality line 2 (ferrous ion uptake rate constraint to -140 mmol/gDW/h) represents the flux distribution ratio plot in Figure 1-10B. Oxygen uptake rate continues to increase after reaching the optimum growth. This was to neutralize the incoming proton from ferrous ion and produce water from oxygen and protons by CYTAA31. When constraining lower and upper bounds in h2co3 uptake rate, the minimum ferrous ion flux necessary to sustain growth was approximately -120 mmol/gDW/h. Furthermore, EPS production associated to growth was determined. While ferrous ion uptake rate increases, EPS production decreases and biomass increases reaching its optimum when no EPS production was observed (optimality line 2 in Figure 1-10E). The point described by the optimality line 2 (ferrous ion uptake rate constraint to -140 mmol/gDW/h) represents the flux distribution ratio plot in Figure 1-10C. According to the third simulation, by not constraining metabolic co2 transport outside the cell, the same results are obtained as in the first simulation. Discrepancies arose in h2co3 consumption and metabolic co2 evolution (purple line in Figure 1-10D). In this case h2co3 uptake was always to -2 mmol/gDW/h, and co2 decreases until reaching the optimum growth, where the cell uses all h2co3 for biomass production. No EPS production was observed, since from an energetic point of view, EPS precursors are harder or energetically more expensive to produce compared to metabolic co2. This suggests that there might be some carbon losses due to metabolic co2 production during cell growth, depending the ferrous ion availability.

1.4.5 EPS production potential analysis

As mentioned in the previous section, due to network rigidity in terms of carbon metabolism, and in order to study the effect of carbon leaking through the cell due to metabolic co2 on EPS production, transport reactions allowing the metabolic co2 transport through the periplasmic (CO2tpp) and extracellular (CO2tex) membrane were added to *i*MC507. Two different analyses were performed. First, using FBA, a graph describing the effect of external metabolic co2 transport addition and the individual reaction knock-down on the predicted growth rate and EPS production was generated (Figure 1-11). When reactions were not constrained to a certain percentage (i.e. knock-down = 0 %), no effect on the growth rate nor on the EPS production was observed. This result was expected, since the system was left to freely take all ferrous ions and oxygen are needed. Also, no co2 production was associated and all carbon was utilized for biomass production. When constraining reactions to 25 % of its optimum value (i.e. knock-down = 25 %) and closing the metabolic co2 transport, several knock-down reactions showed that when perturbing them EPS production appears and growth rate diminishes. This perturbation directly affects the carbon and energy metabolism by diminishing the growth rate and forcing the system to produce EPS. For cases when metabolic co2 was allowed to leave the cell 50 % and 100 % from its optimum, EPS production decays up to 0, showing the preference of the cell to produce co2 instead of EPS, as mentioned before. Nevertheless, two reactions (i.e. MDH and FUM) regardless of co2 transport activation, were able to disrupt the system for the production of EPS. It also can be appreciated that the greater the knock-down for the reactions, the higher the EPS production is (i.e. moving from the bottom to the top). When reactions were constrained to 100 % (i.e. knock-down = 0 %), this means that no flux was allowed to pass through the reactions, simulating a single reaction knock-out for all reactions in the network. As shown in Figure 1-11, some reaction knock-outs were lethal for the cell where no growth rate and EPS production were calculated. This behavior happened regardless of the co2 transport. As mentioned before 2 different single reaction knock-outs were consistent in all of the simulated conditions, where biomass was produced while producing EPS. This reaction knock-out allows the insilico design of A. ferrooxidans strains for growth-coupled EPS production, which is crucial for increasing

the bioleaching capacity. A subset of reactions exist, that when knocked-down higher EPS production is achievable. By inspecting their corresponding subsystems association, it appears that the majority of them are related to electron transfer metabolism.
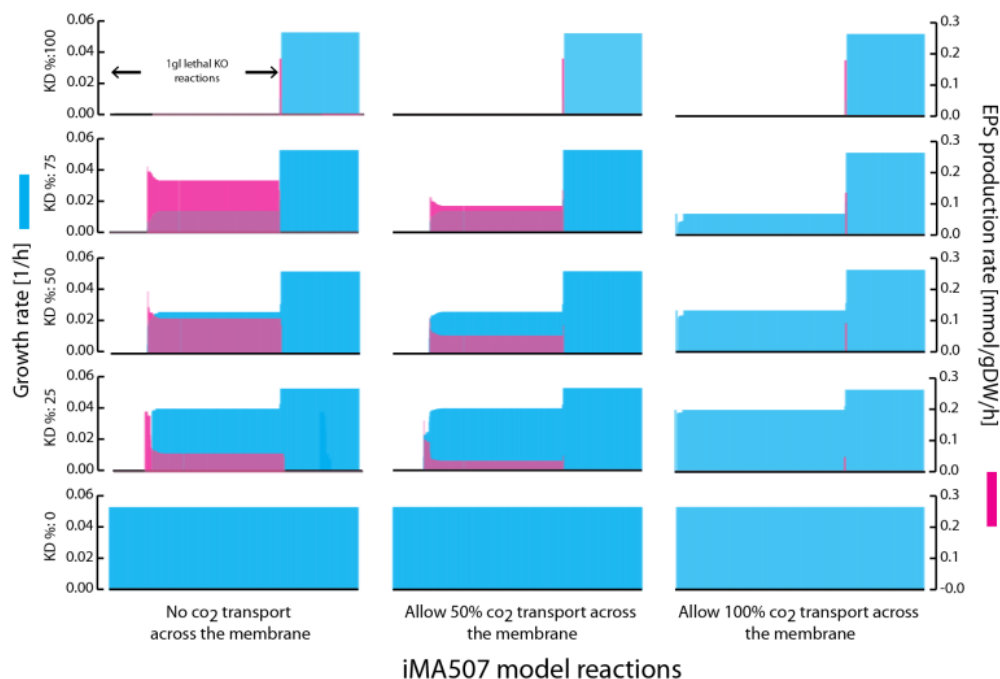
Figure 1-11: Knock-down and external metabolic co2 transport effect on growth rate and EPS production. A graph describing the external metabolic co2 transport and the individual reaction knock-down effect on the predicted growth rate and EPS production is outlined. Simulations were executed by adding an external co2 transport and constraining it in 0, 50 and 100% from the maximum capacity (x-axe), all reactions in the model were constraint between 0, 25, 50, 75 and 100% from their optimum capacities (y-axe). For each simulation the carbon uptake rate in the form of h2co3 was set to 2 mmol/gDW/h, oxygen and ferrous ion uptake rate were left unconstrained, and FBA was ran to maximize the flux through BOF. For each one of the simulations, growth rate (red bars) and EPS (light blue bars) were plotted. In each sub-graph the corresponding biomass and EPS production rate was plotted for each one of the reactions (x-axe) in the model.

Furthermore, an analysis on the EPS growth-coupled potential for additional double and triple knock-outs was evaluated. Results from this analysis are shown in Figure 1-12. A total of 134 growth-coupled designs were calculated (see Table A1-3). Two of them correspond to single knock-outs, 12 to double knock-outs and the rest to triple knock outs. Out of the 134 growth-coupled designs 4 different production envelopes were identified, which described the Pareto frontier in the EPS flux vs. the biomass flux plane. The four different regions were plotted in Figure 1-12A. Approximately 72 % of all the calculated designs correspond to region 2 (blue). Most of the knock-out designs in region 2 correspond to triple knock-outs (Figure 1-12B). Region 3 shows the higher theoretical EPS achievable flux, a total of 12 designs were

calculated, where all of them correspond to triple knock-outs. In all of the calculated designs two different reactions were always found, either fumarase (FUM) or malate dehydrogenase (MDH), which described the production enveloped in region 1. Actually, each of them was found independently 67 times among all designs (see Figure 1-12C), representing the reactions with more instances among all growth-coupled designs reactions. Despite the latter, the subsystem with more reaction instances in growth-coupled designs is the Nucleotide Salvage Pathway (see Figure 1-12D), which does not contain MDH and FUM that are part of the citric acid cycle.
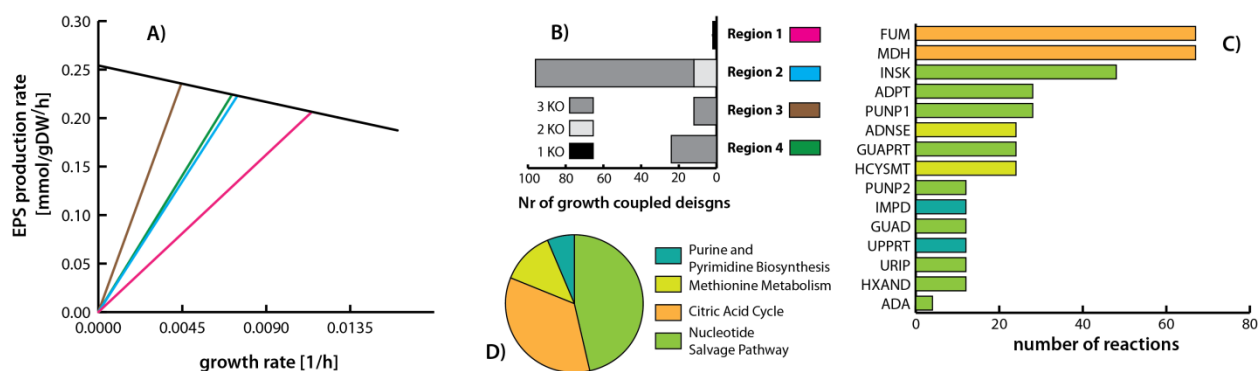
Figure 1-12: Growth-coupled EPS production strain design results. A) A graph that shows the production envelopes for different KO combinations that were calculated during the analysis. The EPS production rate is shown in the y-axe and the growth rate is given on the x-axe. 4 different production envelopes were characterized and described in different colors. B) A bar plot describing the number of growth coupled designs associated with the each characterized region. For each region the information about the number of knock-outs reactions associated with the corresponding growth coupled designs was incorporated. C) A bar plot showing the number of instances that each reaction in the y-axe was found in all growth coupled designs. Bar colors represents the subsystem associated with the corresponding reaction. D) A pie chart descrying the mayor subsystems associated with the predicted knock-out reactions.

1.5 Conclusions

A fundamental goal in biology is to understand and predict the genotype-phenotype relationships in the cell. COBRA methods organize biochemical, genetic and genomic knowledge into a mathematical framework which enables the quantitatively description of metabolic physiology. In this work, the first genome-scale metabolic reconstruction of *Acidithiobacillus ferrooxidans ATCC 23270, i*MC507 is presented, characterized, validated and utilized. The bottom-up metabolic network reconstruction of *i*MC507 represents a comprehensive knowledge base that summarizes and categorizes the information currently available for *A. ferrooxidans*, and serves as a framework for computational analysis. Based on experimental evidence and bioenergetics principles,

model parameters such as GAM, NGAM and proton translocation stoichiometry of key reactions involved in the electron transfer metabolism were predicted. iMC507 was validated for aerobic chemolithoautotrophic conditions by fixing co2 and using three different electron donors: ferrous ion, tetrathionate and thiosulfate. Based on the retrospective model-driven analysis and investigation of the electron transfer and central carbon metabolism, i) key electron transfer reaction proton translocation stoichiometry were predicted, ii) quantitatively phenotypic description of specific organism key properties (i.e. uphill vs downhill pathway and effect of growth rate due to proton exchanged) in the electron transfer metabolism under aerobic growth and using ferrous ion, tetrathionate and thiosulfate as electron donors was determined, iii) quantitative phenotypic description of central carbon metabolism for aerobic growth and using ferrous ion was determined, and iv) based on the knock-down and knock-out analysis, several growth-coupled designs able to produce EPS were calculated and evaluated for further experimental implementation.

iMC507 represents the firs genome-scale reconstruction of *A. ferrooxidans* specie and accounts for 507 ORFs (16%) of the current *A. ferrooxidans* genome annotation. iMC507 represent a major advance compared to the small scale metabolic network models predecessor, which only account for 62 reaction in the first case(Hold, Andrews and Asenjo 2009) and 190 in the second case(Sepúlveda, Cortéz, Abarca, Valdecantos, Iglesias and Roa 2011). iMC507 average confidence score was equal to 2.02. This means that most of the network content was based on genome annotation. In fact only 58% of the reactions were based on the genome annotation and 19% of all reactions in iMC507 have been very well or well-studied. The next step in the expansion of the *A. ferrooxidans* metabolic network will required research efforts directed to poorly described pathways. The current knowledge reflects the traditional use of *A. ferrooxidans* as a chemolithoautotrophic ferrous ion and RISCs oxidation-model bacterium, where subsystems such as co2 fixation, sulfur metabolism, oxidative phosphorylation, and EPS biosynthesis have been well studied. In contrast, metabolisms involved in amino acid fixation, central carbon, transport reactions, among others are poorly described and in need of further characterization. Furthermore, in order to increase the accuracy and utility of the model, BOF should be reevaluated in

future efforts, since some of the precursors were determined based on previous work non-related to *A. ferrooxidans* genome-scale models, specifically *i*AF1260. According to the latter, problems might arise when conditionally essential metabolites are inappropriately included in the BOF, leading to false reaction essentiality and growth-coupled design predictions.

Since its discovery *A. ferrooxidans* has been extensively studied under chemolithoautotrophic aerobic condition, by using mainly ferrous ion and RISCs compounds as electron donors. In this study the proton translocation stoichiometry for key enzymes involved in the ferrous ion and RISCs metabolism together with the GAM and NGAM values were predicted. Specifically for the CYTAA31 reaction (cytochrome c oxidase aa3-type), involved in ferrous ion metabolism, much debate has been generated during the years whether it actually translocates protons across the membrane. Based on the genetic-algorithm base approach, no proton translocation stoichiometry for this reaction was predicted. Furthermore, ATP synthase proton translocation stoichiometry was determined, showing how COBRA methods in conjunction with experimental phenotypic data can be useful for elucidation of chemiosmotic parameters of the cell. It is worthwhile to notice that all of these predictions were performed assuming an external pH equal to 2. Since thermodynamic bound on proton translocation stoichiometry changes depending on the pH difference between the cytoplasm and the periplasm, different results might be obtained when changing the pH conditions.

A quantitatively phenotypic description of specific organism key properties in the electron transfer metabolism (i.e. uphill vs downhill pathway and effect of growth rate due to proton exchanged) under aerobic growth and using ferrous ion, tetrathionate and thiosulfate as electron donors was determined. For FIM, a complete and accurate description of the electron transfer metabolism was performed, showing that most of the electrons (i.e. more than 90 %) go to the proton neutralization by o2 reduction into water through CYTAA31. The rest of the electrons go directly to the Calvin cycle via the "uphill pathway". Furthermore, media alkalization, a characteristic behavior during ferrous iron growth was also predicted. With these results model prediction capacity was validated for FIM. More interesting was the study of TTM and TSM, where the TSM efficiency over the TTM was quantitatively demonstrated. Furthermore, a "RISC uphill pathway" and a

"RISC downhill pathway" were described and quantitatively characterized. Specifically the "RISC downhill pathway" is described by two independent systems (i.e. CYTBO3/CYTBD and CYTAA32 in conjunction with CYTRED). Attention must be paid in future work to elucidate if those pathways actually work independently from each other or in parallel. When comparing the "downhill pathway" in FIM and RISC metabolism, proton translocations in reactions involved in the RISC metabolism (i.e. CYTBO3/CYTBD and CYTAA32 in conjunction with CYTRED) were predicted, no proton translocation in CYTAA31 was found. By inspecting the reactions in more detail, in the case of RISC metabolism and FIM, protons used for o2 reduction come from cytosolic protons integrated through TSQOC and CYTAA31 respectively. According to the latter, in order to neutralize the protons in the cytoplasm the "downhill pathway" in FIM and RISC metabolism uses the protons to generate water. Furthermore, this process is aided by proton translocation through the "RISC downhill pathways" in the RISC metabolism.

In this work the central carbon metabolism was outlined by incorporating three major pathways: the Calvin cycle, glycolysis and the incomplete TCA cycle. The junction point was determined by 3-phosphoglyceric acid, which connects glycolysis with the Calvin cycle. The Calvin cycle showed a higher flux activity compared to the rest of the central carbon metabolism, mainly to sustain the co2 fixation through the RUBISCO reaction. Approximately 96 % of the carbon flux through the Calvin cycle is used for co2 fixation, and the rest is used for generating biomass precursors. For these biomass precursors the carbon flux is diverged to glycolysis and the Calvin cycle in a 2.5/1 ratio. The central carbon metabolism is tightly coupled to the energy metabolism through the Calvin cycle, specifically in two reactions, first through the GAPD reaction, where almost all nadh is used for anabolic porpoises, and secondly through the PGK reaction, which uses 19 % of all ATP produced by the ATP synthase. By constraining the ferrous ion uptake rate to suboptimal values and forcing the cell to metabolize all incoming h2co3, a decrease in the growth rate, together with an EPS production was observed. This shows how rigid the metabolic network is in terms of the carbon metabolism backbone. Thus, in order to allow more degrees of freedom, regarding the internal carbon fluxes a metabolic co2 transport was hypothesized and incorporated into the metabolic network. Result

showed that instead of producing EPS, the system generates co2. Although these results are reasonable, no experimental evidence regarding metabolic co2 evolution, or small compound secretions (i.e. ethanol, lactate, acetate, etc) has been reported. This shows how COBRA methods might aid the process of hypothesis-driven discovery by simulating insilico phenotypes candidates for experimental validation.

EPS production sensitivity was evaluated using two variables. The first was EPS production due to individual reaction knock-down. The second variable was the presence or absence of a metabolic co2 transport mentioned above. Overall, when no metabolic co2 was allowed to leave the cell, by increasing the knock-down strength, a subset of perturbed reactions was identified to promote EPS production. Due to the network rigidity in terms of carbon metabolism, this result was expected. On the other hand, when metabolic co2 was allowed to leave the cell, EPS production was observed only when FUM and MDH were perturbed. Furthermore, when these reactions were knocked-out, growth coupled EPS production was observed. Evidence regarding the incomplete TCA cycle disruption leading to EPS overproduction (Sepúlveda, Cortéz, Abarca, Valdecantos, Iglesias and Roa 2011), validates this results since FUM and MDH are reactions present in the incomplete TCA cycle. Growth-coupled production of EPS depends on the energy benefit that the cell can obtain through the pathway activation related to EPS production. By knocking-out reactions an energy imbalance is generated. Growth coupled designs are generated when the cell recovers the imbalance by coupling EPS production related pathways to growth. The evaluation of all possible single, double and triple knock-outs for growth coupled design was performed. In total 134 growth-coupled designs were found. In all of the predicted designs either FUM or MDH were found to be involved. Additionally reactions related to the Nucleotide salvage pathway, pyrimidine metabolism and methionine metabolism were found in double and triple knock-outs. Furthermore, growth coupled designs, such as those produced here, provide an extra tool for metabolic engineers by allowing the use of selection pressure to achieve a desired production state. Nevertheless, due to the lack of genetic tools that allows the knock-out implementation in *A. ferrooxidans*, it is not possible to implement the design experimentally yet. Still, competitive inhibition might be used in order to

disrupt the metabolism as utilized elsewhere (Sepúlveda, Cortéz, Abarca, Valdecantos, Iglesias and Roa 2011).

As the field of constraint based modeling and analysis continues to expand by incorporating different cellular processes and interactions into the genome-scale space (e.g. transcription and translation in *Escherichia coli* (O'Brien et al. 2013) and *Thermotoga maritime* (Lerman et al. 2012)), *i*MC507 will serve as a key component for the quantitative study of *A. ferrooxidans* and related organisms by providing an extensive map of the cellular metabolism. Furthermore, *i*MC507 represents a major breakthrough for the *A. ferrooxidans* community, by unifying and describing our knowledge of this unique species and providing a computational platform for further analysis and hypothesis formulation for environmental and biotechnological applications.

Chapter 2

Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path

2.1 Abstract

The production of 75% of the current drug molecules and 35% of all chemicals could be achieved through bioprocessing (Arundel and Sawaya 2009). To accelerate the transition from a petroleum-based chemical industry to a sustainable bio-based industry, systems metabolic engineering has emerged to computationally design metabolic pathways for chemical production. Although algorithms able to provide specific metabolic interventions and heterologous production pathways are available, a systematic analysis for all possible production routes to commodity chemicals in *Escherichia coli* is lacking. Furthermore, a pathway prediction algorithm that combines direct integration of genome-scale models at each step of the search to reduce the search space does not exist. Previous work (Feist et al. 2010) performed a model-driven evaluation of the growth-coupled production potential for *Escherichia coli* to produce multiple native compounds from different feedstocks. In this study, we extended this analysis for non-native compounds by using an integrated approach through heterologous pathway integration and growth-coupled metabolite production design. In addition to integration with genome-scale model integration, the GEM-Path algorithm developed in this work also contains a novel approach to address reaction promiscuity. In total, 245 unique synthetic pathways for 20 large volume compounds were predicted. Host metabolism with these synthetic pathways was then analyzed for feasible growth-coupled production and designs could be identified for 1271 of the 6615 conditions

evaluated. This study characterizes the potential for *E. coli* to produce commodity chemicals, and outlines a generic strain design workflow to design production strains.

2.2 Introduction

The global chemical industry has been driven by petroleum feedstocks for the past 100 years, where synthetic organic chemistry played a key role. Today, the global chemical market landscape is beginning to change, based on new possibilities for bio-based product and process development. The renewed interest in industrial biotechnology is due to several reasons. First, the increases in petroleum prices squeeze commodity chemical production margins, increasing economically attractiveness of bio-based processes. Second, there is a strong socio-economic driver towards green chemistry and renewable feedstocks (Keasling 2012). Third, due to technological developments, the past 20 years has seen the successful demonstration of metabolic engineering enabling the generation of microbial strains for the production of a wide range of chemical compounds (Atsumi and Liao 2008; Lee, Na, Park, Lee, Choi and Lee 2012; Peralta-Yahya et al. 2012). The availability of high-throughput technologies, the advances of computational methods, and emergence of genome-scale systems analysis to analyze large amount of omics data, has given rise to the concept of 'systems metabolic engineering' (Jang et al. 2012; Lee, Na, Park, Lee, Choi and Lee 2012; Palsson and Zengler 2010) where the focus has shifted from perturbing individual pathways to manipulating the organisms as a whole. Genome-scale models (GEMs) can now be used as query platforms to examine new strategies and interventions as they contain a parts list of cellular components and their interactions (Feist, Henry, Reed, Krummenacker, Joyce, Karp, Broadbelt, Hatzimanikatis and Palsson 2007; Feist, Herrgard, Thiele, Reed and Palsson 2009; Orth, Conrad, Na, Lerman, Nam, Feist and Palsson 2011). By using constraint-based reconstruction and analysis (COBRA) approaches (Schellenberger, Que, Fleming, Thiele, Orth, Feist, Zielinski, Bordbar, Lewis, Rahmanian, Kang, Hyduke and Palsson 2011), outcomes of cellular metabolism have been predicted successfully for the production of various compounds (Bordbar et al. 2014a; Kim et al. 2008; Lee, Na, Park, Lee, Choi and Lee 2012; McCloskey, Palsson and Feist 2013; Yim, Haselbeck, Niu, Pujol-Baxley, Burgard, Boldt, Khandurina, Trawick, Osterhout, Stephen, Estadilla, Teisan, Schreyer, Andrae, Yang, Lee, Burk and

Van Dien 2011). Moreover, model-driven evaluations for the production potential for growth-coupled native products in *E. coli* have been performed (Feist, Zielinski, Orth, Schellenberger, Herrgard and Palsson 2010). However, a comprehensive computational analysis for the production of valuable non-native *E. coli* metabolites has not been performed. Therefore, we developed a systematic workflow in order to evaluate the production potential of 20 industrially relevant chemicals (Assary and Broadbelt 2011; Curran and Alper 2012; Fischer, Klein-Marcuschamer and Stephanopoulos 2008; Lee, Na, Park, Lee, Choi and Lee 2012; Paster, Pellegrino, Carole, Energetics, U.S. Department of Energy and Renewable Energy 2003; Werpy, Petersen, Aden, Bozell, Holladay, White, Manheim, Eliot, Lasure and Jones 2004; Zeng and Sabra 2011) in *E. coli*, by integrating a combination of computational methods and developing a new pathway prediction algorithm, GEM-Path (Genome-scale Model Pathway Predictor).

Computational approaches for the prediction of non-native pathways exist, but are limited in their design and scope. Different approaches have been implemented for pathway prediction (Arita 2000; Carbonell, Planson, Fichera and Faulon 2011; Cho, Yun, Park, Lee and Park 2010; Dale, Popescu and Karp 2010; Greene, Judson, Langowski and Marchant 1999; Hatzimanikatis, Li, Ionita, Henry, Jankowski and Broadbelt 2005; Heath, Bennett and Kavraki 2010; Hou, Wackett and Ellis 2003; McShan, Rao and Shah 2003; Pharkya, Burgard and Maranas 2004), where increasing attention has been focused mainly on retrosynthetic algorithms (Carbonell, Planson, Fichera and Faulon 2011; Cho, Yun, Park, Lee and Park 2010; Henry, Broadbelt and Hatzimanikatis 2010; Yim, Haselbeck, Niu, Pujol-Baxley, Burgard, Boldt, Khandurina, Trawick, Osterhout, Stephen, Estadilla, Teisan, Schreyer, Andrae, Yang, Lee, Burk and Van Dien 2011) based on Biochemical Reaction Operators (BROs). In these analyses, BROs are used to go from a target compound to a predefined set of metabolites in an iterative backward search. In summary, all of these methods shared basically the same workflow, first calculating all structurally possible pathways and then scoring them using different kinds of metrics. During the synthetic pathway calculation, these algorithms unnecessarily expand the reaction space, generating all possible pathways that link a specific metabolite to a final specific product without performing pathway integration with content known to exist in a given production host. Furthermore, previous algorithms do

not integrate the bioprocessing condition-specific cofactor usage/generation, substrate usage, strain/oxygenation conditions, and related energy balances during the computation of pathways. In order to address these problems, we developed GEM-Path, by integrating retrosynthetic algorithms based on BROs and filtering procedures with GEMs at each iteration step. Furthermore, a novel reaction promiscuity analysis is introduced, which is based on known reaction substrate similarities. These two features distinguish GEM-Path from other computational approaches.

Once a synthetic pathway is successfully established, additional approaches can be taken to further engineer the host strain and synthetic pathways for enhanced production of a desired chemical. Adaptive laboratory evolution together with COBRA methods and organism-specific models has proven successful for the calculation of wild type *E. coli* optimal growth rates (Ibarra, Edwards and Palsson 2002), native *E. coli* metabolite production through knock-outs (Fong, Burgard, Herring, Knight, Blattner, Maranas and Palsson 2005), and for non-native *E. coli* metabolite production through heterologous pathway incorporation and knock-out implementations (Yim, Haselbeck, Niu, Pujol-Baxley, Burgard, Boldt, Khandurina, Trawick, Osterhout, Stephen, Estadilla, Teisan, Schreyer, Andrae, Yang, Lee, Burk and Van Dien 2011). Furthermore, the use of adaptive laboratory evolution together with growth-coupled knock-outs design, allows to select for strains with higher target compound production rates by coupling them to the selection for faster growth (Portnoy, Bezdan and Zengler 2011). Here, we integrate each of the predicted pathways under several different substrates/strain/oxygenation conditions with growth-coupled designs generated through reaction knock-outs by utilizing the RobustKnock (Tepper and Shlomi 2010) and GDLS (Lun, Rockwell, Guido, Baym, Kelner, Berger, Galagan and Church 2009) algorithms. Finally, in order to characterize *E. coli*'s potential production landscape for the studied compounds and for designs implementation purposes, a productivity analysis for maximum theoretical yield and maximum theoretical growth-coupled yield was performed.

2.3 Materials and Methods

2.3.1 Model and flux balance analysis

The metabolic reconstruction of *Escherichia coli* *i*JO1366 was utilized as a basis for synthetic pathway calculations, yield analysis, and further strain designs. This model has been proven to be predictive for computations of growth rates and metabolite excretion rates on a range of substrates and genetic conditions (Feist, Henry, Reed, Krummenacker, Joyce, Karp, Broadbelt, Hatzimanikatis and Palsson 2007; Orth, Conrad, Na, Lerman, Nam, Feist and Palsson 2011). For all phenotype simulation, flux balance analysis (FBA) was used. The biomass objective function (BOF_{core}), maintenance energy, and basic constraints were set according to the reported values in the reconstruction. FBA used the assumption of steady-state metabolic flux as described elsewhere (Orth, Thiele and Palsson 2010). All computations were performed using MATLAB® (The Mathworks Inc., Natick, MA, USA) and the COBRA Toolbox (Schellenberger, Que, Fleming, Thiele, Orth, Feist, Zielinski, Bordbar, Lewis, Rahmanian, Kang, Hyduke and Palsson 2011) software packages with TOMLAB (Tomlab Optimization Inc., San Diego, CA, USA) solvers.

2.3.2 GEM-Path Algorithm: Chemoinformatics tools and techniques

Throughout the process of synthetic pathway generation, chemoinfomatic tools were essential for integrating computational chemical analysis into genome-scale model theory. In order to properly handle molecular structures, a range of chemoinfomatic techniques were incorporated into the COBRA Toolbox MATLAB® environment. For this purpose, in-house methods and functions, which are described below, were developed based on ChemAxon (ChemAxon Ltd., Budapest, Hungary) software package libraries.

Chemical Representation: for compound and reaction representation MDL Molfiles (Dalby et al. 1992) were used. A Molfile contains information about the atoms, bonds, connectivity, and coordinates of a molecule. The Molfile consists of some header information, the connection table containing atom information, then bond connections and types, followed by sections for more complex information.

SMIRKS & SMARTS: For BRO representation, SMIRKS (James et al. 2004) was used as a language for describing generic reactions by using a SMARTS (James, Weininger and Delany 2004) representation of the reaction's substructures. A SMARTS pattern may include not only a specification of reaction center but also a specification of a local structure that must occur or is necessarily absent based on our best understanding of the relevant biochemistry (Silverman 2002). BROs were constructed based on the smallest substructure related to the structural change of the main substrates and products in the reaction. Based on previous studies (Henry, Broadbelt and Hatzimanikatis 2010; Mu et al. 2011; Yim, Haselbeck, Niu, Pujol-Baxley, Burgard, Boldt, Khandurina, Trawick, Osterhout, Stephen, Estadilla, Teisan, Schreyer, Andrae, Yang, Lee, Burk and Van Dien 2011), a set of 443 irreversible BROs were defined to generate novel biochemical reactions and pathways. Approximately 76% of the reactions in KEGG (Kanehisa, Goto, Hattori, Aoki-Kinoshita, Itoh, Kawashima, Katayama, Araki and Hirakawa 2006) and 72% of the reactions in BRENDA (Curran and Alper 2012) involved a transformation captured in this defined BRO set. Furthermore, depending on the BROs's nature, three different types of metabolic transformation were defined: i) '1-1' BROs simulate the substrate conversion without including any co-products and co-substrates in the BRO, ii) '2-1' BROs simulate anabolic conversions, merging the substrate with a cosubstrate, and iii) '1-2' BROs simulate catabolic conversions, where the substrate breaks into the corresponding product and a co-product. 2-2 transformations were ignored since they can be represented by a 2-1 transformation followed by a 1-2 transformation. Co-products and co-substrates were selected from *E. coli's* metabolome information. This formulation allows a host-specific integration at the reaction prediction level.

Standardization and Mass Balance: since MDL Molfiles might come from different sources, a standardization procedure was performed. For each molecular structure, stereochemical information was removed and the major protonation form at pH 7 was determined. For each reaction, mass balance was performed using previously standardized molecular structures. If hydrogen did not reach the balance, reaction stoichiometry was corrected.

Substrate Fingerprint: Substrates were represented by chemical fingerprints. A chemical fingerprint (CFP) is a simple record of the fragments present in a chemical structure. The chemical fingerprint (CFP) of a molecule is defined as CFP = (Fi), where Fi refers to a molecular fragment with real occurrence in a molecule. Fi is obtained by the molecular fragmentation method. Each Fi in the fingerprint is represented in bit string where each position of the sequence is represented by '1' or '0' digits, depending on the presence or absence of the structural pattern predefined by Fi. Previous studies have shown good results by using linear fragments from 5 to 6 bonds (Hu et al. 2012; Latino and Aires-de-Sousa 2009). In this study linear fragmentation up to 6 bonds was used.

Tanimoto Coefficient (TC): The premise of similarity searching is that similar structures have similar fingerprints. Here, we used the TC dissimilarity (TC_{diss}) metric to determine how similar two fingerprints were. Values of this metric are non-negative numbers. A zero dissimilarity value indicates that the two fingerprints are identical, and the larger the value of the dissimilarity coefficient the higher the difference between the two structures. In its original form, the Tanimoto metric is a similarity metric (TC_{sim}):

$$T_{sim} = \frac{B(a\&b)}{B(a) + B(b) - B(a\&b)}$$

Where *a* and *b* are two binary fingerprints, & denotes binary bit-wise and-operator, | denotes bit-wise or-operator, and *B(x)* is the number of 1 bits in any binary fingerprint x:

$$B(x) = |\{x_i = 1\}|x_i \in \{0,1\}; i = 1, \dots \dots, n\}| = \sum_{i=1}^{n} x_i$$

From that it is straightforward to obtain a dissimilarity measure:

$$T_{diss} = 1 - T_{sim}$$

It is worth noting that if the TC_{diss} between two fingerprints is 0, it means that both molecules share the exact same fingerprint.   While this doesn't mean that both molecules are the same, it does mean that both molecules share the same bonds

according to the fragmentation process, since the molecular fingerprint only represents the presence or absence of a given particular bond pattern.

Exact topology matching: Molecular graphs consist of nodes and edges, with atoms corresponding to the nodes and bonds corresponding to the edges. When we compare structures represented as graphs, the graph patterns must match. The type of atoms and bonds must be similar during the structural search. In this study, no stereochemical information was used for matching compounds, only bond and atom connectivity for structural matching was analyzed. A full structure search solution in MolSearch is based on a substructure search algorithm (Ullmann's algorithm) combined with various heuristics and an additional check to verify that the number of heavy atoms are the same in the query and target molecules.

2.3.3 GEM-Path Algorithm: Databases

The *E. coli* metabolome was defined based on the GEM *i*JO1366. Metabolites were extracted from the model and downloaded from PubChem´s (Bolton et al. 2008) compound database. Metabolites were saved as molfiles and named after their BiGG (Schellenberger et al. 2010) identifier. For reaction existence and reaction promiscuity analysis, the BRENDA (Scheer et al. 2011) database file and molecular structure molfiles were downloaded. Three digit EC number databases were generated by lumping together all reactions with similar third level EC numbers. Each entry in the database specifies the corresponding known biochemical reaction formula, the corresponding four digit EC number association, reaction-organism association, and substrate structure file. In cases where a specific reaction-organism association reported affinity for more than one substrate, an entry specifying all substrates was generated. For this purpose all reactions were assumed to be reversible and cofactors were not assigned as substrates.

2.3.4 GEM-Path Algorithm: Thermodynamic Analysis

Thermodynamic analysis was performed by calculating the $\Delta_r G'$ (KJ/mol) where $\Delta_r G'^0$ was estimated based on the group contribution method (Jankowski et al. 2008). Intracellular concentrations were defined based on previous studies (Bennett et al.

2009). For unknown concentrations, estimations were calculated based on the non-polar surface area and compound charge (Bar-Even et al. 2011).

2.3.5 GEM-Path Algorithm: Promiscuity Analysis

This analysis takes into account only substrate reaction promiscuity. Based on the similarity (TC) of the native and non-native substrates, a reaction promiscuity space can be generated and potential promiscuous activities determined depending on the distance between the promiscuous space and the metabolite to analyze. Thus, a similarity matrix based on the TC was calculated between every possible metabolite that the specific reaction-organism association could catalyze; cofactors were excluded from the matrix. Then, the reaction promiscuity space was defined by performing multilinear regression analysis on the similarity matrix, and an average distance between each native metabolite and the space centroid was calculated. By dividing the potential promiscuous target substrate distance from the centroid over the average native distance from the centroid, the reaction promiscuity score (PS) was calculated. If the score was lower than 1.2, the reaction is considered to be promiscuous for the target substrate (Figure 2-5). The reaction promiscuity score was tested and validated by using *E. coli's* promiscuous reaction information from *i*JO1366 (Figure 2-1 and Figure 2-2).
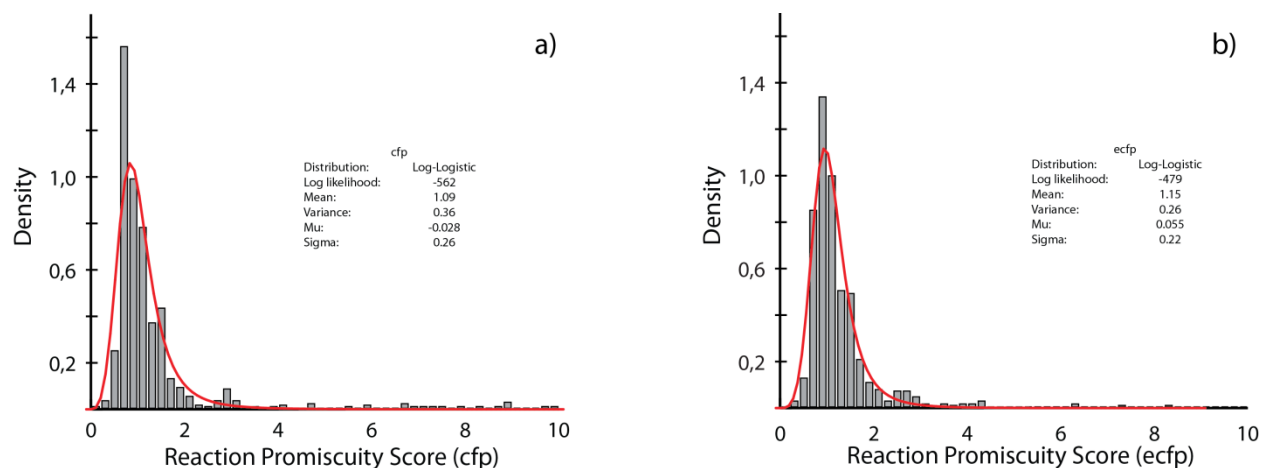
Figure 2-1: *E. coli* iJO1366 reaction promiscuity score analysis. For this analysis promiscuous enzymes reported in *E. coli* GEM iJO1366 (Orth, Conrad, Na, Lerman, Nam, Feist and Palsson 2011) were used (Nam et al. 2012). Only enzyme substrate promiscuity was taken into account addressing only the substrates that participate in the same chemical transformation. Promiscuous enzymes were tested by analyzing the substrate promiscuity score for each substrate. This was done by leaving one substrate out for each enzyme analysis. For comparing molecules and for the similarity matrix construction, two different fingerprints were analyzed, A chemical fingerprint (cfp) and B extended-connectivity fingerprint (ecfp). Both diagrams show the distribution of the reaction promiscuity score along the x-axis. Red lines represent the lo-logistic curve adjusted to the data. For each curve statistical information and parameters regarding the adjusted curve are specified.
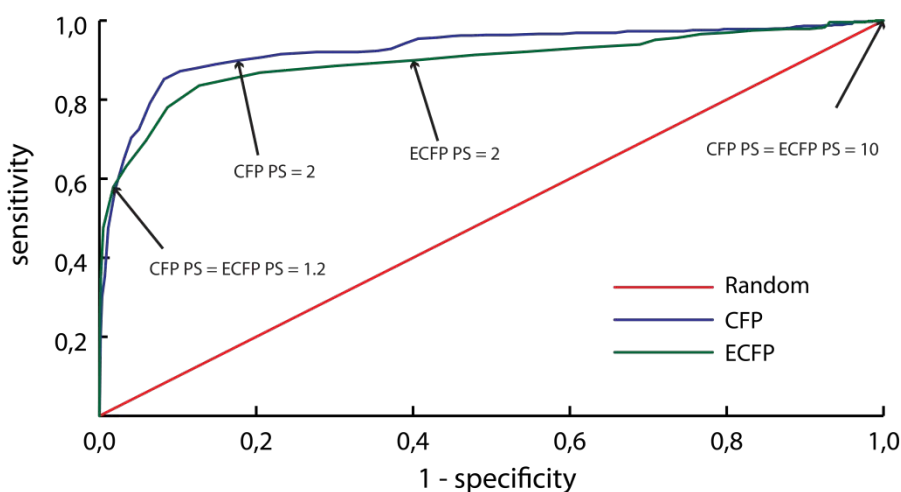
Figure 2-2: *E. coli* iJO1366 reaction promiscuity score ROC analysis. In this analysis the promiscuity score sensitivity was determined. For different promiscuity score threshold values (between 0 and 10) the promiscuity score was calculated for each promiscuous enzyme in *E. coli* GEM iJO1366 and each compound in the metabolome. Depending on the threshold value, different false-positive, true-positive, false-negative and true-negative distributions were obtained. For each threshold value the sensitivity and specificity were calculated and plotted. This analysis was performed for two different fingerprints, chemical fingerprints (blue) and extended-connectivity fingerprints (red). Thresholds between 1.2 and 2 showed a good agreement between predictions and known data.

2.3.6 Theoretical analysis of the production potential in *E. coli*

To evaluate the production efficiency of each product under different metabolic conditions and to determine the most predominant metabolic subsystems that work as precursor sources for product formation, an initial theoretical analysis was performed calculating the maximum theoretical yield in *E. coli for* all predicted pathways. This analysis was executed by: i) incorporating the heterologous pathways to the model, ii) setting an uptake rate to 120 C-mmol $gDW^{-1}h^{-1}$ for each carbon source, 20 mmol $gDW^{-1}h^{-1}$ O_2 (Varma et al. 1993) when specified, iii) setting the reactions CYTBDpp, CYTBD2pp, and CYTBO3_4pp to 0 mmol $gDW^{-1}h^{-1}$ for the ECOM strain (Portnoy et al. 2008), iv) setting a minimal growth rate to sustain growth as 0.1 h^{-1} (as set by the amount of flux necessary through the BOF_{core}), and v) using FBA to maximize the flux through each of the exchange reactions in the model for the target compound. For each predicted pathway, phenotypic results were reported in terms of yield; specific product yield ($Y_{p/s}$) defined as the maximum amount of carbon product that can be generated per unit of carbon substrate.

$$Y_{p/s} = \frac{C_{product} * production\ rate_{product}}{C_{susbtrate} * production\ rate_{susbtrate}} \left(\frac{cmmol_{product}}{cmmol_{susbtrate}} \right)$$

Where *C* is the number of carbons in the substrate and product. This metric provides a proper comparison between pathways productivities, since it standardizes the carbon consumption for each substrate.

2.3.7 Strain Design Computations

Before strain design, the model was preprocessed based on the problem formulation described by Feist et al. (Feist, Zielinski, Orth, Schellenberger, Herrgard and Palsson 2010). Preprocessing was condition specific and was performed for each pathway/substrate/oxygenation combination. The method utilizes six steps in which the model was reduced and target reactions were selected for knock-out simulations. By reducing the model and constraining the reaction set that could serve as a target for a reaction knock-out, computation time was effectively reduced when performing Robust Knock and GDLS algorithms.

RobustKnock and GDLS were implemented in the COBRA Toolbox framework as described in their original documentation. First, RobustKnock was utilized to design strains of *E. coli* for each target/substrate/oxygenation combination for a maximum of 2 and 3 reaction knock-outs. RobustKnock predicts reaction deletion strategies that lead to the over-production of compounds of interest by accounting for the presence of competing pathways in the network. Specifically, this method extends OptKnock to pinpoint specific enzyme-catalyzed reactions that should be removed from a metabolic network, such that the production of the desired product becomes an obligatory byproduct of biomass formation. The predicted set of reaction knock-outs eliminates all competing pathways that may hinder the chemical's production rate, resulting in more robust predictions than those obtained with OptKnock. This is achieved by searching for a set of reaction knock-outs under which the minimal guaranteed production rate of a chemical of interest is maximized, instead of simply assuming that the maximized production rate would be achieved by chance, as in OptKnock. The method is based on a bi-level max–min optimization problem that is efficiently solved via a transformation to a standard mixed-integer linear programming (MILP) problem. If the solution exists, this

algorithm finds the global optima set of knock-outs that evaluate the maximum achievable yield for a specific target compound. Because of the nature of this algorithm and the large amount of combinations to simulate, a search with four knock-outs makes the computational time of the simulations intractable. Because of this, GDLS was used to evaluate the maximum theoretical yield for four knock-outs. GDLS is a scalable, heuristic, algorithmic method that employs an approach based on local search with multiple search paths (k=2), that results in an effective, low-complexity search of the space of genetic manipulations. Still, solutions found with this method do not assure a global optimum. Consumption rate for the main carbon substrate in each simulation was set to 120 C-mmol gDW^{-1}h^{-1}. If aerobic conditions were used, an oxygen uptake rate of 20 mmol gDW^{-1}h^{-1} was also set. For the ECOM strain reactions CYTBDpp, CYTBD2pp, and CYTBO3_4pp were set to 0 mmol gDW^{-1}h^{-1}.

2.4 Results and Discussion

A systematic workflow was developed and organized into three phases (Figure 2-3). First, a synthetic pathway algorithm was developed which integrates GEMs directly into computation and industrially relevant target compounds for simulation were defined. Second, pathway production capabilities were examined in a number of production environments. Each pathway was incorporated into the *E. coli* GEM and analyzed in terms of maximal theoretical yield under different substrate, oxygenation, and strain conditions. Third, strain design computations was performed through a maximum yield analysis, utilizing the RobustKnock (Tepper and Shlomi 2010) and GDLS (Lun, Rockwell, Guido, Baym, Kelner, Berger, Galagan and Church 2009) algorithms . The result was a compendium of candidate synthetic pathways leading to 20 large volume commodity chemicals and strain designs to couple their production to growth.
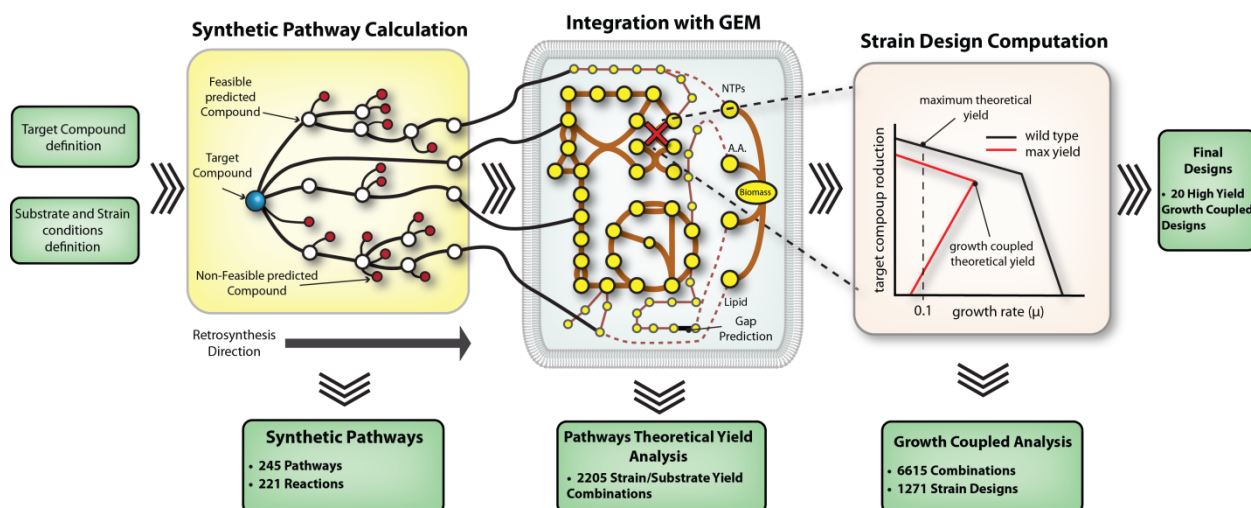
Figure 2-3: Synthetic Pathway Calculation and strain design pipeline. This workflow outlines the integrated process of synthetic pathway prediction (yellow box), constraint-based modeling with the *E. coli* GEM (blue), and strain design computation with design algorithms (pink box). Green boxes represent framework inputs (entry arrows) and general result outputs (exit arrows). From the left, target compounds and substrates/strain conditions were defined to generate synthetic pathways. Synthetic pathways were calculated by using the developed GEM-Path algorithm integrated with GEM computation. Following GEM-Path, each pathway leading to a specific target compound was evaluated for growth-coupled feasibility under previously defined substrate/strain conditions. This workflow was used to outline the production routes from a distance of 4 reaction steps from *E. coli's* metabolome to 20 commodity chemicals.

2.4.1 Synthetic pathway prediction algorithm development

GEM-Path combines and integrates different computational approaches (Table A2-1). The motivation for generating this new framework was that no existing tool combined a comprehensive search of the biochemical space through reaction operators, a thermodynamic analysis of each step, and a filtering of possible reactions at each step through integration with a strain-specific GEM.

2.4.1.1 Biochemical Reaction Operators (BROs) formulation.

An initial step in the design process was to define the set of Biochemical Reaction Operators (BROs) that accurately describes the biochemical reaction space. A total of 443 BROs were defined (see Methods). For use in GEM-Path , each BRO was assigned a specific cofactor use based on the BiGG database (Schellenberger, Park, Conrad and Palsson 2010) terminology, and the corresponding third-level EC number.

2.4.1.2 Pathway Predictor (GEM-Path) Algorithm

The pathway predictor algorithm was developed in an iterative manner (Figure 2-4). The process can be broken down into four major steps:
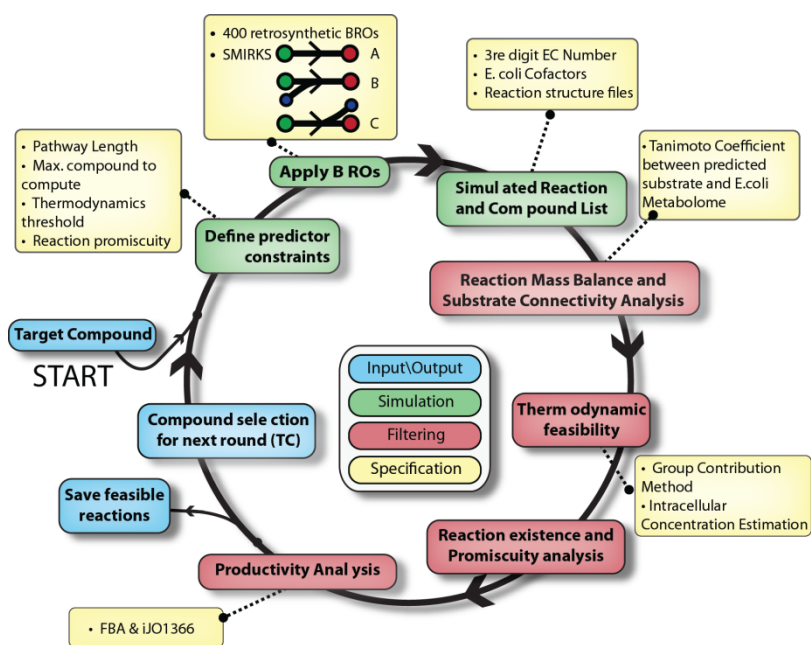
Figure 2-4: GEM-Path Algorithm. The iterative loop describes the synthetic pathway calculation process. Colored boxes show process inputs/outputs (blue), synthetic reaction generation (green), and filtering (red) procedures. Furthermore, detailed specifications were assigned to each stage (yellow). Starting with the target compound, BROs were applied and predictor constraints were defined. Mass-balanced generated reactions were filtered in terms of thermodynamic feasibility, reaction existence, and potential promiscuous activity. Then, productivity analysis was performed using FBA. Predicted reactions able to theoretically produce the target compound were saved and non-producing nodes (metabolites) were fed again into the algorithm to continue the pathway generation up to four iterations (reaction steps).

a)   Starting from the target metabolite, predictor constraints were set, such as maximal pathway length, metabolites to compute at each iteration, a thermodynamic threshold, and a reaction promiscuity threshold.

b)   Predefined BROs were applied to the target in a retrosynthetic manner for generating the corresponding substrates. After BROs application, the corresponding cofactors and third level EC numbers were assigned together with reaction structure files for further analysis. All predicted reactions were then checked for mass balance. If mass balance was not fulfilled, reactions were discarded from the process. Next, predicted metabolites were structurally compared against *E. coli*'s metabolome. Substrate dissimilarities were sorted in terms of the TC (see Methods), and an exact match analysis was performed for TCs equal to 0 since this does not necessarily mean that the compared molecules are the same. If the predicted metabolite matches any compound in the metabolome, FBA was performed in order to validate the potential production.

c) A thermodynamic analysis was performed by calculating the ΔrG (kJ/mol). Each predicted reaction was checked in terms of thermodynamic feasibility for existence of further reactions and potential promiscuity analysis. Reactions with ΔrG lower than or equal to 25 kJ/mol were defined as feasible reactions and saved to continue the checking process. The threshold was set based on estimated variability calculated elsewhere (Henry et al. 2007).

d) As shown in Figure 2-5, in order to determine reaction existence, predicted reactions were compared against BRENDA. The database was structured by lumping together all reactions with similar third level EC numbers. Each level contains known biochemical reactions with the corresponding four digit EC number association, reaction-organism association, and substrate structure file. The third level EC number association for the predicted reaction facilitates the identification of the third level EC class BRENDA sub group for substrate comparison. By calculating the TC, predicted substrates could be compared against all corresponding substrates present in the BRENDA subgroup. The results were sorted and analyzed starting with the most similar compound. Dissimilarities equal to 0 were structurally compared by performing an exact match comparison (see Methods). If the substrates were structurally similar, reaction cofactors were compared. In cases where the predicted reaction matches a reaction in BRENDA, a specific reaction–organism association was assigned to the reaction and the pathway prediction procedure was continued. Otherwise, a substrate promiscuity analysis was performed by considering the reaction-organism association substrate information. If the reaction is considered to be promiscuous, the algorithm saves the reaction, otherwise, it proceeds by analyzing the potential promiscuity for the next sorted substrate. In order to decide whether a reaction might be promiscuous or not, a reaction promiscuity score was calculated based on the similarity between the reaction native substrate and the predicted substrate (Figure 2-5, step 4). The reaction promiscuity score was calculated and analyzed by using *E. coli's* promiscuous reaction information from iJO1366 (Figure 2-1 and Figure 2-2). Based on the previous analysis, the reaction promiscuity score threshold was set to 1.2.
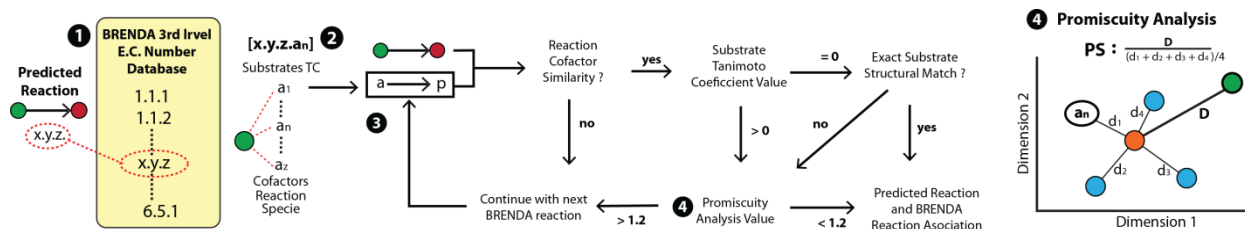
Figure 2-5: Reaction existence and promiscuity analysis. The first three steps outline the main processes for reaction existence and promiscuity analysis, while the forth step shows specifically how the promiscuity analysis was performed. First, for a predicted reaction the third level BRENDA EC number database was identified (yellow box). For each reaction in the databases structural information regarding substrates, cofactor uses and species were determined. Second, the predicted reaction substrate (green circle) was compared to the corresponding third level BRENDA EC number database substrates by calculating the TC. From bottom to top, substrate pairs of TCs were sorted in decreasing order. Third, starting from the lowest TC (a1) until a predicted reaction and BRENDA reaction association was found (an), an iterative decision making algorithm determines whether the predicted reaction exists in BRENDA or there is any reaction in the database able to show promiscuous activity. Fourth, when a specific reaction is sent to promiscuous analysis, non-specific substrates (blue circles) for the reaction/species association are assigned according to BRENDA databases. By calculating the TC between all of the substrates a reaction promiscuity space was generated. From this space, distances from the centroid for each substrates and promiscuity score were calculated.

After the filtering steps, only the 120 predicted compounds closest to *E. coli's* metabolome were allowed to continue the algorithm. This process was repeated 4 times, which means pathways of a maximal length of 4 were obtained. The GEM-Path algorithm overcame the disadvantages of previous methods by not setting a specific metabolite source for the target compound formation, instead leaving open the possibility to reach any metabolite in the metabolome. Furthermore, structural comparison gives the ability to focus on the retrosynthesis direction most similar to the corresponding region of the host metabolome. It should be noted that these characteristics could be extended to other organisms, predicting synthetic heterologous pathways in a host-context specific manner. After completion of this computational procedure, the resulting pathways were characterized and used for theoretical yield analysis under different strain, oxygenation, and substrate conditions. All of the predicted pathways are given in Figure A2-1. Thus, a comprehensive list of feasible biochemical pathways leading to the target compound formation was established.

2.4.2 Description of substrate and product selection

Important production capabilities of the synthetic pathways predicted by GEM-Path were assessed using the *E. coli* GEM. For theoretical yield analysis, three primary substrates were evaluated based on the cost and availability of suitable feedstock (Sauer et al. 2008; Vickers et al. 2012), *E. coli's* metabolic capacity for catalyzing such carbon sources, and unique design potential (e.g., glucose and fructose are not unique

and are examples of interconverted substrates with little to no cost to the cell, (Feist, Zielinski, Orth, Schellenberger, Herrgard and Palsson 2010)). The first two substrates were glucose and xylose, five- and six-carbon sugars, present in lignocellulosic biomass, representing about 40-50% and 20-30% by dry weight of plant material, respectively (Wyman et al. 2005). The use of this type of feedstock is expected to increase with the incentive to produce biofuel and bio–based chemicals (Perlack and Stokes 2011). The third substrate was glycerol, a three-carbon molecule and a byproduct of biodiesel production (Ma and Hanna 1999), whose availability is expected to increase in the coming years (Yang et al. 2012a). Additionally, three different starting strains and oxygenation conditions were analyzed for each product during the synthetic pathway calculations procedure. These are a wild-type strain under aerobic conditions, a wild-type strain under anaerobic conditions, and the 'ECOM' (*E. coli* cytochrome oxidase mutant) strain under aerobic conditions (Portnoy, Herrgard and Palsson 2008). The ECOM strain has the advantage of "aerobic fermentation" as the strain cannot use oxygen as a terminal electron acceptor. The list of targeted overproduction metabolites included 20 different bulk chemicals with biological production potential and precursors for commercially valuable chemical production are shown in Figure 2-6. The selection of the 20 target compounds was determined by evaluating reports generated by the US Department of Energy (Paster, Pellegrino, Carole, Energetics, U.S. Department of Energy and Renewable Energy 2003; Werpy, Petersen, Aden, Bozell, Holladay, White, Manheim, Eliot, Lasure and Jones 2004), which includes chemicals that are currently being produced on an industrial scale (Zeng and Sabra 2011) and metabolites that are described as precursors to or potential target biofuel compounds (Assary and Broadbelt 2011; Curran and Alper 2012; Fischer, Klein-Marcuschamer and Stephanopoulos 2008; Lee, Na, Park, Lee, Choi and Lee 2012). By comparing the target compound list with *i*JO1366 *E. coli's* metabolome, 4 out of 20 products were assigned as native and 16 as non-native. Synthetic pathways for native products were calculated in order to explore the possibility of more productive pathways for their synthesis.

2.4.3 Predicted pathways and reaction specifications

The synthetic pathway calculation procedure using GEM-Path was applied to all selected target compounds of interest and validated by comparing the output pathways

with previous computationally calculated and experimentally implemented pathways. In summary, 245 pathways, 221 reactions, and 59 non-native intermediate metabolites were calculated after 4 iterations of the algorithm (i.e., a maximal pathway length of 4).
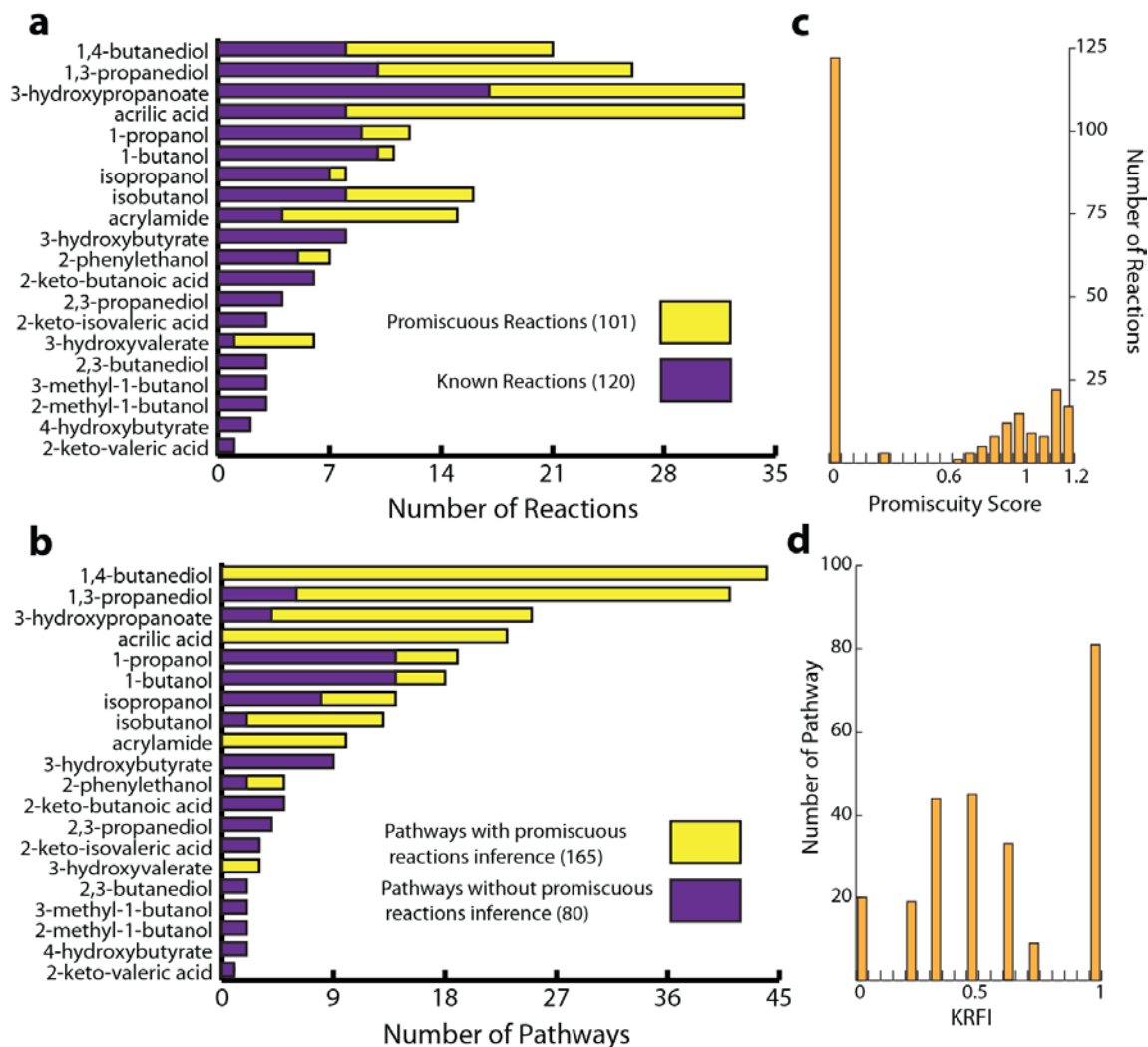
Figure 2-6: Predicted reaction and pathway analysis. For each target compound, the total number of synthetically generated reactions (a) and pathways (b) were plotted. 'Promiscuous' predicted and 'known' reactions are differentiated with yellow and purple sub segments, respectively. In the case of pathways, those containing one or more promiscuous reactions and those with no promiscuous reactions involved were differentiated by corresponding yellow and purple sub segments. Reaction promiscuity score distribution (c) and Known Reaction Fractional Index (KRFI) for each pathway (d) are also shown. A value of 0 for the promiscuity score indicates 'known' reactions, and a value of 1 for KRFI indicates the predicted pathway is constituted by only 'known' reactions.

In total, 25%, 39%, 28%, and 8% of the pathways were of length 4, 3, 2 and 1, respectively. For each product, pathways combining potential promiscuous reactions, already known reactions (i.e., in BRENDA), and different co-factor (i.e., using NAD^+ or $NADP^+$) uses were generated. In total, 44 different precursors from the native *E. coli* metabolome were determined that connected to the synthetic pathways. Furthermore,

42 gap-filling reactions interconnecting native *E. coli's* metabolites were identified which enabled production of a targeted compound. This set includes reactions which may be the reverse reaction of a native enzymatic step in the existing network or have completely unique chemistry acting on a native metabolite.

The number of reactions and pathways predicted using GEM-Path varied across the 20 target compounds analyzed. For 1,4-butanediol, 1,3-propanediol, 3-hydroxypropanoate, and acrylic acid, the number of reactions and pathways were the highest (Figure 2-6). In total, approximately 51% of all predicted reactions were categorized as 'known', which means that each predicted reaction has an exact biochemical reaction association according to BRENDA (Figure 2-6a purple sub-segments). Reactions represented by yellow sub-segments in Figure 2-6 correspond to predicted biochemical reaction steps assigned as 'promiscuous' from the promiscuity analysis. Furthermore, for each of these reactions, a potential reaction from BRENDA that might carry flux through the synthetic pathway was assigned. The promiscuity score distribution is represented in Figure 2-6c. As expected, all 112 'known' reactions were represented with a promiscuity index equal to 0 and 'promiscuous' reactions were distributed around 1.

A predicted pathway can be either entirely 'known' (Figure 2-6b purple sub-segments), meaning every reaction in the pathway has an exact biochemical reaction association according to BRENDA, or partially known, where one or more reactions in the pathway were predicted as 'promiscuous' (Figure 2-6b yellow subsegments). According to the classification in Figure 2-6b, all of the pathways able to generate 1,4-butanediol, acrylic acid, acrylamide, and 3-hydroxyvalerate in *E. coli* contain at least one promiscuous reaction. In order to analyze the fraction of known reactions present in a pathway, we defined the Known Reaction Fractional Index (KRFI) between 0 and 1, where 1 means that the pathway has been completely reconstructed from 'known' reactions and 0 means that it has been completely reconstructed from 'promiscuous' reactions. Based on the previous definition, 30% of all predicted pathways were entirely 'known'. In Figure 2-6 d, entirely 'known' pathways are represented with a KRFI equal to 1, and the rest of the pathways correspond to partially known pathways distributed from 0 to 1. In total, approximately 40% of the predicted reactions were oxidoreductases,

acting on hydroxyl or aldehyde groups with NAD^+ or $NADP^+$ acceptors. Carbon-oxygen and carbon-carbon lyases correspond to around 20%, and transferases, specifically CoA-transferases, were 8% of all of the reactions. This set of generic biochemical transformations details the chemical nature of the predicted reactions that most often enable the production of the targeted non-native compounds in *E. coli* (see Table A2-3).

2.4.4 GEM-Path validation

In order to validate the proposed algorithm, previous work examining computational and experimentally implemented heterologous pathways in *E. coli* were compared to the GEM-Path calculated pathways. According to a bibliographic search, 14 out of 20 target compounds were found to be referenced and targeted by patents or scientific publications. The maximum theoretical yield calculated by GEM-Path for the targeted compounds was then compared to production levels from the bibliographic search set (Table 2-1). In order to determine the production potential for the novel pathway calculated using GEM-Path, a maximum theoretical production comparison was performed for experimentally and computationally reported pathways (Table 2-1). The analysis was performed by calculating the target production ratio between the highest flux carrying novel pathways predicted by GEM-Path over the experimentally and computationally reported pathways. Simulations were run under aerobic and anaerobic conditions, by using glucose, xylose, and glycerol as a carbon source. Values over 1 indicated that GEM-Path's novel pathways have higher production potential than already referenced pathways. Considerable improvements over experimentally implemented pathways were found in the GEM-Path set, specifically under anaerobic conditions, for 1,4-butanediol, 1,3-propanediol, isopropanol, and 3-hydroxybutyrate on various substrates. Distinct, but equal yield pathways were calculated for 1,3-propanediol and 1-butanol. In addition, already known implemented pathways for 1-propanol, 2-phenylethanol, 2,3-propanediol, 2,3-butanediol, 3-methyl-butanol, 2-methyl-butanol, and 4-hydroxybutyrate were found (Table 2-1). These findings revealed that GEM-Path calculated pathways contained experimentally-implemented pathways found in the literature screen and that the selected reaction rules were able to represent the known biochemistry and serve as validation of the approach.

Table 2-1: Comparison of GEM-Path predictions to previously identified pathways from literature

Target Compound	Experimental anaerobic			Experimental aerobic			Ref.	Computational anaerobic			Computational aerobic			Ref.
	gluc	xyl	glyc	gluc	xyl	glyc		gluc	xyl	glyc	gluc	xyl	glyc	
1,4-butanediol	1.3	1.3	1.3	1.1	1.1	1.1	(Yim, Haselbeck, Niu, Pujol-Baxley, Burgard, Boldt, Khandurina, Trawick, Osterhout, Stephen, Estadilla, Teisan, Schreyer, Andrae, Yang, Lee, Burk and Van Dien 2011)	1.3	1.3	1.3	1.1	1.1	1.1	(Yim, Haselbeck, Niu, Pujol-Baxley, Burgard, Boldt, Khandurina, Trawick, Osterhout, Stephen, Estadilla, Teisan, Schreyer, Andrae, Yang, Lee, Burk and Van Dien 2011)
1,3-propanediol	1.2	1.3	1	1	1	1	(Laffend et al. 1997; Nagarajan and Nakamura 1998; Tang et al. 2009; Zeng and Sabra 2011)	---	---	---	---	---	---	---
3-hydroxypropanoate	1	1	1	1	1	1	(Lynch 2011; Suthers and Cameron 2005; Wang et al. 2012)	1	1	1	1	1	1	(Henry, Broadbelt and Hatzimanikatis 2010)
1-propanol	1	0.9	0.9	1	1	1	(Pharkya 2011; Shen and Liao 2008; Shen and Liao 2013)	2.4	2.8	3.3	1.2	1.2	1.1	(Cho, Yun, Park, Lee and Park 2010)
1-butanol	1	1	1	1	1	1	(Atsumi et al. 2008; Bramucci et al. 2008; Lee and Park 2008; Shen et al. 2011)	1	1	1	1	1	1	(Cho, Yun, Park, Lee and Park 2010)
isopropanol	1.2	1.2	1.9	1	1	1.1	(Hanai et al. 2007; Jojima et al. 2008; Pharkya 2011)	---	---	---	---	---	---	---
isobutanol	0.8	0.7	0.8	1	1	1	(Atsumi et al. 2010; Trinh 2012)	0.8	0.7	0.8	1	1	1	(Cho, Yun, Park, Lee and Park 2010)
3-hydroxybutyrate	1.2	1.2	1.5	1	1	1	(Tseng et al. 2009; Valentin and Dennis 1997)	---	---	---	---	---	---	---
2-phenylethanol	0.9	0.9	0.9	1	1	1	(Hwang et al. 2009; Koma et al. 2012)	0.9	0.9	0.9	1	1	1	(Cho, Yun, Park, Lee and Park 2010)

2,3-propanediol	1	1	1	1	1	1	(Altaras and Cameron 1999; Soucaille et al. 2008)	---	---	---	---	---	---	---
2,3-butanediol	1	1	1	1	1	1	(Ji et al. 2011; Lu et al. 2012b; Yan et al. 2009)	---	---	---	---	---	---	---
3-methyl-1-butanol	1	1	1	1	1	1	(Connor et al. 2010)	1	1	1	1	1	1	(Cho, Yun, Park, Lee and Park 2010)
2-methyl-1-butanol	1	1	1	1	1	1	(Cann and Liao 2008)	1	1	1	1	1	1	(Cho, Yun, Park, Lee and Park 2010)
4-hydroxybutyrate	1	1	1	1	1	1	(Zhou et al. 2012)	---	---	---	---	---	---	---

For each target compound, the maximum theoretical productivity ratio between novel pathways generated by GEM-Path and experimentally implemented or computationally generated pathways is shown. Empty spaces (---) indicate that no referenced pathways for the corresponding target compound were found.

For the synthetic design of biochemical pathways, much attention has been focused on BRO-based computational tools (Medema et al. 2012). As such, the pathways predicted from GEM-Path were compared against computationally-predicted pathways from three different BRO based algorithms; BioPath for the production of 1,4-butanediol (Yim, Haselbeck, Niu, Pujol-Baxley, Burgard, Boldt, Khandurina, Trawick, Osterhout, Stephen, Estadilla, Teisan, Schreyer, Andrae, Yang, Lee, Burk and Van Dien 2011), BNICE for the production of 3-hydroxypropanoate (Hatzimanikatis, Li, Ionita, Henry, Jankowski and Broadbelt 2005), and the one developed by Cho *et al.*, for the production of several alcohols (Cho, Yun, Park, Lee and Park 2010).

The first comparison was for the synthetic pathway prediction of 1,4-butanediol by using the BioPath algorithm. When analyzing individual reactions, 91% off all reactions were able to be predicted by GEM-Path independently. Furthermore, through FBA analysis, novel pathways generated with GEM-Path were able to achieve higher theoretical productivity compared to BioPath reported pathways. Specifically for pathways 13 and 14 (see 1,4-butanediol pathways map in Supplemental Figure 9), under aerobic condition and using glucose, xylose, and glycerol, a 10% theoretical productivity increase over BioPath predicted pathway was calculated. Moreover, by using the same substrates under anaerobic conditions, an approximately 30% increase over BioPath predicted pathways was calculated. The second case studied was for the synthetic pathway prediction of 3-hydroxypropanoate by using the BNICE algorithm.

This framework is able to produce all thermodynamically feasible pathways from a source metabolite to a target compound. In this case, GEM-Path was able to generate 11% of all predicted pathways by this algorithm, and 87% of all reactions. This result was expected since both algorithms share similar BROs. By applying the reaction existence and promiscuity analysis based on BRENDA, GEM-Path was able to constrain the predicted pathways by reporting only a feasible subset of pathways. According to FBA simulations, novel pathways generated with GEM-Path were able to achieve the same maximum theoretical production rates compared to BNICE generated pathways, specifically for pathways 12, 13, and 3 (see 3-hydroxypropanoate pathways map in Supplemental Figure 9). When using xylose and glucose as substrates, production rates were 76% higher than glycerol. Under aerobic conditions, no substantial increments in theoretical production rates between GEM-Path and previously generated BNICE pathways were identified. Finally, the third case analyzed after the synthetic pathway generation was Cho, et. al. Here, the author introduces a novel scoring algorithm in order to extract the most feasible pathways. The framework was validated for the production of 1-propanol, 1-butanol, 2-methyl-1-butanol, 3-methyl-1-butanol, isobutanol, and 2-phenylethanol from a variety of 2-ketoacids. When comparing the results between GEM-Path and Cho's predictions for each product, the same pathways were found for each case. Still, according to the simulations, none of the remaining pathways predicted by GEM-Path were able to achieve the production rates of pathways previously generated by Cho´s algorithm. Pathway and reaction prediction discrepancies were due to the filtering procedure, specifically during the promiscuity analysis, where only the most promising reactions were allowed to constitute a pathway in GEM-Path. However, of note is that the vast majority of reactions predicted in the referenced work was also predicted with GEM-Path. Specifically, GEM-Path was able to simulate 92% and 32% of all reactions and pathways, respectively. Furthermore, discrepancies arose due to a lack of connectivity between the host metabolic network and the predicted synthetic pathways in the referenced work and also from the predefined pathway length which allowed a maximum pathway length of four. A number of differences can be the result of the GEM-Path algorithm immediately stopping the search through each branch when it reaches the metabolome; the three other algorithms mentioned above do not have this stipulation.

When comparing GEM-Path with other computational tools (see Table A2-1), its most characteristic features are its capability to shrink the biochemical reaction solution space by calculating the closest pathways to the metabolome and its ability to select mechanistically-feasible reactions from BRENDA. These properties rely on the integration of the promiscuity analysis and GEMS into the reaction prediction algorithm. Furthermore, GEM-Path is able to systematically integrate physiological conditions (e.g., carbon source and oxygenation) into the pathway generation procedure, allowing for the consideration of the active content in a given condition and not reactions or nodes that cannot be reached in a desired media condition. Furthermore, when comparing GEM-Path to previous tools, it shows a wider predictive capacity as it, i) takes into account more cofactors, ii) does not constrain the search to only one compound source, instead every metabolite in the metabolome might work as a source, and iii) allows generation of anabolic and catabolic reactions. Nevertheless, some solutions might be hindered as not all nodes (i.e., predicted compounds) were allowed to continue through the prediction algorithm when compared to the *E. coli* metabolome. However, GEM integration into GEM-Path allows the algorithm to find more than one precursor present in the metabolic network without constraining the search to only one compound.

2.4.5 Theoretical Yield analysis of the production potential in *E. coli*

The production potential landscape in *E. coli* was outlined by calculating and plotting the maximum theoretical yield for each target compound in terms of carbon moles captured (i.e., C-mol). Simulations were performed by combining all predicted pathways with the corresponding substrate utilization and oxygenation conditions (see Methods). In total, 2205 flux balance analysis (FBA) combinations were calculated (Figure 2-7). Maximum theoretical yields (Figure 2-8a-c) and the corresponding pathways were tabulated for each target compound (Table 2-2).

Results were grouped together based on strain and oxygenation conditions and a yield interval was applied to plot the number of pathways for different substrates (Figure 2-8a-c). Furthermore, in order to determine the most efficient subsystem for product formation, results were clustered in terms of yield and *E. coli's* precursor metabolic

subsystems (Figure 2-8d). Overall, the average yields for WT/aerobic, ECOM/aerobic, and WT/anaerobic were 0.68, 0.53, and 0.38, respectively.
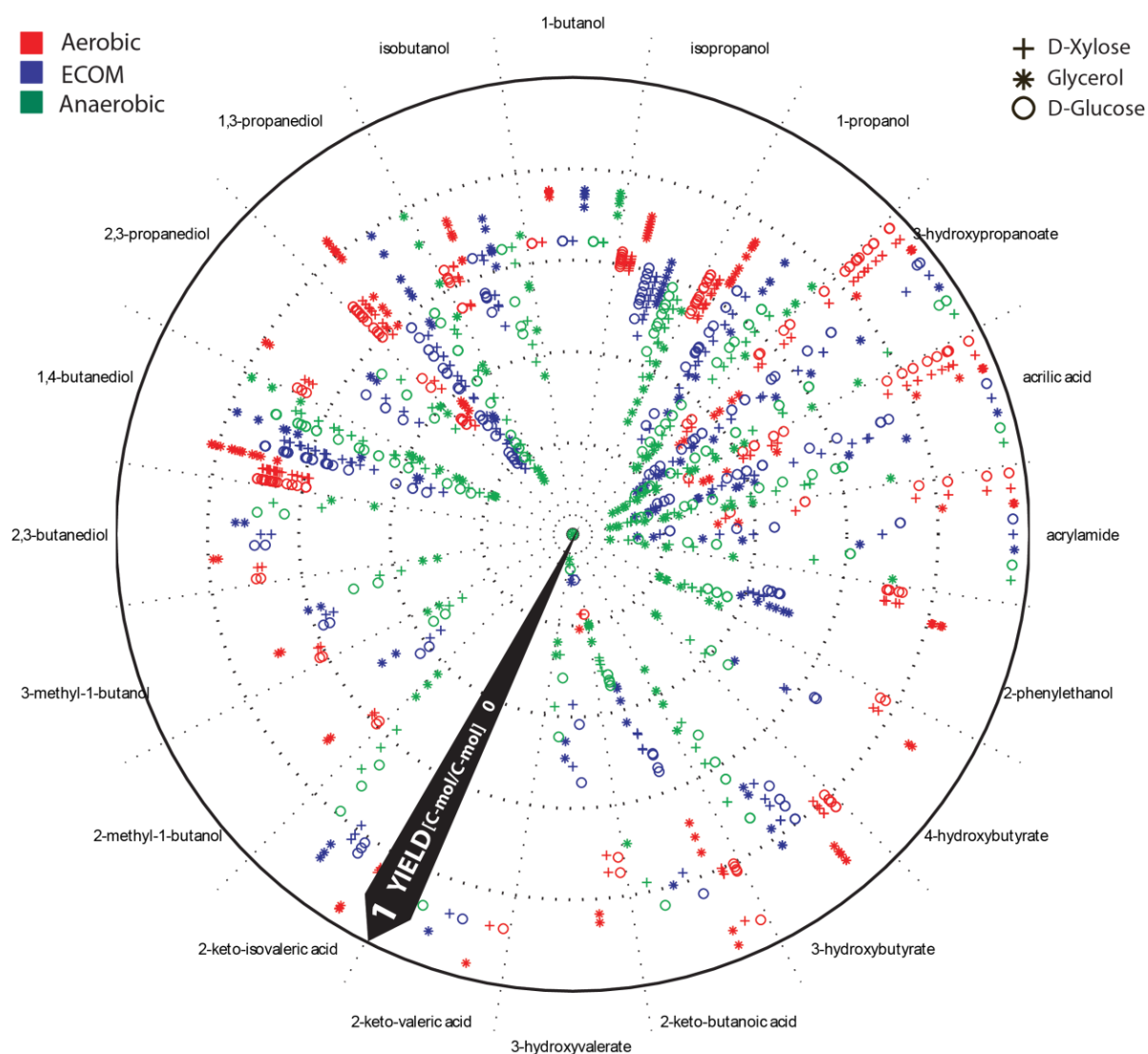
Figure 2-7: Theoretical maximal yield. The circumference is divided in 20 different segments. Each one of them corresponds to a specific target compound. The yield is represented along the radius, where the center and the perimeter represent yields equal to 0 and 1 respectively. The theoretical maximal yield for different strain/substrate conditions was plotted for each target compound. Colored points represent the strain condition for wild type/aerobic (red), ECOM/aerobic (blue), and wild type/anaerobic (green). The shape defines a specific substrate use for xylose (+), glycerol (*) and glucose (o). The pink box shows a schematic representation of the production envelope and describes the points that were represented in the diagram, specifically the maximum theoretical yield at 0.1 growth rate.

By defining the ECOM/aerobic condition as an intermediate state of aerobiosis between WT/aerobic and WT/anaerobic, a correlation between the aerobiosis state of the cell and the production potential can be drawn. As shown in Figure 2-8a-c, a pronounced displacement of maximum theoretical yield distributions towards lower yields is directly correlated with the extent of anaerobiosis. This trend is also shown in

74

Figure 2-8d, where a gradual shift from lower anaerobic yield to higher aerobic yields can be visualized. Furthermore, this pattern is shown together with a preference for glycerol as a substrate under aerobic conditions and for glucose under anaerobic conditions. Pathways near central carbon metabolism subsystems are able to achieve higher yields (Figure 2-8d). The specific analysis for each strain / oxygenation condition was described below.

2.4.5.1 Wild type aerobic

For wild type *E. coli* under aerobic condition, 77% of the 2205 simulations were greater than 0.6 C-mol Product/C-mol Substrate (Figure 2-8a) and the overall average yield was 0.68. All products can potentially be made from either glucose, xylose, or glycerol under aerobic conditions. Glycerol is found to be the most efficient substrate for aerobic production. The average yields were identical for glucose and xylose (aprox. 0.66) and 11% higher on glycerol. For maximum theoretical yield (Table 2-1) the same trend as describe above was identified for most of the products, where glycerol was found to be the most efficient substrate. 3-hydroxypropanoate, acrylic acid, and acrylamide, were shown to be generated at similar yields under any of the three predefined substrates. Furthermore, these compounds showed one of the highest yields of the list, together with 1,3 propanediol, 1,4 butanediol, 2-keto-isovaleric acid, 2-keto-valeric acid, 3-hydroxyvalerate, 2-keto-butanoic acid, 3-hydroxybytyrate and 2-phenylethanol, where the maximum production potential was achieved using glycerol as a substrate (Table 2-2).

2.4.5.2 ECOM aerobic

Compared to the wild type aerobic, the ECOM strain shows a shift in the simulation distribution pattern (Figure 2-8b), where 69% of all the simulations are between 0.4 and 0.8. For higher yields, glycerol is the most efficient substrate. By examining the average yields, glycerol was only about 4% higher than glucose, and glucose 5% higher than xylose. 3-hydroxypropanoate, acrylic acid, acrylamide, 2-keto-isovaleric acid and 2-keto-valeric acid yields were found to be higher than 0.8 with any

substrate. With the exception of 2-ketobutanoic acid, 3-hydroxyvalerate and 4-hydroxybutyrate, all products archive the highest yield by using glycerol as a substrate.

Table 2-2: Targeted compounds and theoretical maximum yield analysis

Target Compound	Native or non-native	No. of carbons	No. of computed pathways	No. of unique reactions in each pathway	aerobic (C-mol Yield / pathway ID)			aerobic ECOM (C-mol Yield/ pathway ID)			Anaerobic (C-mol Yield/ pathway ID)		
					gluc	xyl	glyc	gluc	xyl	glyc	gluc	xyl	glyc
1,4-butanediol	non-native	4	44	21	70 / 13	69 / 13	82 / 13	70 / 13	66 / 13	78 / 13	70 / 13	66 / 13	78 / 14
1,3-propanediol	non-native	3	41	26	69 / 16	69 / 14	84 / 7	57 / 14	54 / 16	79 / 34	57 / 16	52 / 16	79 / 7
3-hydroxypropanoate	native	3	25	33	96 / 12	96 / 13	97 / 3	96 / 12	96 / 13	97 / 12	96 / 12	96 / 13	71 / 12
acrilic acid	non-native	3	23	33	96 / 2	96 / 2	97 / 1	96 / 2	96 / 2	97 / 2	96 / 1	96 / 2	71 / 1
1-propanol	non-native	3	19	12	64 / 3	64 / 2	75 / 3	64 / 1	64 / 2	75 / 1	64 / 3	64 / 4	75 / 1
1-butanol	non-native	4	18	11	64 / 5	64 / 15	75 / 10	64 / 4	64 / 4	75 / 10	64 / 11	64 / 5	75 / 5
isopropanol	non-native	3	14	8	63 / 4	62 / 4	72 / 1	61 / 3	59 / 3	63 / 4	58 /3	56 / 3	54 / 1
isobutanol	non-native	4	13	16	64 / 2	64 / 1	74 / 1	64 / 2	64 / 1	72 / 1	64 / 1	64 / 1	66 / 1
acrylamide	non-native	3	10	15	96 / 3	96 / 3	97 / 3	96 / 3	96 / 3	97 / 3	96 / 3	96 / 2	71 / 2
3-hydroxybutyrate	non-native	4	9	8	83 / 5	82 / 1	93 / 5	79 / 2	77 / 1	82 / 1	73 / 2	68 / 1	49 / 1
2-phenylethanol	non-native	8	5	7	73 / 9	73 / 9	83 / 9	47 / 1	44 / 2	50 / 6	36 / 5	31 / 1	36 / 6
2-keto-butanoic acid	native	4	5	6	94 / 1	93 / 1	97 / 1	84 / 1	78 / 1	80 / 1	84 / 1	78 / 1	69 / 1
2,3-propanediol	non-native	3	4	4	68 / 1	68 / 1	79 / 2	55 / 1	51 / 1	56 / 2	55 / 1	51 / 1	40 / 1
2-keto-isovaleric acid	native	5	3	3	86 / 3	85 / 3	97 / 3	84 / 3	82 / 3	90 / 3	79 / 3	69 / 3	49 / 3
3-hydroxyvalerate	non-native	5	3	6	75 / 2	74 / 2	85 / 2	54 / 2	51 / 2	49 / 2	44 / 2	38 / 2	27 / 2
2,3-butanediol	non-native	4	2	3	70 / 1	69 / 1	79 / 1	70 / 1	68 / 1	74 / 1	69 / 1	66 / 1	53 / 1
3-methyl-1-butanol	non-native	5	2	3	62 / 1	61 / 1	70 / 1	58 / 1	57 / 1	60 / 1	49 / 1	43 / 1	53 / 1
2 -methyl-1-butanol	non-native	5	2	3	60 / 1	59 / 1	70 / 1	42 / 1	39 / 1	50 / 1	36 / 1	31 / 1	36 / 1
4-hydroxybutyrate	native	4	2	2	78 / 1	77 / 1	88 / 1	64 / 1	58 / 1	45 / 1	44 / 1	39 / 1	26 / 1
2-keto-valeric acid	non-native	5	1	1	88 / 1	88 / 1	97 / 1	88 / 1	87 / 1	92 / 1	88 / 1	87 / 1	70 / 1

For each target compound, maximum theoretical yields (C-mol) were reported for different strain and substrate conditions. Shown next to the yield is the corresponding pathway ID shown in Figure A2-1.

2.4.5.3 Wild type anaerobic

Two different kinds of shift from aerobic to anaerobic, passing through aerobic ECOM conditions were identified (Figure 2-8c). As shown for the ECOM strain, the first shift shows a displacement of the overall simulation towards lower yields. Approximately

74% of the simulations were between 0 and 0.6. And the second shift, regarding substrate preference, shows a direct association between the use of glucose as a substrate and higher yields, compared with previous analysis where glycerol was associated with higher yields. Specifically, in this case glucose was 9% higher than xylose, and xylose 4% higher than glycerol. The highest yields under anaerobic conditions were achieved for 3-hydroxypropanoate, acrylic acid, and acrylamide, by indiscriminately using glucose or xylose as a substrate. The highest yields for 1,3-propanediol, 1,4-butandeiol, 1-propanol, 1-butanol and isopropanol were generated on glycerol.
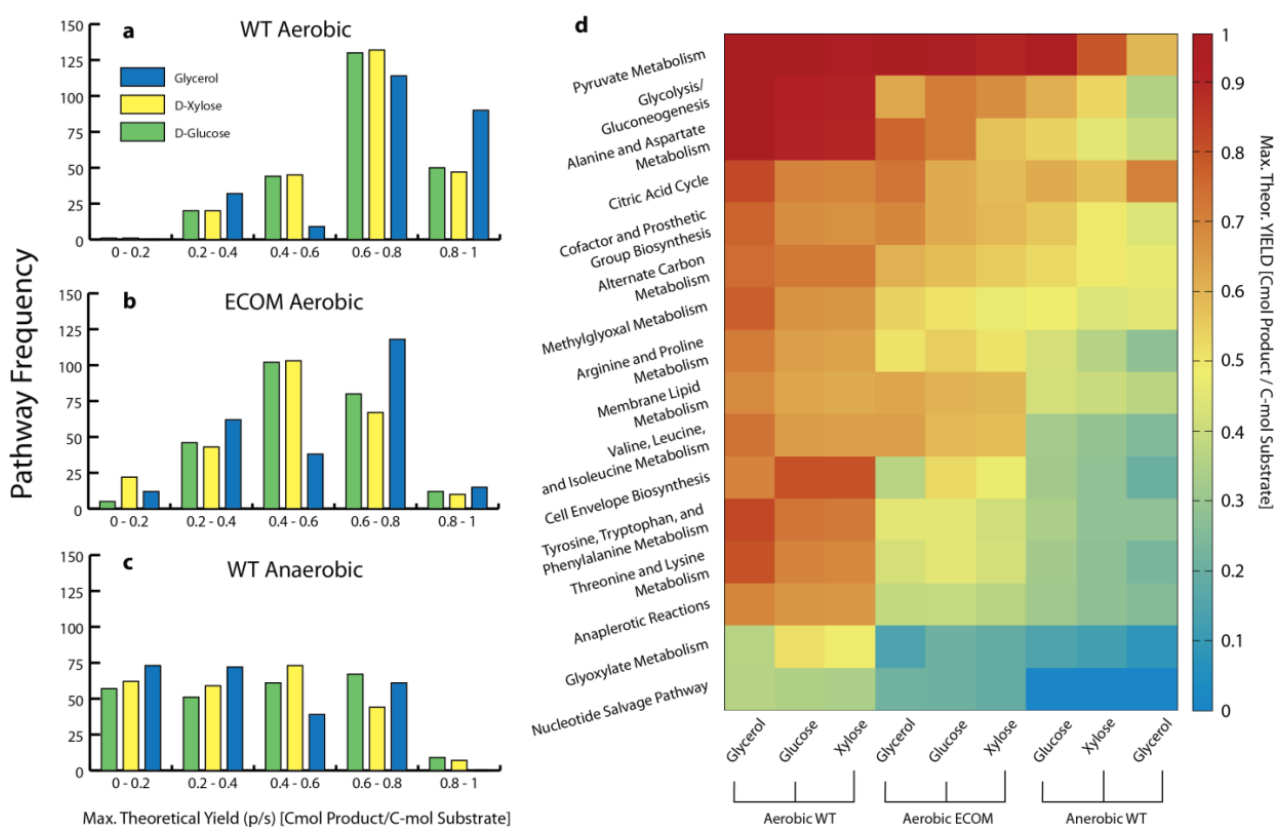
Figure 2-8: Theoretical maximal yield distribution for different strain/substrate conditions and subsystems. FBA was performed for each predicted pathway and strain/substrate condition using the E. coli GEM. At each yield interval, the number of pathways was plotted for each specific substrate: glucose (green), xylose (yellow), and glycerol (blue). This analysis was performed for wild type/aerobic (a), ECOM/aerobic (b), and wild type/anaerobic (c). In total, 2205 simulations were performed. (d) The subsystem form which each precursor metabolite were from were determined by analyzing the reaction that connects the network with the corresponding synthetic pathway. The yield average was calculated for each precursor subsytem and clustered by each strain / substrate condition.

2.4.6 Strain Design

Utilizing the synthetic pathways identified for each target compound, strain design simulations were performed to determine if reaction knock-outs could increase production. The predicted synthetic pathways were independently incorporated into the *E. coli* GEM and further model preprocessing was executed according to a previously developed approach (Feist, Zielinski, Orth, Schellenberger, Herrgard and Palsson 2010). Growth-coupled designs, which couple the optimal production of biomass and energy generation to the production of the compound of interest, were chosen as objectives for the strain design performed here. A combination of the RobustKnock (Tepper and Shlomi 2010) and the GDLS (Lun, Rockwell, Guido, Baym, Kelner, Berger, Galagan and Church 2009) algorithms with the conditioned model of *i*JO1366 (Orth, Conrad, Na, Lerman, Nam, Feist and Palsson 2011) was used. First, RobustKnock was utilized to design strains of *E. coli* for each target/substrate/oxygenation combination for a maximum of two and three reaction knock-outs allowed. GDLS was used in order to decrease computational time and to evaluate the maximum theoretical growth-coupled yield for four knock-outs.

All reactions which were identified in the strain design process for elimination were collected and analyzed (seeTable A2-4). Growth-coupled designs could be found for 1271 different target/substrate/oxygenation/knock-out combinations (Campodonico et al. 2014). Overall, this number was 19% out of the 6615 possible conditions examined. The results of the design analysis are given in Table 2-3. Result landscapes of maximum growth-coupled yield for each target compound are shown in Figure 2-10. Overall, production could be growth-coupled in 75% of the targeted compounds and 43% of all predicted pathways. Targets which could not be growth-coupled were 2,3-propanediol, 3-methyl-1-butanol, 2-methyl-1-butanol, 4-hydroxybutytare, and 2-phenylethanol. In total, 84 different reaction knock-outs were identified across all selected target reactions, some of them participating more frequently in strain designs. Pyruvate formate lyase and ATP synthase occurred 12 times more often than the average 44 knock-outs per reaction in all designs (1271). Pyruvate kinase occurred 7.4 times more and acetate kinase, pyruvate dehydrogenase, triose-phosphate isomerase, glucose-6-phosphate isomerase, ribulose 5-phosphate 3-epirase, glutamate

78

dehydrogenase, alcohol dehydrogenase, and malate dehydrogenase occurred approximately 2.8 times more often than the average. As stated earlier (Feist, Zielinski, Orth, Schellenberger, Herrgard and Palsson 2010), this uneven distribution of reaction knock-out occurrences suggests that certain reactions are critical for diverting carbon flux.
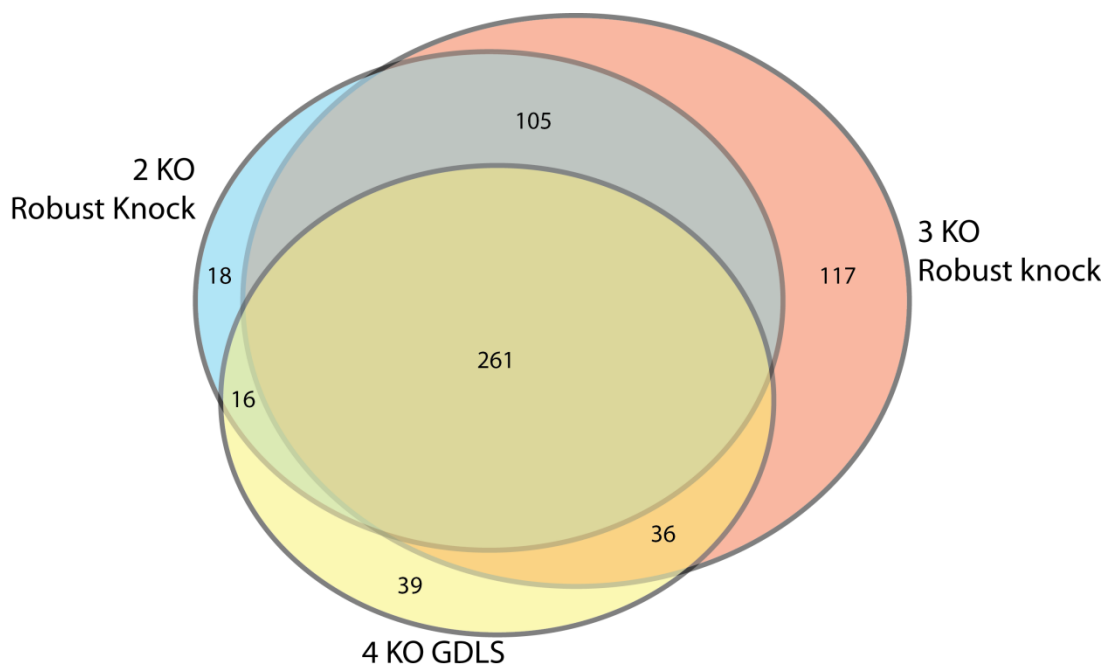
Figure 2-9: Growth-Coupled Venn Diagram sorted by number of knock-outs. Shown in the plot is a Venn diagram of instances of overlapping designs found for each oxygen condition / substrate / predicted pathway combination. For example, there were 261 conditions where a strain design was determined for the same conditions with 2, 3, or 4 reaction knock-outs. In theory, the 2 knock-out designs should be a subset of the 3 knock-out designs, however, algorithm run time likely prevented this from occurring in all instances. The same could be said for 3 knock-out designs being a subset of 4 knock-out designs. If one were to run the algorithm longer for each condition, this property may be achieved.

Approximately 8% of all designs were above a molar yield for carbon of 0.6, and this corresponded to designs for 9 out of 20 targeted compounds. When comparing different oxygenation conditions, most of the designs were calculated under wild type/anaerobic conditions (40%), followed by wild type/aerobic (33%), and ECOM/aerobic (23%). The highest average yield for all possible designs was calculated for wild type/anaerobic as being approximately 17% and 91% higher than the ECOM/aerobic and wild type aerobic, respectively. The predominant substrate for growth-coupled designs was glycerol (40%), then xylose (32%), and glucose (28%). The average growth-coupled yield distribution follows a similar trend where glycerol was 21% and 56% higher than xylose and glucose, respectively. According to the previous study

(Feist, Zielinski, Orth, Schellenberger, Herrgard and Palsson 2010), the larger the number of allowable knock-outs for a given target compound, the greater the maximum achievable yield. This trend was observed when comparing RobustKnock for two and three knock-out designs, where the average maximum growth-coupled yield was 21% higher for three versus two knock-outs.
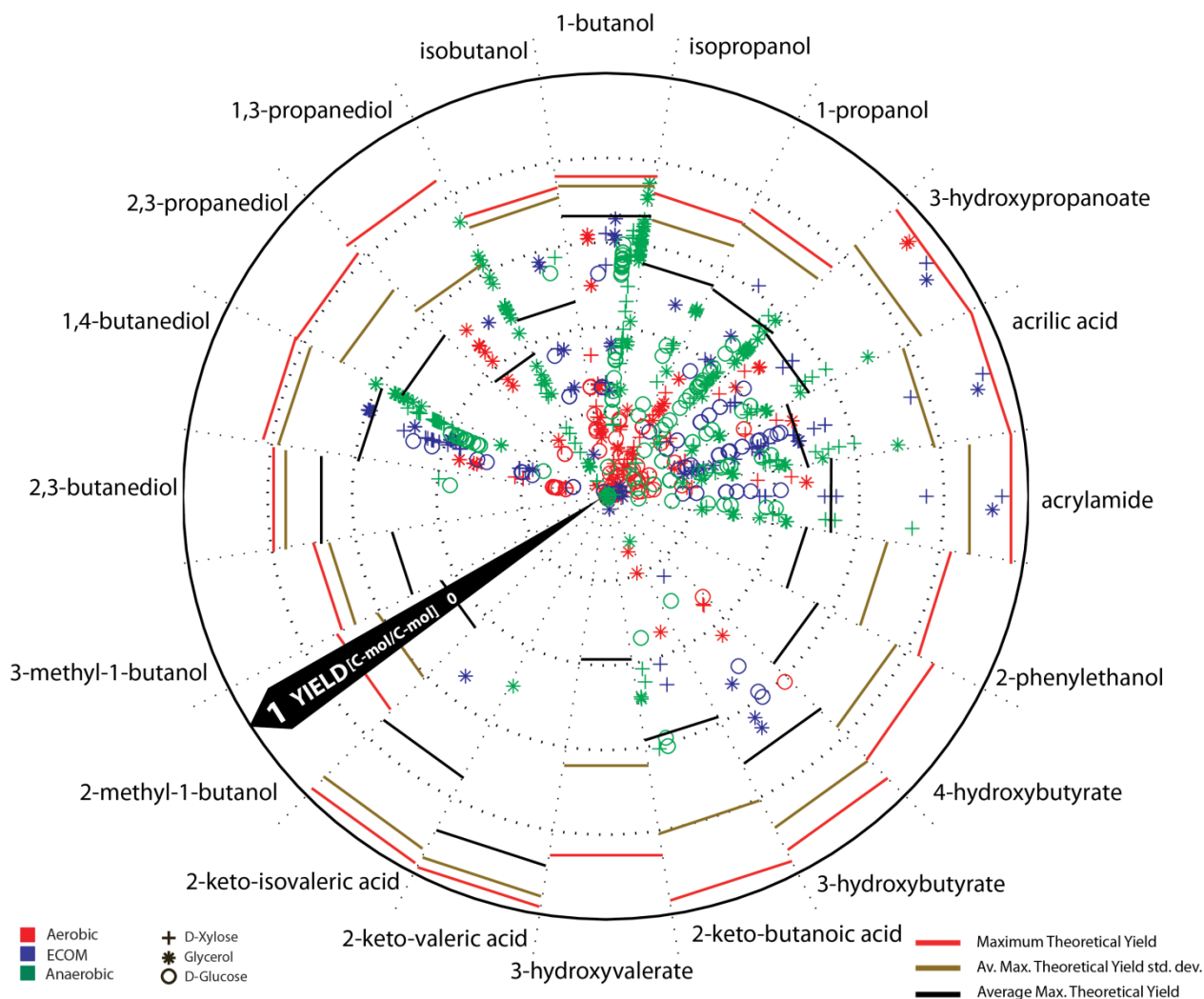
Figure 2-10: Analysis of predicted yields for identified strain designs. The circumference of the plot is divided into 20 different segments, each corresponding to a specific target compound. The yield is represented along the radius, where the center and the perimeter represent C-mol yields equal to 0 and 1, respectively. Two different kinds of results are plotted in this diagram. First, the theoretical maximal growth-coupled yield for different knock-out and strain/substrate combinations were plotted for each target compound. Colored points represent the strain condition for wild type/aerobic (red), ECOM/aerobic (blue), and wild type/anaerobic (green). The shape defines a specific substrate use for xylose (+), glycerol (*), and glucose (o). The second set of results corresponds to the average maximal theoretical yield (black line) for each compound (each compound can have multiple predicted pathways) with the corresponding standard deviation (brown line) added to this mean. These values were calculated from the theoretical yield analysis, where all the simulations regarding strain/substrate conditions were taken into account. Finally, the highest maximal theoretical yield value is represented by the red line.

When comparing GDLS strain designs for four knock-outs to RobustKnock strain designs for three knock-outs, 37% of all designs were able to achieve higher C-mol yield when allowing four knock outs. Furthermore, for 9% of all designs, GDLS was able to find a growth-coupled design when RobustKnock could not (see Figure 2-9). However, when comparing GDLS output for four knock-outs to RobustKnock for 3 knock-outs, no increase in the average maximum achievable yield was observed (i.e., for 519 3-KO designs, the average C-mol yield was 0.35 whereas for 352 4-KO designs, the average yield was 0.34). This can be because the GDLS algorithm is not guaranteed to find an optimal solution (Lun, Rockwell, Guido, Baym, Kelner, Berger, Galagan and Church 2009), but this value could increase given a longer run time or different starting parameters.
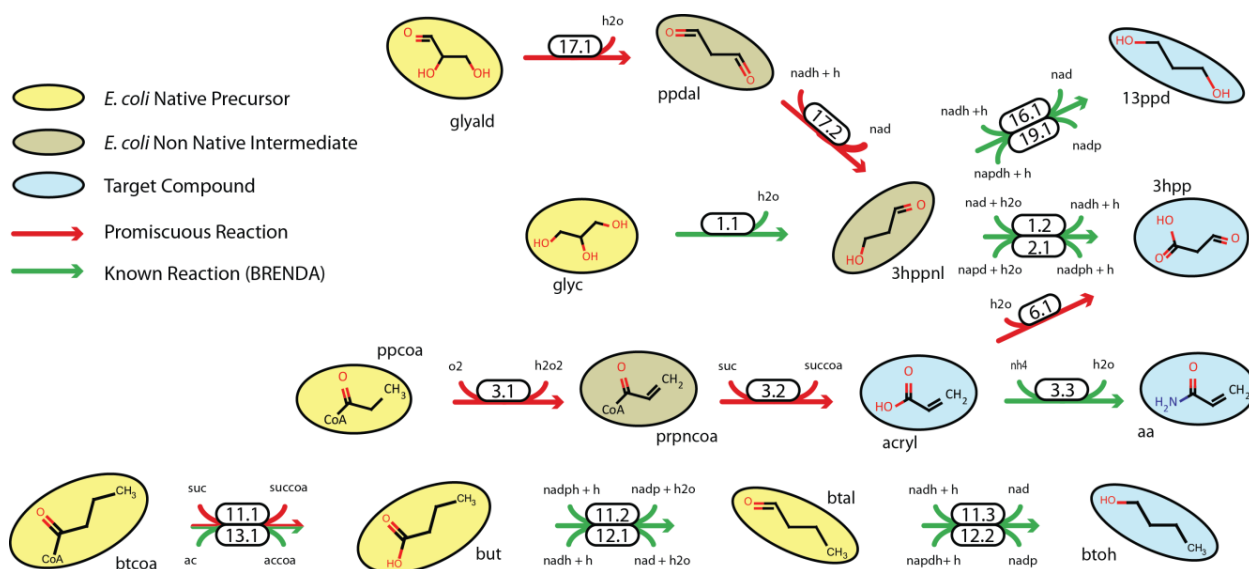
Figure 2-11: A pathway map of synthetic pathways identified in high yield growth-coupled designs. The top 20 growth-coupled pathways carrying high fluxes for product formation are shown. Reactions are represented by arrows. The color specifies whether the reaction is predicted as promiscuous (red) or known (green). Compounds might be E. coli´s native metabolite (yellow), nonnative intermediate (grey), or a target compound (light blue). Each reaction has a two digit index which is specified in Table A2-2. Compound abbreviations are: 3hppl: 3-hydroxypropanal, acryl: acrylic acid, prpncoa: propenoyl-CoA, ppdal: propanedial, aa: acrylamide, btoh: 1-butanol, 13ppd: 1,3-propanediol, 3hpp: 3-hydroxypropanoate, ppcoa: propanoyl-CoA, ppap: propanoyl phosphate, mmcoa-S: (S)-methylmalonyl-CoA, but: butanoic acid, and btcoa: butanoyl-CoA.

Overall, the growth-coupled yield analysis revealed a positive correlation between the total number of strain designs and the number of predicted pathways for each target compound (Table 2-3). The same correlation was observed when comparing the number of independent growth-coupled pathways and the number of predicted pathways for each target compound. When examining specific targeted products,

approximately 40% of all predicted pathways were able to couple the target compound production with growth across any of the predefined oxygenation/substrate/knock-out conditions. Strain design C-mol yield averages for the production of 1-butanol were higher than the corresponding medial yield for other target compounds.

Table 2-3: Predicted Yields for Growth-coupled strain designs by production Interval and product

Yield Interval for growth-coupled designs	Total No. of Strain Designs	Oxygenation Nr. of designs/Avg. C-mol Yield			Substrate Nr. of designs/Avg. C-mol Yield			Knock Outs Nr. of designs/Avg. C-mol Yield		
		aerobic	ECOM	anaerobic	glucose	xylose	glycerol	2 KO	3 KO	4 KO
0.8 - 1.0	10	2 / 0.93	8 / 0.93	0 / 0	0 / 0	5 / 0.94	5 / 0.92	0 / 0	8 / 0.93	2 / 0.94
0.6 - 0.8	89	17 / 0.61	16 / 0.64	56 / 0.64	3 / 0.61	17 / 0.64	69 / 0.64	15 / 0.65	43 / 0.64	31 / 0.63
0.4 - 0.6	342	35 / 0.47	70 / 0.49	237 / 0.50	38 / 0.51	121 / 0.48	183 / 0.51	105 / 0.5	146 / 0.50	91 / 0.48
0.2 - 0.4	461	123 / 0.29	130 / 0.32	208 / 0.32	169 / 0.31	138 / 0.31	154 / 0.32	133 / 0.31	185 / 0.31	143 / 0.32
0 - 0.2	369	248 / 0.1	59 / 0.05	62 / 0.08	149 / 0.1	121 / 0.10	99 / 0.06	147 / 0.07	137 / 0.11	85 / 0.09
Overall	1271	425 / 0.21	283 / 0.34	563 / 0.4	359 / 0.25	402 / 0.32	510 / 0.39	400 / 0.29	519 / 0.35	352 / 0.34
Percentage	19%	33%	22%	44%	28%	32%	40%	31%	41%	28%

Target Compound	Total No. of Strain Designs	Oxygenation Nr. of designs/Avg. C-mol Yield			Substrate Nr. of designs/Avg. C-mol Yield			Knock Outs Nr. of designs/Avg. C-mol Yield		
		aerobic	ECOM	anaerobic	glucose	xylose	glycerol	2 KO	3 KO	4 KO
acrylamide	77	23 / 0.13	26 / 0.31	28 / 0.33	29 / 0.20	25 / 0.39	23 / 0.21	27 / 0.22	31 / 0.30	19 / 0.28
acrilic acid	152	38 / 0.13	67 / 0.30	47 / 0.31	56 / 0.21	49 / 0.34	47 / 0.24	44 / 0.22	65 / 0.28	43 / 0.27
3-hydroxypropanoate	110	34 / 0.19	36 / 0.28	40 / 0.32	38 / 0.18	36 / 0.36	36 / 0.26	35 / 0.22	45 / 0.31	30 / 0.26
1-propanol	143	55 / 0.15	6 / 0.41	82 / 0.39	41 / 0.25	47 / 0.25	55 / 0.38	63 / 0.26	50 / 0.33	30 / 0.36
isopropanol	34	18 / 0.11	1 / 0.48	15 / 0.39	10 / 0.19	11 / 0.22	13 / 0.30	12 / 0.19	12 / 0.26	10 / 0.28
1-butanol	211	62 / 0.35	25 / 0.44	124 / 0.49	53 / 0.36	56 / 0.39	102 / 0.51	76 / 0.41	96 / 0.43	39 / 0.52
isobutanol	23	1 / 0.01	17 / 0.32	5 / 0.23	6 / 0.28	8 / 0.22	9 / 0.35	1 / 0.00	18 / 0.36	4 / 0.03
1,3-propanediol	106	40 / 0.26	3 / 0.50	63 / 0.35	4 / 0.14	30 / 0.20	72 / 0.38	31 / 0.30	41 / 0.37	34 / 0.28
2,3-propanediol	0	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00
1,4-butanediol	375	144 / 0.21	88 / 0.35	143 / 0.44	110 / 0.23	129 / 0.31	136 / 0.43	100 / 0.30	145 / 0.34	130 / 0.34
2,3-butanediol	2	0 / 0.00	0 / 0.00	2 / 0.38	1 / 0.37	1 / 0.40	0 / 0.00	0 / 0.00	2 / 0.38	0 / 0.00
3-methyl-1-butanol	0	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00
2-methyl-1-butanol	0	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00
2-keto-isovaleric acid	1	0 / 0.00	1 / 0.54	0 / 0.00	0 / 0.00	0 / 0.00	1 / 0.54	0 / 0.00	0 / 0.00	1 / 0.54
2-keto-valeric acid	1	0 / 0.00	0 / 0.00	1 / 0.50	0 / 0.00	0 / 0.00	1 / 0.50	0 / 0.00	1 / 0.50	0 / 0.00
3-hydroxyvalerate	2	0 / 0.00	2 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	2 / 0.00	1 / 0.00	0 / 0.00	1 / 0.00
2-keto-butanoic acid	17	5 / 0.14	3 / 0.31	9 / 0.50	4 / 0.39	6 / 0.39	7 / 0.31	4 / 0.36	7 / 0.34	6 / 0.38
3-hydroxybutyrate	17	5 / 0.41	8 / 0.53	4 / 0.11	7 / 0.42	4 / 0.36	6 / 0.40	6 / 0.30	6 / 0.43	5 / 0.46
4-hydroxybutyrate	0	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00
2-phenylethanol	0	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00	0 / 0.00

The number of strain designs and the corresponding average yield are separated by "/". Results were tabulated for each yield interval under different oxygenation/substrate/knock-out conditions. Overall values were added at the bottom. The number of strain designs and the corresponding average yield are separated by "/". Results were tabulated for target compounds under different oxygenation/substrate/knock-out conditions. Furthermore, the number of predicted strain designs and the number of growth-coupled pathway for each target compound were tabulated.

For acrylamide, acrylic acid, and 3-hydroxypropanoate, the average yield was higher, only when compared to other targets on xylose. 1-propanol, isopropanol, and 1,3-propanediol yield averages were higher under ECOM/aerobic, and 1,4-butanediol under wild type/anaerobic using glycerol as a substrate. Specifications regarding the number of predicted strain designs and average yield for each pathway are shown in Table A2-4.

Table 2-4: Comparison of GEM-Path growth coupled design to previously identified pathways from literature

Target Compound	aerobic			ECOM			anaerobic		
	glucose	xylose	glycerol	glucose	xylose	glycerol	glucose	xylose	glycerol
1,4-butanediol	0.13*	0.22*	0.36*	0.47*	0.50*	0.60*	0.40*	0.51*	0.60*
1,3-propanediol	0.14*	0.16*	1,15[E]	---	---	1,00[E]	---	1,12[E]	0,86[E]
3-hydroxypropanoate	1,76[C]	2,16[C]	0,51[E]	1,21[C]	2,48[C]	22,5[C]	1,67[C]	1,46[C]	2,29[C]
1-propanol	1,85[E]	0,78[E]	0,79[E]	0.36*	0,65[E]	0.48*	0,82[E]	0,68[E]	0,85[E]
1-butanol	1,34[E]	1,57[E]	2,32[E]	2,03[E/C]	2,31[E/C]	1,06[E]	1,05[E/C]	1,03[E]	1,16[E]
isopropanol	0.18*	0.17*	0.21*	---	---	0.48*	0.38*	0.40*	0.50*
isobutanol	---	---	0.01*	0.35*	0.37*	0.58*	0.54†[E/C]	0,01†[E/C]	---
3-hydroxybutyrate	0.61*	0.35*	0.43*	0.60*	0.52*	0.66*	0.30*	---	0.12*
2-phenylethanol	---	---	---	---	---	---	---	---	---
2,3-propanediol	---	---	---	---	---	---	---	---	---
2,3-butanediol	---	---	---	---	---	---	0.37†[E]	0.40†[E]	---
3-methyl-1-butanol	---	---	---	---	---	---	---	---	---
2-methyl-1-butanol	---	---	---	---	---	---	---	---	---
4-hydroxybutyrate	---	---	---	---	---	---	---	---	---

For each target compound, the growth-coupled ratio between novel pathways generated by GEM-Path and experimentally implemented (E) and/or computationally generated pathways (C) are shown. Empty spaces (---) indicate that no referenced pathways for the corresponding target compound were found.
* no experimentally implemented nor previous computationally predicted pathways were able to growth couple the target compound production. Maximum growth-coupled yield associated with new pathway predicted by GEM-Path is reported.
† only experimentally or previous computationally predicted pathway were able to growth couple the target compound production. Maximum growth-coupled yield is reported.

As expected, depending on the pathway precursor, intermediates, stoichiometry, and cofactors involved, specific combinations for oxygenation/substrate/knock-out lead to different productivities. As shown in Figure 2-10, for each target compound, most of the production potentials were under the maximum theoretical yield average. This behavior is due to the resulting strain designs being predicted as heterofermentative strains and also because some knock-outs significantly constrain the production potential.

In order to compare and determine the growth coupled production potential for the novel pathway calculated using GEM-Path and the already reported pathways (computationally or experimentally), an analysis was performed by calculating the ratio between the highest growth-coupled production for the novel pathways predicted by GEM-Path over the experimentally or computationally reported pathways. Results were displayed under aerobic, ECOM, and anaerobic conditions, by using glucose, xylose, and glycerol as a carbon source. Values over 1 indicated that GEM-Path's novel pathways have higher growth-coupled production potential than already referenced pathways (Table 2-4). Specifically, for 3-hydroxypropanoate, 1-propanol, and 1-butanol,

considerable improvements were found on various substrates and oxygenation conditions. Furthermore, for 1,4-butanediol, isopropanol, isobutanol, and 3-hydroxybutarte, only growth-coupled designs associated to novel pathways from GEM-Path were found.

2.4.7 Analysis of Pathways and Strain Designs with the Highest Yield

The final decision to implement any of the predicted pathways will be subject to the production potential. Promising strain designs with growth-coupled yield between the average theoretical yield plus standard deviation and the highest theoretical yield values were found in silico. The pathways are outlined in Figure 2-11 and strain designs were specified in Table A-2. Specifically, pathways for the production of 1,3-propanediol, 1-butanol, 3-hydroxypropanoate, acrylic acid, and acrylamide were identified as potential candidates for experimental implementation. Of these top performers, 10% of the designs were predicted as being wild type/aerobic, 40% under ECOM/aerobic, and 50% wild type/anaerobic. As for the substrates, 25% of the designs used xylose as a carbon source and 75% used glycerol. No designs from glucose were found among the top producer designs, which indicate how robust *E. coli*'s metabolism is when using glucose as substrate. From all the selected pathways, three similar pathway structure patterns were identified. The first has as its intermediate compound 3-hydroxypropanal (3hppnl) and the precursors glycerol (glyc) and glyceraldehyde (glyald) (see Figure 2-11). Through these pathways, production of 3-hydroxypropanoate and 1,3-propanediol were predicted (deviation values were extracted for analysis. *E. coli* genes for reactions labeled as 1.2, 2.1, and 19.1 have been reported as being able to catalyze such enzymatic transformations (Jo et al. 2008; Zhuge et al. 2010). Pathways for designs 1, 2, 16, and 17 (see deviation values were extracted for analysis. The pathways showed that all reactions are "known" (no promiscuous inference). Furthermore, designs 1 and 2 showed an interesting phenotype, diverting most of the glycerol uptake flux through the production of 3-hydroxypropanoate, while fixing carbon dioxide for growth. In this design, approximately 5% of the carbon intake must come from carbon dioxide. This assumption could be valid as it has been reported in *E. coli* that carbon dioxide uptake representing around 10% of the entire carbon uptake has been observed (Lu et al. 2009). The second common pathway structure, starting from design identifier 3 to 10

(deviation values were extracted for analysis. The pathways combines predictions that are able to produce acrylic acid, acrylamide, and 3hydroxypropanoate. These pathways have propenoyl-CoA (prpncoa) as a common intermediate and propanoyl-CoA (ppcoa) as a precursor. None of the pathways related to these designs were completely known, which means that at least one reaction in the pathway was predicted as promiscuous when comparing against BRENDA. All of the designs in this set were predicted under ECOM/aerobic conditions and the pyruvate kinase knock-out was involved in all of them. Finally, the third common structured pathway pattern, starting from design identifier 11 to 15. For the linear pathway regarding the production of 1-butanol from butanoyl-CoA, all of the intermediates were native *E. coli* metabolites. All of these designs were predicted under wild type/anaerobic conditions using only glycerol as a substrate. Reactions 12.1 and 13.1 were already in the E. coli model (ALDD4 and BUTCT respectively) (Orth, Conrad, Na, Lerman, Nam, Feist and Palsson 2011), but defined as irreversible in the opposite direction. These designs represent a subset of a total of 1271 designs identified in this work which are provided in detail in the online paper resources.

2.4.8 GEM-Path output example

2.4.8.1 Case I: Production of 1,3-propanediol

A output example for the production of 1,3-propanediol using GEM-Path is given in Figure 2-12. Two different GEM-Path calculated pathways were outlined: Pathway #7 (reactions 6 and 3) that has already been experimentally implemented, and pathway #16 (reactions 17, 12, 16, and 3). For pathway #16, specific output relating to the existence of catalyzing reactions from BRENDA and the promiscuity analysis are shown. For reaction 3, 6, and 17, exact matches in the BRENDA database were found, sharing identical cofactors, substrates, and products. This is represented by a substrate TC equal to 0 during the search. Furthermore, the species and EC number were reported (Ishikura et al. 2005; Kajiura et al. 2007; Wang, Liu, Xian, Zhang and Zhao 2012). It is worthwhile to note that for homologous enzymes, there was no ranking in terms of species shown to carry out a given reaction. The algorithm reports only the first hit, associated with the corresponding species and the predicted reaction. For experimental purposes, it may be necessary to use the predicted EC Number and a database such as

BRENDA to find multiple possible enzyme options that can be evaluated. The concept of sorting in terms of species distances between the host and the species associated with the predicted reaction was not considered. Attempts have been made to mine content for homologous expression in *E. coli* (Bayer et al. 2009), but this is an active field of research. According to the iterative algorithm shown in Figure 2-5, no exact substrate structural matches were found for reactions 12 and 16. Instead, a promiscuity analysis was performed, obtaining the corresponding reactions from BRENDA. The promiscuity space is represented by a multi-dimensional space, obtained from the multi-linear regression analysis. For simplicity, and in order to describe the promiscuity analysis output,, a two dimensional space was outlined in Fig 6 and the native BRENDA reaction substrate's (blue circles) distances from the centroid (red circle) were normalized to 1, and the tested substrate (green circle) was outlined at a distance equal to the promiscuity score. For reaction 12 and 16, a PS equal to 1.1 and 0.69 were calculated, respectively. The BRENDA predicted reaction was reported showing the substrate difference in terms of the TC, the corresponding species, and EC number(Furuyoshi et al. 1991).

In order to avoid extensive computation, the algorithm chooses and saves the first possible solution for the particular species and related promiscuity score, leaving behind a number of additional feasible solutions able to fulfill the PS threshold. Reaction 17 represents a reaction gap filled by the algorithm. As shown in Figure 2-12, the only reaction that connects the D-malate (mal-D) metabolite to *E. coli* metabolism are transport reactions through the periplasmic and external membrane (iMALDt2_2pp and MALDtex, respectively). Since mal-D was not set as a media constituent, there was no option for it to be generated by the network. By calculating reaction 17, it was feasible to connect a heterologous pathway to central carbon metabolism, specifically to oxaloacetate (oaa). Note that *E. coli* does contain malate dehydrogenase (MDH) which reversibly converts L-malate to oaa, but it is not implied that it can convert it to D-malate(Sutherland and McAlister-Henn 1985).

Following synthetic calculation of heterologous pathways for each target compound, strain design computations were performed to engineer host cell metabolism. Continuing with the current example for 1,3-propandiol production,

production envelopes of growth-coupled designs for pathways #7 and #16 were outlined in Figure 2-16a. Furthermore, a productivity analysis under different conditions was performed (Figure 2-16b), where shaded areas represent the maximum theoretical production rate by setting the computational minimal growth rate to 0.1 h-1, and solid areas represent the maximum growth-coupled production rate. As mentioned before, the overall trend shows that under aerobic conditions, maximum theoretical production is higher flux compared to anaerobic. Moreover, by using glycerol as a substrate instead of glucose, higher productivities were calculated for both aerobic and anaerobic conditions. Specifically, when comparing the maximum theoretical production for pathways #16 and #7 under aerobic conditions, an increase of 17% and 25% was observed, and under anaerobic conditions a 6% and 67% increase was observed over glucose, respectively.
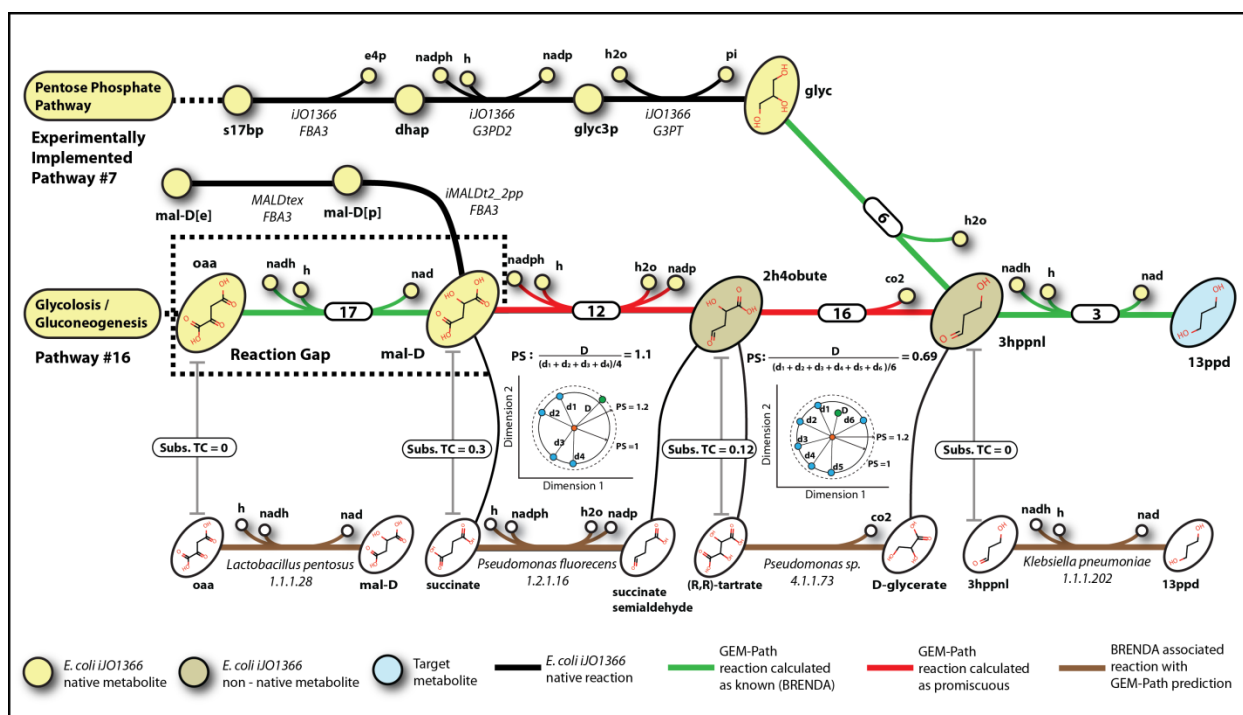
Figure 2-12: GEM integrated synthetic pathway calculation (GEM-Path) output for 1,3-propanediol. Two different GEM-Path calculated pathways are shown: Pathway #7 (top, reactions 6 and 3) which has been experimentally implemented, and pathway #16 (bottom, reactions 17, 12, 16, and 3). For each pathway, reactions leading from the host metabolome are shown in black. Native and non-native *E. coli* metabolites are represented in yellow and brown, respectively. The corresponding target compound is shown in light blue. Reactions calculated as known in BRENDA and reaction calculated as promiscuous are shown in green and red, respectively. For each predicted reaction in pathway #16, specific values of the tanimoto coefficient (TC), promiscuity score (PS) and the corresponding BRENDA reaction (brown lines) are shown. For reactions predicted as promiscuous, the corresponding promiscuity space was outlined with the number of metabolites associated with the specific reaction found in BRENDA. For simplicity, a two dimensional space was plotted, where each of the native BRENDA metabolites (green) are separated from the centroid (red circle) in 1 dimensionless unit and the predicted metabolite is shown in green, with the corresponding distance equal to the PS. A promiscuity score threshold was plotted at a distance equal to 2. Each BRENDA reaction shows the corresponding associated EC number and a species known to catalyze the specific reaction.

Depending on the inserted heterologous pathway, different flux distributions were calculated. For pathways #16 and #7, the flux solution ratio for maximum theoretical

production when using glycerol over glucose as substrates was calculated and qualitatively outlined (see

Figure 2-13). For both pathways under aerobic conditions, there was a decrease in the carbon dioxide evolution when using glycerol as a substrate (approximately 50% less carbon dioxide was produced). Looking at the flux distributions, for pathway #16, no activity was predicted for the pentose phosphate pathway (PPP) when using glycerol as a substrate, and a higher target production rate when using glycerol was observed. This was due to the glycerol uptake metabolism, which is able to produce nadh and nadph similar to the PPP, but without generating carbon dioxide in the process, leading to more efficient carbon metabolism. For anaerobic conditions, a similar trend and a mixed acid fermentation behavior was observed. By comparing the maximum theoretical production for both pathways under the same conditions (same substrate and oxygenation), pathway #16 is able to achieve higher productivity by using glucose as a substrate, approximately 4% and 22% more under aerobic and anaerobic conditions, respectively. Still, by using glycerol as a carbon source, the productivity decreases approximately, 3% and 22% less under aerobic and anaerobic conditions, respectively. This result demonstrated that the novel GEM-Path predicted pathways #16 is more suited to implement linked to a glucose based fermentation process. According to Figure 2-16a and Figure 2-16b, growth-coupled designs were only found for glycerol under both anaerobic and aerobic conditions. No growth-coupled designs associated with glucose under anaerobic conditions were found, and under aerobic conditions, only a low productivity growth coupled design for pathway #16 was found.
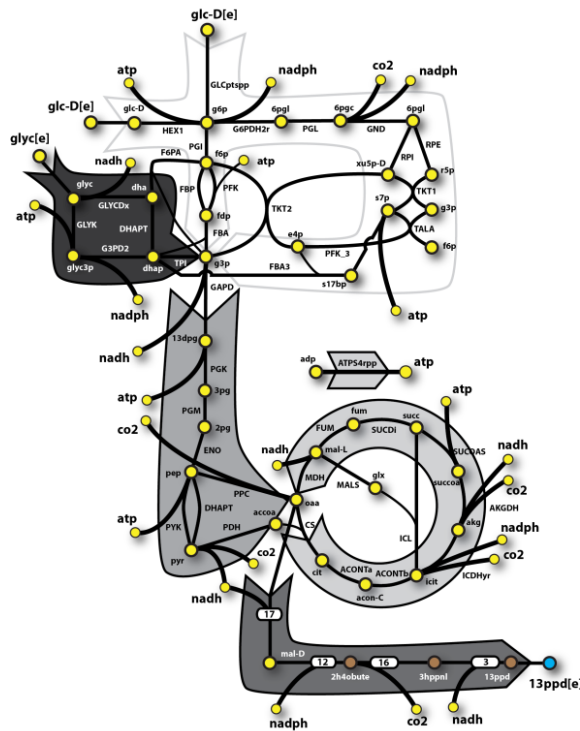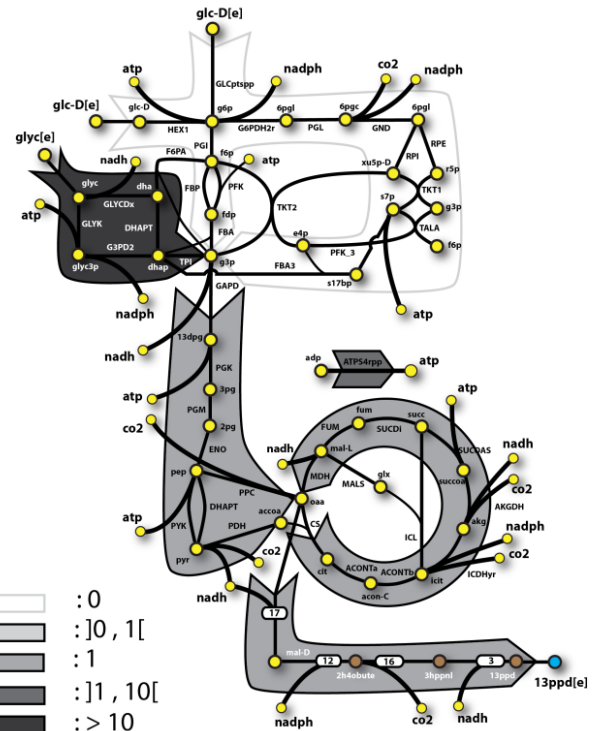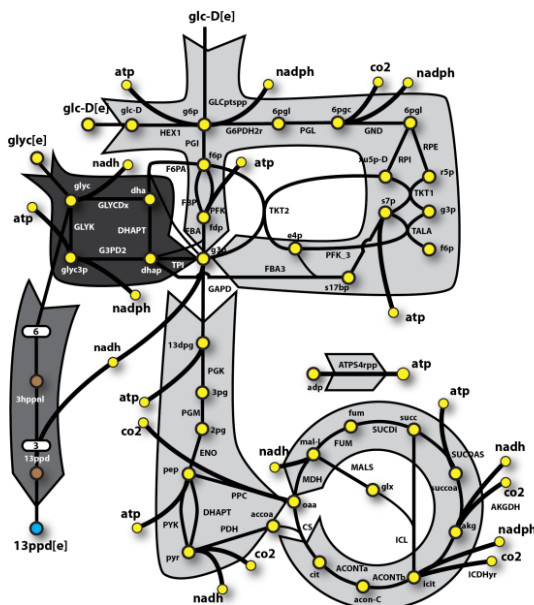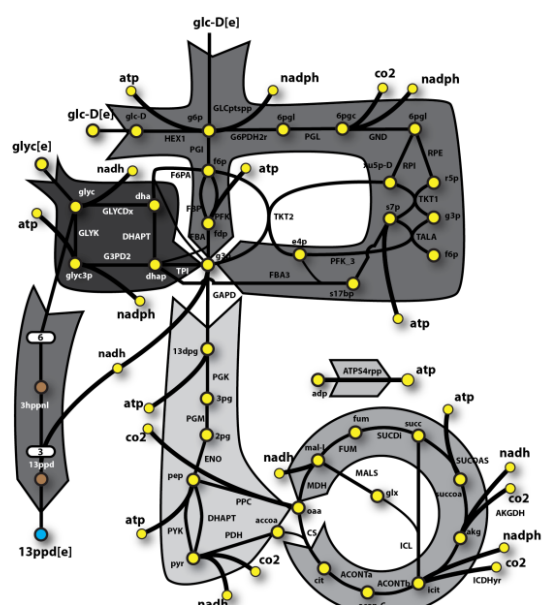
Figure 2-13: Mayor pathways FBA solution comparison for 1,3-propanediol production. Two different pathways (Pathways #16 and #7) where used to compared FBA solutions under aerobic and anaerobic conditions. The substrates used for the analysis were glycerol and glucose. FBA was executed by : i) incorporating the heterologous pathways to the model, ii) setting an uptake rate to 120 C-mmol gDW-1h-1 for each carbon source, 20 mmol gDW-1h-1 O2 when specified, iii) setting a minimal growth rate to sustain growth as 0.1 h-1 (as set by the amount of flux necessary through the BOFcore), and iv) using maximizing the flux through each of the exchange reactions in the model for the target compound. Afterwards, the ration between the glycerol over the glucose FBA solution was calculated. Reactions associated for each pathway were qualitatively analyzed, and pathways were highlighted in terms of the FBA solution ration. The intensity of the highlighted pathways increase with the value of the ration. Five different ratios intervals were chosen to analyze the difference between the conditions: approximately 0, between 0 and 1, approximately 1, between 1 and 10 and higher than 10. These conditions represents: no flux through the pathway by using glycerol as a substrate, higher flux through the pathway by using glucose as a substrate, no difference on the flux distribution, higher flux through the pathway by using glycerol as a substrate, and extremely higher flux through the pathway by using glycerol as a substrate, respectively.

89

2.4.8.2 Case-study II: Production of isopropanol

GEM-Path predicted pathways for the production of isopropanol are given in Figure 2-14. For the production of isopropanol, two different GEM-Path calculated pathways were outlined: Pathway #13 (reactions 7, 8, and 1) which has been experimentally implemented, and Pathway #1 (reactions 1 and 3). For Pathway #1, specific output relating to a promiscuity analysis is shown. For reaction 1, an exact match in the BRENDA database was found, sharing the same cofactors, substrates, and products. This is represented by a substrate TC equal to 0 during the search. Furthermore, a species known to carry out this reaction and EC number were reported (Drewke and Ciriacy 1988). Although the reaction was reported to proceed in the reverse direction, no evidence showing reaction irreversibility was found. Conversely, no exact substrate structural match was found for reaction 3. Thus, a promiscuity analysis was performed, obtaining the corresponding reaction from BRENDA. The promiscuity space is represented by 11 different native substrates (n =11). In order to describe the promiscuity analysis output and for simplicity, a two dimensional space was outlined (as described above). According to the results, a PS equal to 0.88 was calculated. The BRENDA predicted reaction was reported showing the substrate difference in terms of the TC along with a corresponding species and EC number. For the other pathway (i.e. #13), reaction 7 represents a reaction gap filled by the algorithm. As shown in Figure 2-14, two different reactions connect acetoacetate (acac) to other metabolites in the network. The first reaction is ACACCT, which is an irreversible reaction on pathway #13 opposite the isopropanol production direction, and the second reaction is a transport reaction. Since acac was not set as a media constituent, there was no option for it to be generated by the network. By calculating Reaction 7, it was feasible to connect the heterologous pathway to central carbon metabolism, specifically to acetyl-coa (accoa). It should be noted that there is experimental evidence for the existence of a reversible ACACT1r reaction (Fujii et al. 2010; Gulevich et al. 2012), but there is also contradictory evidence indicating that operation in this direction could be highly unfavorable (Lan and Liao 2012; McCloskey et al. 2014). Nonetheless, the GEM-Path algorithm uses the content as defined in the model (Orth, Conrad, Na, Lerman, Nam, Feist and Palsson 2011) and curation is a helpful step after promising production pathways are identified.

Beyond identifying pathways, growth coupled-designs utilizing pathways #1 and #13 were outlined in Figure 2-16b. Furthermore, a max theoretical production analysis under different conditions was performed (Figure 2-16d), where shaded areas represent the maximum theoretical production by setting the computational growth rate to 0.1 h-1, and solid areas show the maximum growth-coupled productivity. As mention before, the overall trend shows that under aerobic conditions, pathways are capable of carrying higher theoretical flux as compared to anaerobic. Moreover, by using glycerol as a substrate instead of glucose, higher productivities were calculated for aerobic conditions. For anaerobic conditions and using glycerol as a substrate, only pathway #1 was able to achieve higher flux compared to glucose.
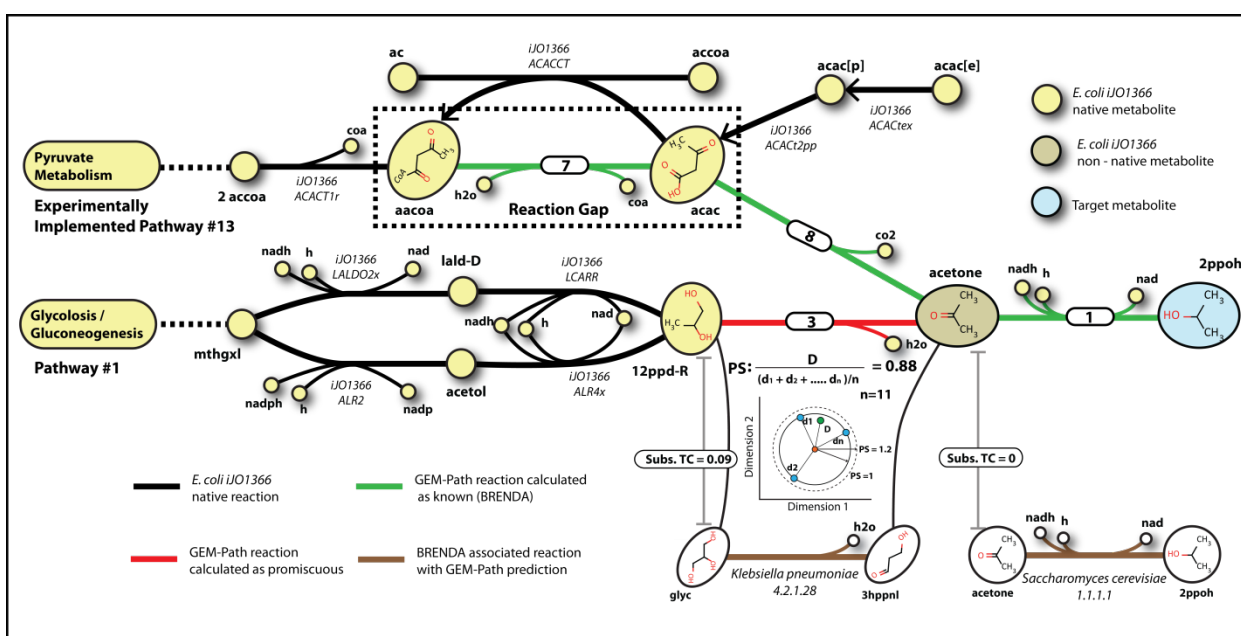
Figure 2-14: GEM integrated synthetic pathway calculation (GEM-Path) output for GEM integrated synthetic pathway calculation (GEM-Path) output for isopropanol. Two different GEM-Path calculated pathways are shown. Pathway #13 (reactions 7, 8, and 1), that has been experimentally implemented, and Pathway #1 (reactions 1 and 3). For each pathway, reactions leading to the host cell metabolome are shown in black. Native and non-native *E. coli* metabolites are represented in yellow and brown, respectively. The corresponding target compound is shown in light blue. Reactions calculated as known in the BRENDA database and reactions calculated as promiscuous are shown in green and red, respectively. For each predicted reaction in pathway #1, specific values of the tanimoto coefficient (TC), promiscuity score (PS), and the corresponding BRENDA reaction (brown) are specified. For reactions predicted as promiscuous, the corresponding promiscuity space was outline with the number of metabolites associated with the specific reaction found in BRENDA. For simplicity, a two dimensional space was plot, where each of the native BRENDA metabolites (green) are separated from the centroid (red circle) in 1 dimensionless units and the predicted metabolite is shown in green, with the corresponding distance equal to the PS. A promiscuity scores threshold was plotted at a distance equal to 2. Each BRENDA reaction shows the corresponding associated EC number and species known to carry out the reaction.

A maximum theoretical production analysis for isopropanol revealed differences in production potential when using glycerol or glucose as a substrate. For both pathways

under aerobic conditions, a decrease in carbon dioxide evolution was observed when using glycerol as a substrate. For both pathways, approximately 25% less carbon dioxide was produced. Looking at the flux distribution (see Figure 2-15), for pathway #1 and #13, no activity in the PPP during glycerol consumption was observed, due to the same reason described in the first 1,3-propanediol case study. For anaerobic conditions, specifically for pathway #1, a mix acid fermentation behavior was observed. Higher productivity was observed compared to glycerol for pathway #13 under anaerobic conditions and using glucose as a substrate. Byproduct formation during glycerol growth was critical in diminishing the productivity.

By comparing the pathway's maximum theoretical production under the same conditions (same substrate and oxygenation), pathway #1 is able to achieve higher productivity by using glucose as a substrate, approximately 1% and 7% more under aerobic and anaerobic conditions, respectively. When utilizing glycerol as a carbon source, the productivity increases approximately 6% and 94% under anaerobic conditions, respectively. This result demonstrated that the novel GEM-Path predicted pathway #1 shows a higher theoretical potential when using a glycerol based fermentation process. As indicated in Figure 2-16c and Figure 2-16d, growth-coupled designs were only found for pathway #1. Growth-coupled productivities similar to the maximum theoretical achievable productivity were found using both glycerol and glucose as a substrate and under anaerobic conditions.
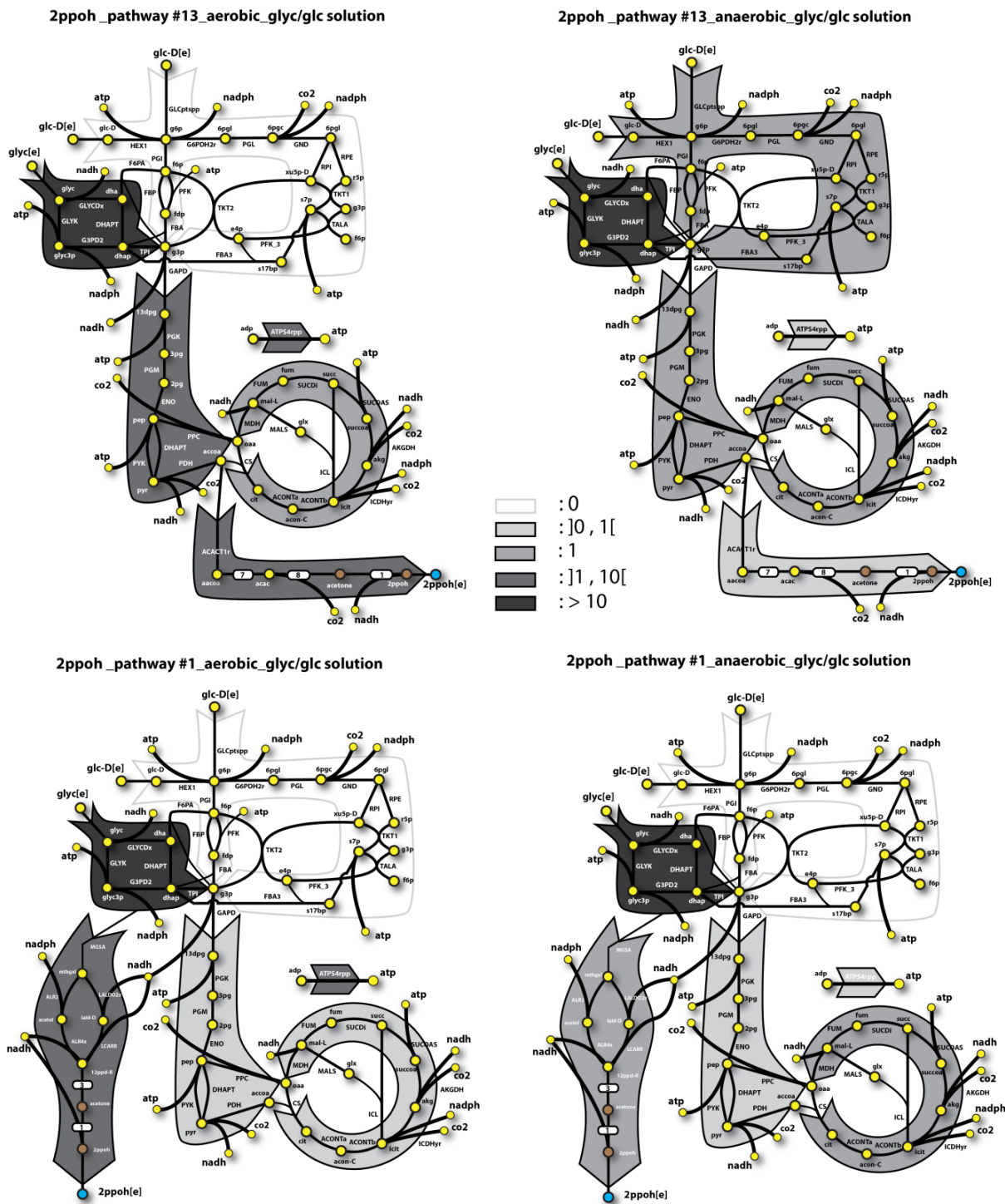
Figure 2-15: Mayor pathways FBA solution comparison for isopropanol production. Two different pathways (Pathways #13 and #1) where used to compared FBA solutions under aerobic and anaerobic conditions. The substrates used for the analysis were glycerol and glucose. FBA was executed by : i) incorporating the heterologous pathways to the model, ii) setting an uptake rate to 120 C-mmol gDW-1h-1 for each carbon source, 20 mmol gDW-1h-1 O2 when specified, iii) setting a minimal growth rate to sustain growth as 0.1 h-1 (as set by the amount of flux necessary through the BOFcore), and iv) using maximizing the flux through each of the exchange reactions in the model for the target compound. Afterwards, the ration between the glycerol over the glucose FBA solution was calculated. Reactions associated for each pathway were qualitatively analyzed, and pathways were highlighted in terms of the FBA solution ration. The intensity of the highlighted pathways increase with the value of the ration. Five different ratios intervals were chosen to analyze the difference between the conditions: approximately 0, between 0 and 1, approximately 1, between 1 and 10 and higher than 10. These conditions represents: no flux through the pathway by using glycerol as a substrate, higher flux through the pathway by using glucose as a substrate, no difference on the flux distribution, higher flux through the pathway by using glycerol as a substrate, and extremely higher flux through the pathway by using glycerol as a substrate, respectively.

93

Figure 2-16: Strain design productivity analysis for GEM-Path case studies. The production envelopes for strain designs of 2-4 gene knockouts were plot (a and c). Glucose and glycerol were used as substrate examples under aerobic and anaerobic conditions. Production envelopes for 1,3-propanediol (pathways #16 and #7) and isopropanol (pathways #13 and #1) are shown in (a) and (c), respectively. Solid blue and red lines represent experimentally implemented pathways and novel GEM-Path calculated pathways, respectively. Production envelopes for growth-coupled designs are shown in dotted lines. Productivity analysis for the production of 1,3-propanediol (b) and isopropanol (d) were outlined. Results were grouped for aerobic and anaerobic conditions, associated with the corresponding pathway number. By using glucose (blue) and glycerol (green) as substrates, maximum theoretical production rate (shaded bars) and growth-coupled production rate (filled bars) were plotted. FBA was used to determine the maximum theoretical productivity by setting the growth rate to 0.1 h-1 and optimizing for the target compound production. Growth-coupled productivity was calculated by knocking out computationally identified reactions and optimizing for growth rate. The maximum value for each condition was reported.

2.5 Conclusions

The aim of this work was to outline the production potential for 20 industrially-relevant chemicals in *E. coli* and generate feasible designs for production strains. The enabling technology generated for the project was a computational pipeline including chemoinformatics, bioinformatics, constraint-based modeling, and GEMs to aid in the process of metabolic engineering of microbes for industrial bioprocessing purposes. The main results from this study are, i) a comprehensive mapping from *E. coli's* native metabolome to commodity chemicals that are 4 reactions or less away from a natural metabolite, ii) sets of metabolic interventions, specifically knock-outs and knock-ins, that coupled the target chemical production to growth rate, iii) the development of a retrosynthetic based pathway predictor algorithm containing a novel integration with GEMs and reaction promiscuity analysis, and iv) a complete strain design workflow integrating synthetic pathway prediction with growth-coupled designs for the production of non-native compounds in a target organism of interest.

For synthetic pathway predictions, much attention has been focused on retrosynthetic algorithms, where a backward search for synthetic pathways is performed by an iterative application of biochemical reaction operators (BROs) from a target compound to a predefined source of metabolites (Medema, van Raaphorst, Takano and Breitling 2012). Based on 443 BROs included in this work, a retrosynthetic pathway predictor algorithm was developed which incorporates GEMs into the procedure. The GEM-Path algorithm is also coupled together with database analysis for reaction existence and reaction promiscuity inference. Predictions were compared to literature, and showed a good agreement with previously reported algorithms. Due to the filtering procedure at each iteration step, specifically the promiscuity analysis, the number of generated pathways was considerably lower as compared with previous algorithms, diminishing the candidates required for further experimental implementation. In total, GEM-Path generated 245 synthetic pathways for the production of 20 different compounds in *E. coli.* The majority of the predicted pathways involved at least one promiscuous reaction. Since the promiscuity analysis is based on E.C. reaction numbers instead of genes, an enzymatic validation step may be necessary to confirm the predicted functionality before introduction into a production host.

Theoretically, all synthetic pathways identified in this work are able to produce the target compound under a given substrate/oxygenation/strain conditions and in total, they characterize the production space. Novel pathways able to achieve high yield were found for a range of commodity chemicals. According to the theoretical maximum yield analysis, pathways implemented under wild type aerobic conditions tend to have a greater production potential compared to the other strain/oxygenation conditions. Furthermore, when changing the anaerobiosis threshold, the more anaerobic the condition of a strain, the less overall production could be achieved. Lower maximum theoretical yields observed in anaerobic conditions (vs. aerobic) are expected, as no oxygen is essentially an additional constraint, limiting the capability of the network (just like the removal of a key reaction in the network). Based on a C-mol Yield basis, under wild-type/aerobic conditions, glycerol is found to be the most efficient substrate for heterologous target compound production. However, for wild-type/anaerobic, xylose and glucose are the most efficient substrates. Additionally, precursor yield analysis reveals that pathways having precursors closest to the central metabolism are able to achieve higher yields which agrees with logic as central metabolic reactions carry the most flux in the network (Almaas et al. 2004).

Growth-coupled production of a specific metabolite depends on the energy benefit that the cell can obtain through the pathway activation related to the growth-coupled metabolite. Growth-coupled design algorithms operate by knocking out reactions, thus generating an energy imbalance that is recovered by then coupling different pathways to growth. The final metabolite involved in these pathways works as a final electron acceptor, thus, under anaerobic conditions, pathways are more susceptible to coupling to growth. The ability to find growth-coupled designs preferentially under anaerobic conditions can be seen by analyzing the overall results, where growth-coupled designs under wild-type anaerobic conditions were found to be present more frequently and were able to achieve higher yields. Further, designs with glycerol as a substrate had the highest yields anaerobically. Thus, under anaerobic conditions, growth-coupled designs are easier to obtain compared to aerobic conditions. Furthermore, for most of the predicted reactions contributing to a growth-coupled design, approximately 40%, were oxidoredutases with NAD or NADP acceptors.

Removal of these reactions facilitates growth-coupling as they shift the flow of electrons in metabolism (King and Feist 2013).

Designs highlighted in this work were selected according to their production potential (i.e., yield, $Y_{p/s}$). Nevertheless, further improvements are needed for the design and production workflow to promote success in experimental implementation. For instance, toxicity due to product or co-product formation was not evaluated during the design pipeline; this might lead to the production of toxic compounds together with cell death. Due to the scope of the GEMs used in this work (i.e., metabolic GEMs), key regulatory steps were not taken into account. Furthermore, the impact of low substrate affinity of predicted promiscuous enzymes might lead to false positive results, decreasing the *in vivo* maximum achievable yield. However, to generate non-native products, it is obvious that new production pathways are necessary and thus that was the focus of this work. Furthermore, growth coupled designs, such as those produced here, provide an extra tool for metabolic engineers by allowing for the use of selection pressure to achieve a desired production state. For reactions predicted as promiscuous, *in vitro* enzyme analysis might be necessary to identify and characterize the potential promiscuous activity. Moreover, in order to avoid undesirable metabolite sinks, a promiscuity analysis regarding native metabolites must be taken into account when reactions are incorporated into metabolism. Lastly, as in any production strain project, enzyme efficiency issues and heterologous codon optimization (Medema, van Raaphorst, Takano and Breitling 2012) must also be considered for product formation.

Taken together, the workflow presented here finds that the 20 major commodity chemicals are within 4 reactions from the metabolic network of *E. coli*. Further, it maps out all the feasible pathways linking the chemical structures of these commodity chemicals to the metabolic network of *E. coli* and their theoretical yields. It also maps out the chemical reactions and enzymatic requirements for building these pathways. Thus, in a way, we have generated a pathway atlas that can guide the global metabolic engineering and strain design efforts needed to convert the petroleum-based industry to a biomass-based industry, and thus forms the basis for a grand challenge undertaking by the community.

Chapter 3

A chemoinformatic approach to predict Gain-of – Function (GoF) oncometabolites.

3.1 Abstract

By using the mutated isocitrate dehydrogenase (IDH) example to produced oncometabolites due to Gain-of-Function (GoF) mutations, that could be used for biomarkers identification, a chemoinformatic based workflow was generated to identify potential chemical oncometaboliotes substructures. Based on 6 mutated enzymatic genes and the corresponding metabolic reactions associations and assuming an IDH type of GoF kind of behavior for each mutated enzyme, synthetic reaction were simulated and further filtered to identify potential promiscuous activities. By using a maximum common substructures (MCS) search on all predicted promiscuous metabolites for each mutated enzyme reaction, 24 chemical substructures of oncometabolites resulting from the GoF analysis were predicted.

3.2 Introduction

The metabolic state coined as "aerobic glycolysis" (Warburg, Posener and Negelein 1924). which has been described as a passive response causing a malignant transformation (Hanahan and Weinberg 2011), is being challenged by recent discoveries. Studies has been proven that altered metabolism by itself can be a driver for oncogenesis (Letouze, Martinelli, Loriot, Burnichon, Abermil, Ottolenghi, Janin, Menara, Nguyen, Benit, Buffet, Marcaillou, Bertherat, Amar, Rustin, De Reynies, Gimenez-Roqueplo and Favier 2013; Lu, Ward, Kapoor, Rohle, Turcan, Abdel-Wahab, Edwards, Khanin, Figueroa, Melnick, Wellen, O'Rourke, Berger, Chan, Levine, Mellinghoff and Thompson 2012a; Xiao, Yang, Xu, Ma, Lin, Zhu, Liu, Liu, Yang, Xu,

Zhao, Ye, Xiong and Guan 2012; Xu, Yang, Liu, Yang, Wang, Kim, Ito, Yang, Xiao, Liu, Jiang, Liu, Zhang, Wang, Frye, Zhang, Xu, Lei, Guan, Zhao and Xiong 2011; Yang, Soga, Pollard and Adam 2012b). Specifically, characterized isocitrate dehydrogenase (IDH1, IDH2) mutations have established a new paradigm in cancer developments in that the heterozygous point mutations confer a new metabolic enzymatic activity that produce an oncometabolite (e.g. 2-hydroxyglutarate (2-HG), from α-ketoglutarate(α-KG)) (see Figure 3-1, GoF square). In fact, 2-HG shows a 100-fold increased concentration in glioma and acute myeloid leukemia's (AML) patients with IDH1 or IDH2 missense mutations. This increased concentration of 2-HG competitively inhibits α-ketoglutarate binding to histone demethylases, thus blocking differentiation of cells (Lu, Ward, Kapoor, Rohle, Turcan, Abdel-Wahab, Edwards, Khanin, Figueroa, Melnick, Wellen, O'Rourke, Berger, Chan, Levine, Mellinghoff and Thompson 2012a; Xu, Yang, Liu, Yang, Wang, Kim, Ito, Yang, Xiao, Liu, Jiang, Liu, Zhang, Wang, Frye, Zhang, Xu, Lei, Guan, Zhao and Xiong 2011).

Recently the Constraint-Based Reconstruction and Analysis (COBRA) approach, an emerging field in systems biology approach, has been proven successful in addressing heterogeneity in cancer by integrating experimental data with the genome scale metabolic reconstructions (GEMs) to tailor the models to the unique gene expression profiles of general cancer tissue, and even individual cell lines and tumors. This enables to study the cancer metabolism and simulate tumor phenotypes from a genome wide perspective(Lewis and Abdel-Haleem 2013). Based on genetic mutation information on a massive scale collected from more than 1,700 cancer genomes into context-specific GEMs of metabolism for nine cancer types, Nam et al., 2014 determined 6 different mutated metabolic enzymatic genes for GoF analysis, which presented missense mutations similar to those presented in the IDH case. According to the latter, in this work, a chemoinformatic based workflow was generated for predicting potential oncometabolites chemical structures due to GoF mutation.

3.3 Materials and Methods

3.3.1 Chemoinformatics resources for gain-of-function (GoF) oncometabolite predictions

Throughout the workflow for GoF oncometabolite predictions, chemoinformatic tools were essential for integrating computational chemical analysis into genome-scale model theory by simulating synthetic reactions, analyzing metabolites databases and results. A range of chemoinformatic in-house methods and functions, which are described below, were incorporated into the MATLAB® environment by linking it with ChemAxon's software package libraries, specifically Marvin, JChem Base, Standardizer and Reactor [(ChemAxon, Budapest,Hungary, www.chemaxon.com)].

3.3.1.1 Defining similarity (Measuring similarities of substrates and reactions)

Based on gene mutations (i.e. GoF), enzymes are capable to carry out new catalytic functions. Two different kinds of changes on the catalytic functions can be distinguished. First, the basic reaction mechanism is changed (e.g. changes up to three digits in the EC number) which directly affects the nature of the substrate to be catalyzed by the mutated enzyme. And on the other hand, the basic reaction mechanism is preserved, but substrate specificity is changed (e.g. changes in the fourth digit in the EC number). This can directly affect the kcat for the native substrate and also affects different molecules which are structurally similar to the native substrate. These promiscuous enzyme descriptions are presented in this work as GoF. Changes between non-mutated and mutated enzymes (GoF) could be measured and compared using a chemoinformatic approach. A number of studies have previously shown the use of chemoinformatic tools and approaches for enzyme promiscuity predictions(de Groot et al. 2009; Hu et al. 2011), reaction Enzyme Commission (EC) number assignments(Hu, Zhu, Li, Zhang, Deng and Yang 2012; Latino and Aires-de-Sousa 2009), heterologous metabolic pathways for small chemicals predictions(Campodonico, Andrews, Asenjo, Palsson and Feist 2014), and reaction database analysis(Holliday et al. 2012). In this study mutated enzyme's native and non-native metabolites and reactions were represented by fingerprints, and the similarity of reactions and metabolites were determined by calculating the corresponding Tanimoto coefficient between fingerprints.

3.3.1.2 Insilico chemical representation

For compound and reaction representation MDL Molfiles (Dalby, Nourse, Hounshell, Gushurst, Grier, Leland and Laufer 1992) were used. A Molfile contains information about the atoms, bonds, connectivity, and coordinates of a molecule. The Molfile consists of some header information, the connection table containing atom information, then bond connections and types, followed by sections for more complex information.

3.3.1.3 Molecular and reactions fingerprints

The chemical fingerprint (CFP) of a molecule is defined as CFP = (Fi), in which Fi refers to a molecular fragment with real occurrences of a molecule. Fi is obtained by molecular fragmentation method. Each Fi in the fingerprint is represented in bit string where each position of the sequence is represented by '1' or '0' digits, depending on the presence or absence of the structural pattern predefined by Fi. Previous studies have shown good results by using linear fragments from 5 up to 6 bonds (Hu, Zhu, Li, Zhang, Deng and Yang 2012; Latino and Aires-de-Sousa 2009). In order to choose optimal parameters for the compounds present in the HMDB, the ChemAxon parameter optimization criteria was used. From this analysis, the fingerprint length, the maximum fragmentation pattern length and number of bit in the string were 1024, 6, and 2 respectively.

In the case of reaction fingerprints, each compound present in the reaction is represented as a CFP. This fingerprint is a segmented fingerprint that is constituted from 8 chemical fingerprint sections. This reaction fingerprint representation allows to compare reactions from its structural and transformational features. Parameters were defines as previously stated for molecules.

3.3.1.4 Tanimoto coefficient

The premise of similarity searching is that similar structures have similar fingerprints. Here, we used the TC dissimilarity (TC$_{diss}$) metric to determine how similar two fingerprints were. Values of this metric are non-negative numbers. A zero dissimilarity value indicates that the two fingerprints are identical, and the larger the value of the

dissimilarity coefficient the higher the difference between the two structures. In its original form, the Tanimoto metric is a similarity metric (TC_{sim}):

$$T_{sim} = \frac{B(a\&b)}{B(a) + B(b) - B(a\&b)}$$

Where *a* and *b* are two binary fingerprints, & denotes binary bit-wise and-operator, | denotes bit-wise or-operator, and *B(x)* is the number of 1 bits in any binary fingerprint x:

$$B(x) = |\{x_i = 1\}|x_i \in \{0,1\}; i = 1, \ldots \ldots, n\}| = \sum_{i=1}^{n} x_i$$

From that it is straightforward to obtain a dissimilarity measure:

$$T_{diss} = 1 - T_{sim}$$

It is worth noting that if the TC_{diss} between two fingerprints is 0, it means that both molecules share the exact same fingerprint. While this doesn't mean that both molecules are the same, it does mean that both molecules share the same bonds according to the fragmentation process, since the molecular fingerprint only represents the presence or absence of a given particular bond pattern.

3.3.2 Synthetic Reaction Construction: SMIRKS & SMARTS

For biochemical reaction operators (BROs) representation, SMIRKS (James, Weininger and Delany 2004) was used as a language for describing generic reactions by using a SMARTS (James, Weininger and Delany 2004) representation of the reaction's substructures. A SMARTS pattern may include not only a specification of reaction center but also a specification of a local structure that must occur or is necessarily absent based on our best understanding of the relevant biochemistry (Silverman 2002). BROs were constructed based on the smallest substructure related to the structural change of the main substrates and products in the reaction. Based on previous studies (Henry, Broadbelt and Hatzimanikatis 2010; Mu, Unkefer, Unkefer and Hlavacek 2011; Yim, Haselbeck, Niu, Pujol-Baxley, Burgard, Boldt, Khandurina, Trawick, Osterhout, Stephen, Estadilla, Teisan, Schreyer, Andrae, Yang, Lee, Burk and

Van Dien 2011), a set of 374 irreversible BROs were defined to generate novel biochemical reactions and pathways. Approximately 76% of the reactions in KEGG (Kanehisa, Goto, Hattori, Aoki-Kinoshita, Itoh, Kawashima, Katayama, Araki and Hirakawa 2006) and 72% of the reactions in BRENDA (Curran and Alper 2012) involved a transformation captured in this defined BRO set. Each BRO is related to specific cofactors and a third-level EC number for further reaction reconstruction and identification. After BRO construction, we generate all possible reactions that may occur and every compound that may be produced given the previous selected list of human metabolites. Then, specific cofactors were assigned, and for filtering purposes mass balance was performed.

3.3.3 Compound Clustering Analysis

Maximum common substructures (MCS), refers to the largest substructure shared by two compounds and provide one of the most accurate similarity measures(Cao et al. 2008). The MCS problem has been well studied as a general graph matching problem, and has found applications in many areas(Bunke 2000). In this study, in order to identify the most common chemical pattern from all predicted products and substrates related to each mutated enzyme, a MCS analysis was used. Basically the algorithm is able to identify from a range of molecular structures, the largest substructure common in all of them.

3.4 Results and Discussions

Results and discussions presented here are in the context of Nam et al., 2014 work in PLoS Computational Biology. At the moment of this thesis presentation, the manuscript was accepted for publication. The general workflow used for predicting oncometabolites via context-specific genome-scale metabolic network is described in Figure 3-1. As shown in Figure 3-1, starting from nine different cancer cell-types, together with the corresponding gene expression and mutation data, 20 recurrently mutated metabolic enzymatic genes were selected for oncometabolite prediction. Oncometabolite metabolite predictions were based on two different metabolic reaction

perturbations due to enzyme mutations. First, native oncometabolites concentrations could change due to metabolic enzymes Loss-of-Function mutations (e.g. FH and SDH LoF in Figure 3-1). And second, oncometabolites could change their concentrations due to metabolic enzymes Gain-of-Function mutations (e.g. IDH GoF in Figure 3-1). The aim of the work presented in this thesis was to completely describe the chemoinformatic approach used in Nam et al., 2014 for predicting oncometabolites due to GoF mutations. Out of the 20 selected enzymes, previously described, 6 mutated enzymatic genes were selected for GoF analysis, presenting missense mutations (i.e. the consensus type of mutation that is observed in IDH oncometabolite studies) and showing potential catalytic activity in cancer. Genes were associated with the corresponding enzymes and reactions in the global human metabolic network Recon 2(Thiele et al. 2013). In total 36 (see Table 3-1) metabolic reactions were extracted from Recon 2 and used for GoF analysis which is described below.
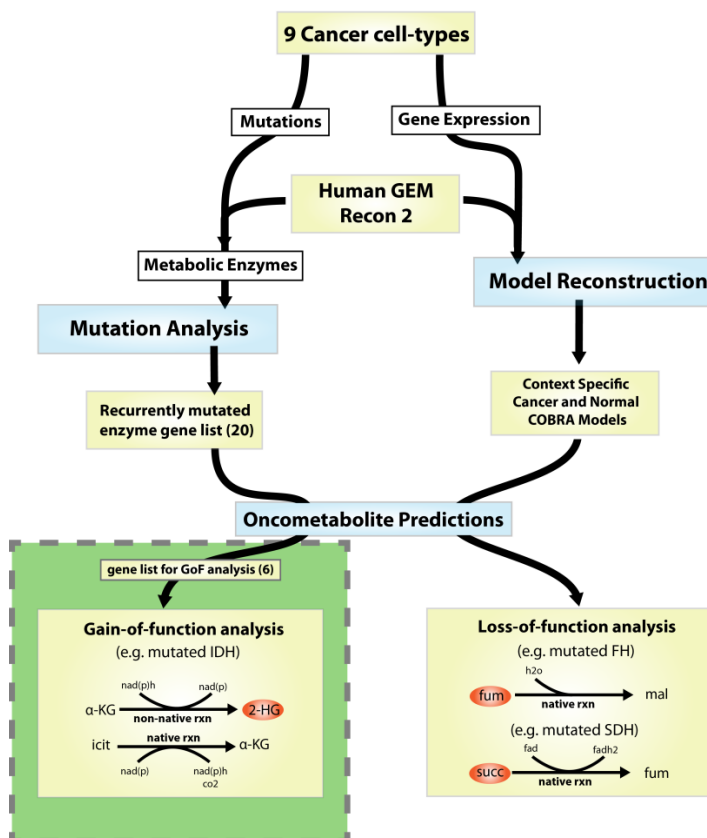
Figure 3-1: Nam et al., 2014 General Workflow to predict oncometabolites via context-specific genome-scale metabolic networks. Across 9 different cancer types, a list of recurrently mutated metabolic enzyme that could produce oncometabolites was generated. Then by using this list as a starting point oncometabolites were predicted by first simulating flux changes between Loss-of-function (LoF) cancer vs, normal *in silico* GEM analysis, and second by using a chemoinformatics approach which predicts potential promiscuous enzyme activities due to Gain-of-function (GoF) mutations. GoF analysis was highlighted in green, and represents the work presented in this thesis.

3.4.1 Gain-of-Function (GoF) analysis workflow

The framework for GoF oncometabolites predictions, involves five major steps, specifically described in Figure 3-2. First, based on the mutated enzyme gene list generated for GoF analysis, 36 metabolic reactions out of 6 genes were obtained from Recon 2. For each reaction substrates and products were extracted and the corresponding Molfiles were generated. For irreversible reactions only the substrate was taken into account, whilst for reversible reactions, both substrates and products were taken into account. Second, in order to further feed the synthetic reaction construction process, metabolites were selected from The Human Metabolome Database (HMDB) (Wishart et al. 2009). The HMDB contains 40.251 metabolites, this large amount of metabolites makes the synthetic reaction reconstruction process and further analysis computationally intractable. Thus, the HMDB was filtered and metabolites for synthetic reaction construction were selected on a reaction specific manner. For each reaction, substrates and/or products were compared to the HMDB. This was performed by using chemical fingerprints (CFP), where a dissimilarity Tanimoto coefficient (TC) matrix was calculated between each substrate and/or product of mutant enzymatic reactions and each metabolite present in the HMDB. Then, by using different TC cutoffs values (0.8 – 0.005), the number of HMDB metabolites to feed the synthetic reaction calculation for Biochemical Reaction Operators (BROs) simulation was determine. For each TC cutoffs values, HMDB metabolites scores of less than or equal to the cutoff values were counted. Then, in order to constraint computational efforts, for each mutated enzyme the number of HMDB for further BRO application was set to a value that allows to further proceed with the analysis in tractable time. The total number of HMDB used selected for each reaction are shown in Table 3-1, numbers in superscripts represents the TC cutoff value used for the metabolite selection. The compound list for BROs simulation was reduced from 40.251 to 550 metabolites in total. The third step in the workflow, correspond to the BRO simulations for calculating all potential synthetic reactions. For this 374 BROs were used to calculate all possible reactions that may occur and every compound that may be produced given the previous set of human metabolites. In total more than 200,000 synthetic reactions were generated.
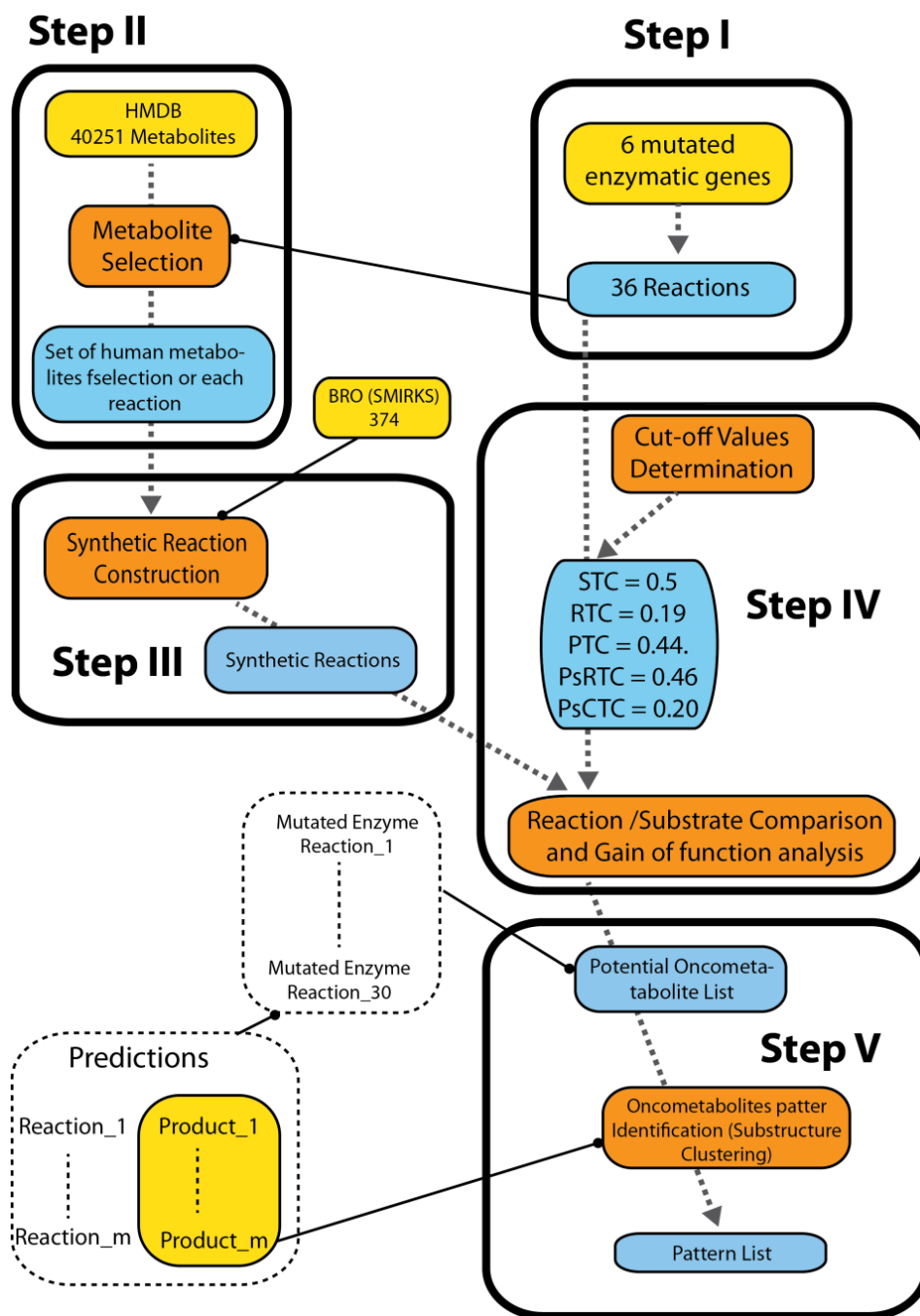
Figure 3-2: Gain-of-function (GoF) oncometabolites prediction pipeline**.** Synthetic reaction reconstruction and filtering pipeline for GoF prediction is described and outlined in five major steps. Process descriptions in yellow represent external inputs to the workflow. The specific processes or algorithm uses for the analysis were highlighted in orange, while inputs or outputs from these processes were depicted in blue.

Table 3-1. Mutated enzyme genes, reaction and general result information

Mut. Enz. ID	No. of HMDB for BROs	Reaction abbreviation	EC num	Gene (Entrez ID)	No. of Filtered rxns	No. of Indep. Prod.	No. of Indep. HMDB Prod.	No. of Indep. Subs.	No. of Indep. HMDB Subs
3	$12^{0.05}$	hco3[c] + atp[c] + accoa[c] -> malcoa[c] + adp[c] + h[c] + pi[c]		ACACB(32)	25	23	5	4	0
4	$12^{0.05}$	hco3[m] + accoa[m] + atp[m] -> malcoa[m] + h[m] + adp[m] + pi[m]	6.4.1.2	ACACB(32)	25	23	5	4	0
5	$8^{0.05}$	ACP[c] + accoa[c] <=> acACP[c] + coa[c]	2.3.1.38	FASN (2194)	2542	491	8	490	484
6	$2^{0.01}$	occoa[c] + 3*h[c] + 2*nadph[c] + malcoa[c] -> dcacoa[c] + h2o[c] + 2*nadp[c] + co2[c] + coa[c]	2.3.1.86	FASN (2194)	822	311	2	303	301
7	$2^{0.01}$	dcacoa[c] + 3*h[c] + 2*nadph[c] + malcoa[c] -> ddcacoa[c] + h2o[c] + 2*nadp[c] + co2[c] + coa[c]	2.3.1.86	FASN (2194)	822	311	2	303	301
8	$2^{0.01}$	ddcacoa[c] + 3*h[c] + 2*nadph[c] + malcoa[c] -> tdcoa[c] + h2o[c] + 2*nadp[c] + co2[c] + coa[c]	2.3.1.86	FASN (2194)	822	311	2	303	301
9	$2^{0.01}$	tdcoa[c] + 3*h[c] + 2*nadph[c] + malcoa[c] -> pmtcoa[c] + h2o[c] + 2*nadp[c] + co2[c] + coa[c]	2.3.1.86	FASN (2194)	822	311	2	303	301
10	$2^{0.01}$	pmtcoa[c] + 3*h[c] + 2*nadph[c] + malcoa[c] -> stcoa[c] + h2o[c] + 2*nadp[c] + co2[c] + coa[c]	2.3.1.86	FASN (2194)	822	311	2	303	301
11	$2^{0.01}$	accoa[c] + 9*h[c] + 6*nadph[c] + 3*malcoa[c] -> occoa[c] + 3*h2o[c] + 6*nadp[c] + 3*co2[c] + 3*coa[c]	2.3.1.86	FASN (2194)	777	301	2	313	311
12	$2^{0.01}$	accoa[c] + 20*h[c] + 14*nadph[c] + 7*malcoa[c] -> hdca[c] + 6*h2o[c] + 14*nadp[c] + 7*co2[c] + 8*coa[c]	2.3.1.85	FASN (2194)	777	301	2	313	311
13	$8^{0.05}$	ACP[c] + malcoa[c] <=> malACP[c] + coa[c]	2.3.1.85	FASN (2194)	1860	436	8	477	471
14	$81^{0.05}$	asp_L[c] + cbp[c] <=> cbasp[c] + h[c] + pi[c]	2.1.3.2	CAD (790)	21	21	4	12	4
15	$57^{0.05}$	gln_L[c] + h2o[c] + 2 atp[c] + hco3[c] -> 2 adp[c] + 2 h[c] + pi[c] + glu_L[c] + cbp[c]		CAD (790)	2886	1664	57	170	137
16	$7^{0.05}$	dhor_S[c] + h2o[c] <=> cbasp[c] + h[c]	3.5.2.3	CAD (790)	0	0	0	0	0
17	$69^{0.08}$	56dura[c] + nadp[c] <=> ura[c] + h[c] + nadph[c]	1.3.1.2	DPYD (1806)	13	2	2	8	8
18	$127^{0.08}$	56dthm[c] + nadp[c] <=> thym[c] + h[c] + nadph[c]	1.3.1.2	DPYD (1806)	25	4	4	12	12
41	$2^{0.01}$	HC01401[x] + nad[x] -> 3oddcoa[x] + h[x] + nadh[x]	1.1.1.35	FASN (2194)	782	285	2	254	252
42	$11^{0.02}$	HC01446[e] + h2o[e] <=> HC01440[e] + glc_D[e]	3.2.1.23	FASN (2194)	0	0	0	0	0
43	$16^{0.005}$	xol7ah3[c] + nad[c] <=> xol7ah2al[c] + h[c] + nadh[c]	1.1.1.1	FASN (2194)	2565	681	17	630	614
44	$16^{0.005}$	xoltriol[c] + o2[c] + h[c] + nadph[c] <=> xoltetrol[c] + h2o[c] + nadp[c]	1.14.13.15	FASN (2194)	4688	636	17	944	928
45	$2^{0.01}$	HC01459[x] + nad[x] -> cholcoaone[x] + h[x] + nadh[x]	1.1.1.35	FASN (2194)	749	285	2	222	220
46	$18^{0.005}$	xol7ah2[c] + nadp[c] <=> xol7ah[c] + h[c] + nadph[c]	1.1.1.50	FASN (2194)	3985	643	19	929	911
47	$17^{0.005}$	xoltriol[c] + nadp[c] <=> xoldioloneh[c] + h[c] + nadph[c]	1.1.1.50	FASN (2194)	3726	656	18	859	842
48	$7^{0.02}$	34dhmald[m] + h2o[m] + nadp[m] <=> 34dhoxmand[m] + 2 h[m] + nadph[m]	1.2.1.5	FASN (2194)	1033	218	15	173	170
49	$11^{0.02}$	3mox4hpac[m] + h2o[m] + nad[m] <=> homoval[m] + 2 h[m] + nadh[m]	1.2.1.5	FASN (2194)	1146	260	24	274	267
50	$4^{0.05}$	HC01590[c] + nadp[c] <=> HC01589[c] + h[c] + nadph[c]	2.3.1.85	FASN (2194)	65	16	4	17	16
51	$5^{0.05}$	HC01590[c] <=> HC01591[c] + h2o[c]	2.3.1.85	FASN (2194)	0	0	0	0	0
58	$2^{0.025}$	pail4p_hs[c] + atp[c] -> pail34p_hs[c] + adp[c] + h[c]	2.7.1.68	PIK3C2B (5287)	542	224	2	217	215
59	$2^{0.025}$	pail4p_hs[r] + atp[r] -> pail34p_hs[r] + adp[r] + h[r]	2.7.1.68	PIK3C2B (5287)	542	224	2	217	215
60	$2^{0.025}$	pail5p_hs[c] + atp[c] -> pail35p_hs[c] + adp[c] + h[c]		PIK3C2B (5287)	559	242	2	216	214
61	$2^{0.025}$	pail5p_hs[r] + atp[r] -> pail35p_hs[r] + adp[r] + h[r]		PIK3C2B (5287)	559	242	2	216	214
62	$2^{0.025}$	pail_hs[c] + atp[c] -> pail3p_hs[c] + adp[c] + h[c]	2.7.1.137	PIK3C2B (PIK3C2B (5287))	407	169	2	169	167
63	$2^{0.025}$	pail_hs[r] + atp[r] -> pail3p_hs[r] + adp[r] + h[r]	2.7.1.137	PIK3C2B (5287)	407	169	2	169	167
64	$2^{0.025}$	pail5p_hs[c] + atp[c] <=> CE5132[c] + adp[c] + h[c]	2.7.1.137	PIK3C2B (5287)	1071	242	2	227	225
56	$16^{0.03}$	pail345p_hs[c] + h2o[c] -> pail45p_hs[c] + pi[c]	3.1.3.67	PTENP1(5728)	5	1	1	5	5
57	$16^{0.03}$	pail345p_hs[n] + h2o[n] -> pail45p_hs[n] + pi[n]	3.1.3.67	PTENP (5728)	5	1	1	5	5

During the fourth step, in order to determine potential promiscuous reactions due to GoF mutations, calculated reactions were compared and filtered based on the IDH GoF mutation case. Five different cut off values were determined according to Figure

3-3. Rxn 1 represents the native reaction mechanism, while Rxn 2 corresponds to the new enzyme activity due to enzyme mutation. In order to characterized and quantify the difference between the reactions and the corresponding metabolites and cofactors, two different pairs of "pseudo" reactions were defined. PsRrxn 1 and 2 (Figure 3-3B) were respectively build from Rxn 1 and Rxn2. In this case only substrates and products were used to form both reactions. Same procedure was used in order to construct PsCrxn 1 and 2, (Figure 3-3C) but in this case only cofactors were taken into account to construct the "pseudo" reactions. From these pairs of reaction different metrics where calculated in order to numerically characterizes the chemical changes in the reactions due to mutation.
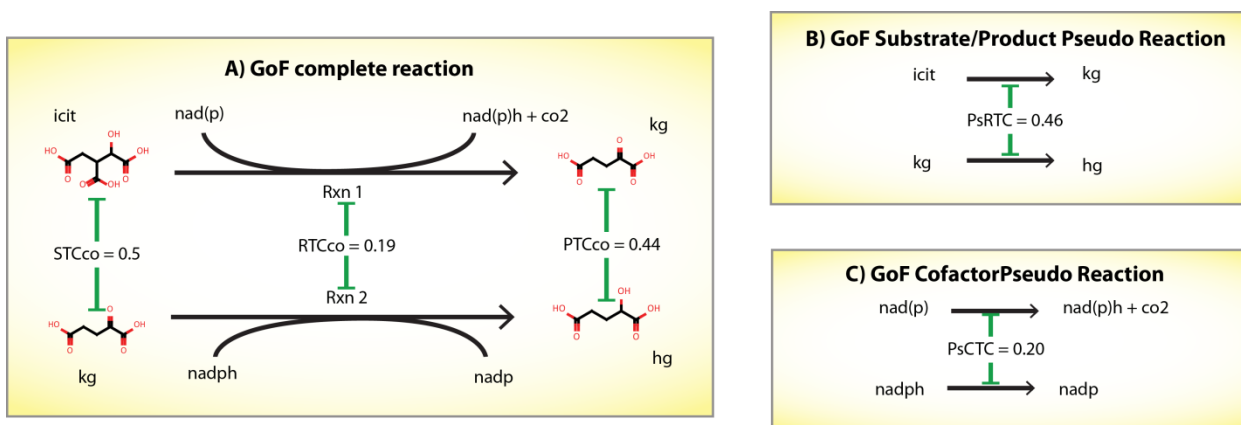
Figure 3-3: Cutoff values for reaction filtering based on the IDH GoF mutation example. A) GoF complete reaction parameters were defined. Specifically for the Substrate Tanimoto Coefficient (STco), Product Tanimoto Coefficient (PTCco) and the Reaction Tanimoto Coefficient (RTCco). B) GoF Substrate/Product Pseudo Reaction metric definition based only on the reactions products and substrates (PsRTC). C) GoF Cofactor Pseudo Reaction metric definition based only on the cofactors used in the reactions (PSCTC).
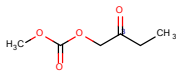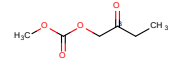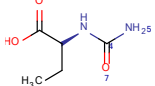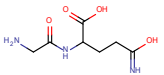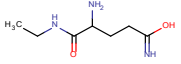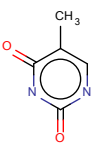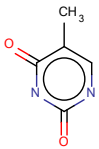
As mention above, these metrics were use as threshold values for filtering mutated enzyme reactions and computationally generated reactions. For the pair of reactions outlined in Figure 3-3A, the substrate Tanimoto coefficient (STCco) and the reaction Tanimoto coefficient (RTCco) were used for filtering. For "pseudo" reactions presented in B only the reaction Tanimoto coefficient (PsRTC) was taken into account. And finally for the last pair of "pseudo" reactions (C), the substrates Tanimoto coefficient (PsTCsub) and the product Tanimoto coefficient (PsCTpro) were chosen as valid thresholds. According to the later, reactions with scores of less than or equal to the cutoff values were chosen for further substructure oncometabolite search. In Table 3-1 was described the output regarding each studied reaction, in which is specified the total
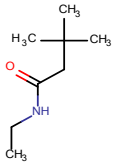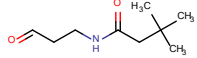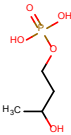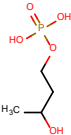
number of potential promiscuous reactions, together with the corresponding number of independent substrates and products. Finally, based on the filtered reactions, for each reaction-enzyme association, an MCS analysis on the predicted substrates and products was performed and the corresponding structural features of oncometabolites were predicted.
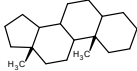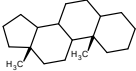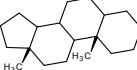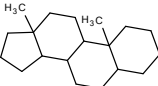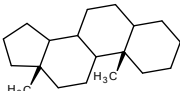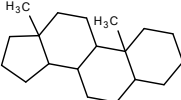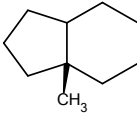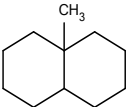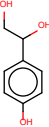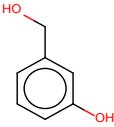
3.4.2 Substructures of potential oncometabolites as a result of GoF mutations predictions
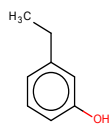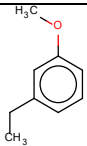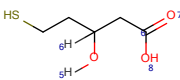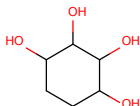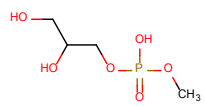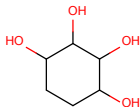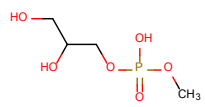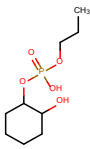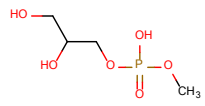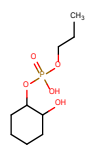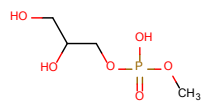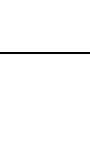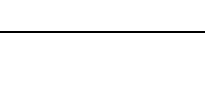
Among the reactions associated with the six mutants, approximately more than 35.000 promiscuous activities were calculated, 10.000 independent products and 9.000 independent substrates (see Table 3-1). In most of the cases, promiscuous substrates showed high similarity with the native substrate, and in order to demonstrate predominant substructures of promiscuous substrates and products a MCS analysis was performed. The chemical substructure shown in Table 3-2, correspond to the predominant substructure observed for promiscuous substrates and products. Notable, several promiscuous substrates and products did not have dominant substrates. Finally, in total 24 chemical substructures were identify as potential promiscuous compound substructures as features of GoF oncometabolites.
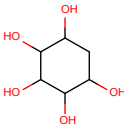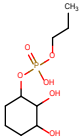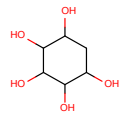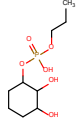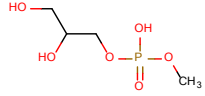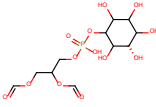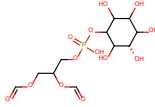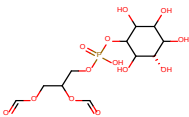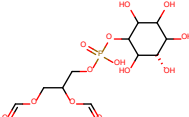
Table 3-2: Mutated enzymes associated synthetic promiscuous reactions and the corresponding predominant substructures of potential promiscuity oncometabolites

MutEnz ID	Gene (Entrez ID)	No. of predicted promiscuous reactions	No. of predicted substrate, dominant substructure	No. of predicted product, dominant substructure
3	ACACB (32)	25	4 N/A	23
4	ACACB (32)	25	4 N/A	23
14	CAD (790)	21	12 	21
15	CAD (790)	2886	170 	1644
17	DPYD (1806)	13	8 N/A	2 N/A
18	DPYD (1806)	25	12 	4
5	FASN (2194)	2542	490	491

6	FASN (2194)	822	303 N/A	311 N/A
7	FASN (2194)	822	303 N/A	311 N/A
8	FASN (2194)	822	303 N/A	311 N/A
9	FASN (2194)	822	303 N/A	311 N/A
10	FASN (2194)	822	303 N/A	311 N/A
11	FASN (2194)	777	311 N/A	301 N/A
12	FASN (2194)	777	311 N/A	301 N/A
13	FASN (2194)	1860	477	436

41	FASN (2194)	782	254 N/A	285 N/A
43	FASN (2194)	2565	630	681
44	FASN (2194)	4688	944	636
45	FASN (2194)	749	222 N/A	285 N/A
46	FASN (2194)	3985	929	643
47	FASN (2194)	3726	859	656
48	FASN (2194)	1033	173	218
49	FASN (2194)	1146	274	260

50	FASN (2194)	65	17 	16 N/A
58	PIK3C2B (5287)	542	217 	224
59	PIK3C2B (5287)	542	217 	224
60	PIK3C2B (5287)	559	216 	242
61	PIK3C2B (5287)	559	216 	242
62	PIK3C2B (5287)	407	169	169

63	PIK3C2B (5287)	407	169	169
64	PIK3C2B (5287)	1071	227	242
56	PTENP1 (5728)	5	5	1
57	PTENP1 (5728)	5	5	1

3.5 Conclusions

Based on a chemoinformatic approach in conjunction with Recon 2, a workflow for predicting potential oncometabolites was established. Starting from a list of mutated enzymes genes, described as GoF mutations, a range of promiscuous catalytic activities were inferred. In total 24 chemical substructures of oncometabolites resulting from the GoF analysis were predicted.

Since all the predictions were based only on the example showed for IDH, where an unexpected changed in the enzyme catalytic activity due to a missense mutation was observed, it is mandatory to experimentally validate these results and test the predictive capacity of the proposed method. Nevertheless, this attempt represents the first example of oncometabolites predictions based on a chemoinformatic approach. Further improvements might be implemented by taking into account structural biology information and analysis in the context of GEMs regarding enzymes spatial and mechanistic changes due to GoF mutation and further connection with oncometabolites. Moreover, more examples such as IDH are needed to expand the predictive capacity of the method and make it more reliable. Finally, despite the drawbacks of the implemented method, this work represents the first attempt to predict oncometabolites structural patterns, due to GoF metabolic enzyme mutations based on a chemoinformatic approach. These predictions, could be used as a basis for cancer diagnosis and prognosis, and also be highly beneficial in biomarkers identification.

Appendices

Appendix 1

Figure A1-1 Carbon fixation and electron transfer metabolism Flux Variability Analysis (FVA) results

FVA was performed under autotrophic conditions using as electron donors ferrous ion (green), tetrathionate (purple), and thiosulfate (red). Carbon fixation and electron transfer metabolism were outlined. Reactions in light blue represent the flow of electrons through the membrane. Each reaction in the figure was associated with a specific graph, in which is represented the calculated FVA range for the three different condition studied.
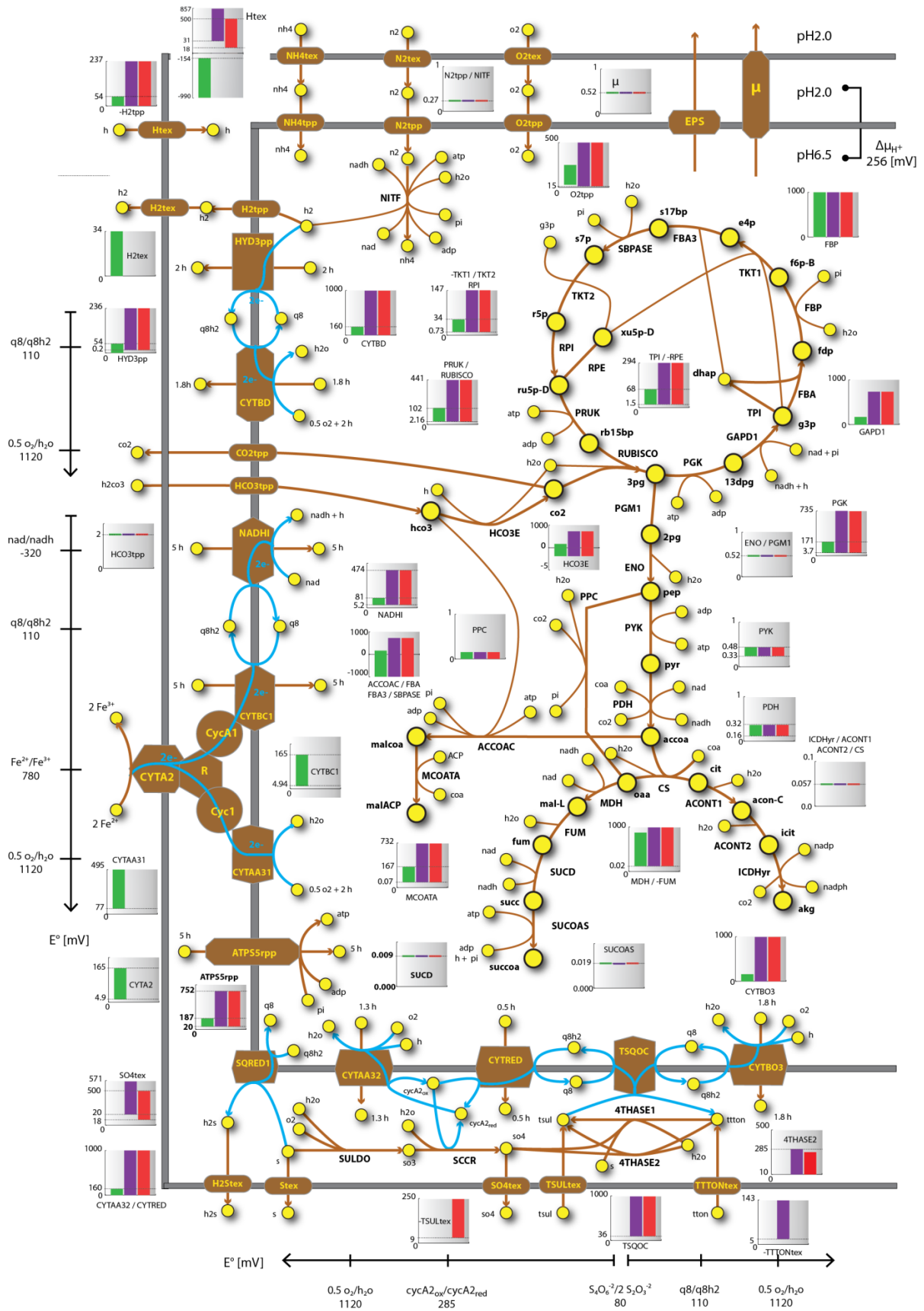
Figure A1-2 Carbon fixation and electron transfer metabolism under oxygen limitations

FBA was performed under autotrophic conditions using as electron donors ferrous ion (green). This analysis was executed by setting the carbon uptake rate in the form of h2co3 to 2.34 mmol/gDW/h, oxygen was varied between 0 and 60 mmol/gDW/h, and running FBA to maximize the flux through BOF. Carbon fixation and electron transfer pathways were outlined. Reactions in light blue represent the flow of electrons through the membrane. Each reaction in the figure was associated with a specific graph, in which is represented the FBA prediction range for the three different condition studied. Lines in red represents two different kinds of simulations: i) simulations were a metabolic co2 transport outside the cell was incorporated into the model and setting h2co3 uptake rate upper and lower bound to -2.34, and ii) where no metabolic co2 transport was incorporated but setting the h2co3 uptake rate lower bound to -2.34 and upper bound to 1000. Lines in purple shows the results without metabolic co2 transport incorporation, and setting the h2co3 uptake rate upper and lower bound to -2.34. Lines in light blue shows the obtained discrepancies between i) and ii).
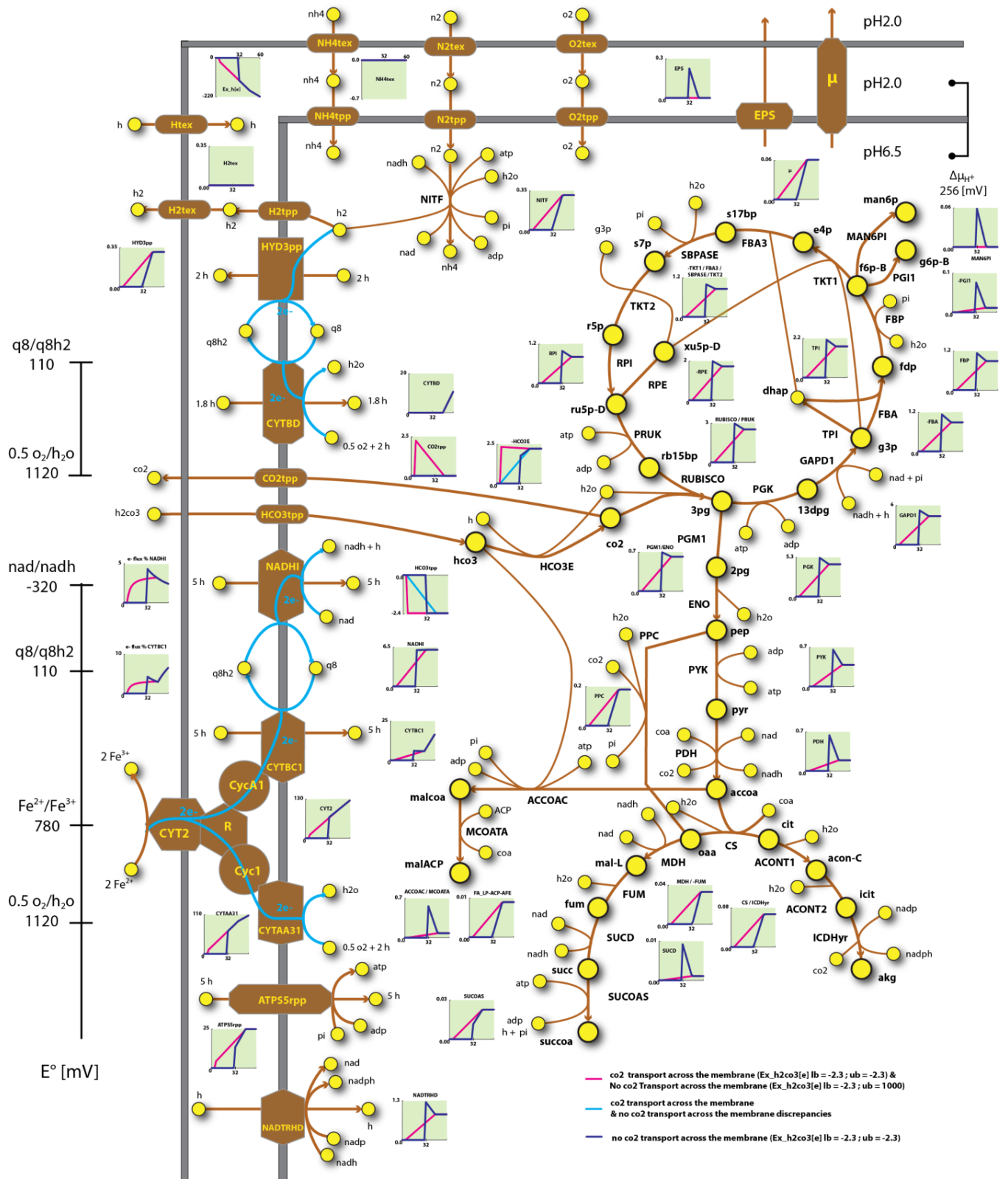
Table A1-1: Biomass objective function specification

Macromolecule	overall [gr/gr DW]	metabolite (BiGG)	Experimental Composition	composition (molar fraction)	mmol/gDW (Calc.)	difference b/w metabolite and structural component	MWcorr (mg/mmol)	mg (if total 1 mmol)	total wt (mg)	composition (weight fraction)	g/gDW
Protein			mg/ 100 mg Protein [1]	mmol AA/mg Protein	**5.57**						
	0.6 [2]	ala-L	7.8	0.0011	0.66	h2o	71				
		arg-L	7.3	0.0005	0.28	h2o	157				
		asn-L	5.15	0.0005	0.27	h2o	114				
		asp-L	5.15	0.0005	0.27	h2o	114				
		cys-L	0	0.0000	0.00	h2o	103				
		gln-L	5.55	0.0004	0.26	h2o	128				
		glu-L	5.55	0.0004	0.26	h2o	128				
		gly	5.2	0.0009	0.55	h2o	57				
		his-L	5.5	0.0004	0.24	h2o	137				
		ile-L	5.4	0.0005	0.29	h2o	113				
		leu-L	9.6	0.0008	0.51	h2o	113				
		lys-L	7.1	0.0005	0.33	h2o	129				
		met-L	3.7	0.0003	0.17	h2o	131				
		phe-L	4.4	0.0003	0.18	h2o	147				
		pro-L	4.4	0.0005	0.27	h2o	97				
		ser-L	3.4	0.0004	0.23	h2o	87				
		thr-L	4.4	0.0004	0.26	h2o	101				
		trp-L	0	0.0000	0.00	h2o	186				
		tyr-L	3.7	0.0002	0.14	h2o	163				
		val-L	6.7	0.0007	0.41	h2o	99				
DNA			mmol /mmol %*		**0.075**		mg dNTP/mmol DNA			mg dNTP / mg DNA	gdNTP/gDW
	0.023 [†]	datp	0.21		0.015	ppi	312	65	307	0.21	0.005
		dctp	0.30		0.022	ppi	286	85		0.28	0.006
		dgtp	0.29		0.022	ppi	328	96		0.31	0.007
		dttp	0.21		0.015	ppi	304	63		0.20	0.005
RNA			mmol /mmol %*		**0.44**		mg NTP/mmol RNA			mg NTP / mg RNA	gNTP/gDW
	0.14 [2]	ctp	0.29		0.13	ppi	304	89	321	0.28	0.039
		gtp	0.30		0.13	ppi	344	104		0.32	0.045

Category	Coefficient	Metabolite	mg/100mg biomass[1] / mg/mmol soluble pool	mmol / mmolsoluble pool	Value	Byproduct				
		utp	0.21		0.09	ppi	305	63	0.20	0.028
		atp*	0.20		0.09	ppi	328	65	0.20	0.028
Glycogen	0.025 [3]	glycogen			**0.154**	none	162			
Murein	0.025 [3]	murein5px4p			**0.013**	none	1893			
LPS	0.034 [3]	lps_AFE			**0.024**	none	1400			
Phospholipid	0.093 [3]	ppl_AFE			**0.067**	none	723			
Inorganic Ions			**mg/100mg biomass [1]**		**0.78**					
	0.03 [1]	k	0.17		0.04	none	39			
		na	0.33		0.14	none	23			
		mg2	0.58		0.24	none	24			
		ca2	0.35		0.088	none	40			
		fe2	0.375		0.067	none	56			
		fe3	0.375		0.067	none	56			
		cu2	0.009		0.001	none	64			
		mn2	0.17		0.03	none	55			
		zn2	0.014		0.002	none	64			
		so4	0.88		0.09	none	96			
Soluble Pool			**mg/mmol soluble pool**	**mmol / mmolsoluble pool**	**0.00498**					
	0.0034 [3]	ptrc	546.31	0.7155	0.00358	none	90			
		spmd	96.08	0.1450	0.00073	none	148			
		accoa	4.45	0.0060	0.00003	none	806			
		coa	0.00	0.0036	0.00002	none	764			
		succoa	0.00	0.0021	0.00001	none	863			
		nad	0.00	0.0384	0.00019	none	662			
		nadp	0.00	0.0024	0.00001	none	740			
		nadph	0.00	0.0072	0.00004	none	741			
		fad	0.00	0.0048	0.00002	none	784			
		thf	0.00	0.0048	0.00002	none	443			
		mlthf	0.00	0.0048	0.00002	none	467			
		5mthf	3.76	0.0048	0.00002	none	469			
		thmpp	2.13	0.0048	0.00002	none	422			
		q8h2	2.19	0.0048	0.00002	none	729			
		pydx5p	2.20	0.0048	0.00002	none	245			
		hemeO	2.26	0.0048	0.00002	none	781			
		pheme	2.03	0.0048	0.00002	none	559			

	Metabolite					
	sheme	3.50	0.0048	0.00002	none	853
	gthrd	2.68	0.0048	0.00002	none	302
	udcpdp	0.80	0.0012	0.00001	none	924
	10fthf	4.44	0.0048	0.00002	none	471
	chor	1.08	0.0048	0.00002	none	224
	amet	1.92	0.0048	0.00002	none	399
	ribflv	1.81	0.0048	0.00002	none	376
		678.82	**mg/mmol**			
maintenance	atp*			139		
maintenance	h2o*			139		
maintenance +RNA	atp*			139.1		
maintenance - AA	h2o*			139.0		
maintenance [‡]	adp			139		
maintenance [‡]	h			139		
maintenance [‡]	pi*			139		
sum of dNTPs	ppi*			0.075		
sum of NTPs	ppi*			0.436		
sum of AA	h2o*			5.569		
NT + dNT	ppi*			0.511		
maintenance [‡]	pi*			139		

(The first column group is labelled **PRODUCTS** for the lower green section.)

*   Nucleotide content according to the genome composition. See Table A1-2.

†   Obtain form correlation with different microorganisms. See Table A1-2.

‡   GAM maintenance coefficients obtain from parameter estimation.

Table A1-2: RNA content in *Acidithiobacillus ferrooxidans* (Peterson et al. 2001)

AA Name	Codon Triplet				Number of Triplet Occurrences	% Triplet Frequency	% Triplet Frequency for AA	A	C	G	U
Ala (A)	GCC	G	C	C	44970	5.01%	44.91%	0	89940	44970	0
Ala (A)	GCA	G	C	A	12969	1.44%	12.95%	12969	12969	12969	0
Ala (A)	GCG	G	C	G	32113	3.58%	32.07%	0	32113	64226	0
Ala (A)	GCT	G	C	T	10086	1.12%	10.07%	0	10086	10086	10086
					100138	11.15%	100%				
Arg (R)	CGA	C	G	A	4271	0.47%	6.77%	4271	4271	4271	0
Arg (R)	AGG	A	G	G	3193	0.35%	5.06%	3193	0	6386	0
Arg (R)	CGT	C	G	T	12099	1.34%	19.18%	0	12099	12099	12099
Arg (R)	CGG	C	G	G	16592	1.85%	26.30%	0	16592	33184	0
Arg (R)	AGA	A	G	A	2037	0.22%	3.23%	4074	0	2037	0
Arg (R)	CGC	C	G	C	24887	2.77%	39.45%	0	49774	24887	0
					63079	7%	100%				
Asn (N)	AAC	A	A	C	13444	1.49%	53.84%	26888	13444	0	0
Asn (N)	AAT	A	A	T	11527	1.28%	46.16%	23054	0	0	11527
					24971	3%	100%				
Asp (D)	GAC	G	A	C	21542	2.40%	47.75%	21542	21542	21542	0
Asp (D)	GAT	G	A	T	23576	2.63%	52.25%	23576	0	23576	23576
					45118	5%	100%	0	0	0	0
Cys (C)	TGC	T	G	C	5946	0.66%	65.13%	0	5946	5946	5946
Cys (C)	TGT	T	G	T	3184	0.35%	34.87%	0	0	3184	6368
					9130	1%	100%				
Gln (Q)	CAA	C	A	A	10167	1.13%	26.99%	20334	10167	0	0
Gln (Q)	CAG	C	A	G	27496	3.06%	73.01%	27496	27496	27496	0
					37663	4%	100%				
Glu (E)	GAG	G	A	G	22882	2.55%	47.88%	22882	0	45764	0
Glu (E)	GAA	G	A	A	24909	2.77%	52.12%	49818	0	24909	0
					47791	5%	100%				
Gly (G)	GGA	G	G	A	10183	1.13%	13.79%	10183	0	20366	0
Gly (G)	GGT	G	G	T	17033	1.90%	23.07%	0	0	34066	17033
Gly (G)	GGG	G	G	G	14788	1.64%	20.03%	0	0	44364	0
Gly (G)	GGC	G	G	C	31827	3.55%	43.11%	0	31827	63654	0
					73831	8%	100%				
His (H)	CAC	C	A	C	11075	1.23%	45.60%	11075	22150	0	0
His (H)	CAT	C	A	T	13212	1.47%	54.40%	13212	13212	0	13212
					24287	3%	100%				
Ile (I)	ATT	A	T	T	13889	1.54%	29.72%	13889	0	0	27778
Ile (I)	ATC	A	T	C	28118	3.13%	60.18%	28118	28118	0	28118
Ile (I)	ATA	A	T	A	4720	0.52%	10.10%	9440	0	0	4720
					46727	5%	100%				
Leu (L)	CTT	C	T	T	8071	0.90%	8.23%	0	8071	0	16142
Leu (L)	CTG	C	T	G	47842	5.33%	48.80%	0	47842	47842	47842
Leu (L)	TTA	T	T	A	3342	0.37%	3.41%	3342	0	0	6684
Leu (L)	CTA	C	T	A	3023	0.33%	3.08%	3023	3023	0	3023
Leu (L)	CTC	C	T	C	19838	2.21%	20.23%	0	39676	0	19838
Leu (L)	TTG	T	T	G	15925	1.77%	16.24%	0	0	15925	31850
					98041	11%	100%				
Lys (K)	AAG	A	A	G	15145	1.68%	55.54%	30290	0	15145	0
Lys (K)	AAA	A	A	A	12126	1.35%	44.46%	36378	0	0	0
					27271	3%	100%				
Met (M)	ATG	A	T	G	22121	2.46%	100.00%	22121	0	22121	22121
					22121	2.46%	100.00%				
Phe (F)	TTT	T	T	T	14760	1.64%	46.47%	0	0	0	44280
Phe (F)	TTC	T	T	C	17001	1.89%	53.53%	0	17001	0	34002
					31761	4%	100%				
Pro (P)	CCG	C	C	G	18418	2.05%	38.51%	0	36836	18418	0
Pro (P)	CCC	C	C	C	17733	1.97%	37.07%	0	53199	0	0
Pro (P)	CCA	C	C	A	5451	0.60%	11.40%	5451	10902	0	0
Pro (P)	CCT	C	C	T	6229	0.69%	13.02%	0	12458	0	6229
					47831	5%	100%				
Ser (S)	AGT	A	G	T	6671	0.74%	13.84%	6671	0	6671	6671
Ser (S)	TCA	T	C	A	2847	0.31%	5.91%	2847	2847	0	2847
Ser (S)	AGC	A	G	C	14340	1.59%	29.75%	14340	14340	14340	0
Ser (S)	TCG	T	C	G	7208	0.80%	14.95%	0	7208	7208	7208

Ser (S)	TCC	T	C	C	12957	1.44%	26.88%	0	25914	0	12957
Ser (S)	TCT	T	C	T	4186	0.46%	8.68%	0	4186	0	8372
					48209	**5%**	**100%**				
STOP	TAG	T	A	G	508	0.05%	15.99%	508	0	508	508
STOP	TGA	T	G	A	1897	0.21%	59.71%	1897	0	1897	1897
STOP	TAA	T	A	A	772	0.08%	24.30%	1544	0	0	772
					3177	**0%**	**100%**				
Thr (T)	ACA	A	C	A	4119	0.45%	9.30%	8238	4119	0	0
Thr (T)	ACT	A	C	T	4246	0.47%	9.59%	4246	4246	0	4246
Thr (T)	ACC	A	C	C	23509	2.62%	53.09%	23509	47018	0	0
Thr (T)	ACG	A	C	G	12404	1.38%	28.01%	12404	12404	12404	0
					44278	**5%**	**100%**				
Trp (W)	TGG	T	G	G	14323	1.59%	100.00%	0	0	28646	14323
					14323	**1.59%**	**100.00%**				
Tyr (Y)	TAT	T	A	T	12914	1.44%	55.06%	12914	0	0	25828
Tyr (Y)	TAC	T	A	C	10539	1.17%	44.94%	10539	10539	0	10539
					23453	**3%**	**100%**	0	0	0	0
Val (V)	GTC	G	T	C	19515	2.17%	30.90%	0	19515	19515	19515
Val (V)	GTG	G	T	G	27826	3.10%	44.06%	0	0	55652	27826
Val (V)	GTT	G	T	T	8036	0.89%	12.72%	0	0	8036	16072
Val (V)	GTA	G	T	A	7780	0.86%	12.32%	7780	0	7780	7780
					63157	**7%**	**100%**				

Grand Totals		896357	0.9964	

	ATP	CTP	GTP	UTP
SUM	534056	783090	812090	559835
In stop codon	3949	0	2405	3177
Total	530107	783090	809685	556658
	ATP	CTP	GTP	UTP
	0.197835076	0.292247923	0.302173	0.207744

Table A1-3: Single, double and tripe knock-out and growth coupled design results

KO Reactions			EPS	Biomass
'FUM'			0.206	0.011
'MDH'			0.206	0.011
'ADA'	'FUM'		0.223	0.007
'ADA'	'MDH'		0.223	0.007
'ADPT'	'FUM'		0.223	0.007
'ADPT'	'MDH'		0.223	0.007
'PUNP1'	'FUM'		0.223	0.007
'PUNP1'	'MDH'		0.223	0.007
'FUM'	'ADA'		0.223	0.007
'FUM'	'ADPT'		0.223	0.007
'FUM'	'PUNP1'		0.223	0.007
'MDH'	'ADA'		0.223	0.007
'MDH'	'ADPT'		0.223	0.007
'MDH'	'PUNP1'		0.223	0.007
'INSK'	'PUNP2'	'FUM'	0.223	0.007
'INSK'	'PUNP2'	'MDH'	0.223	0.007
'INSK'	'FUM'	'PUNP2'	0.223	0.007
'INSK'	'FUM'	'IMPD'	0.236	0.004
'INSK'	'FUM'	'GUAD'	0.223	0.007
'INSK'	'FUM'	'HXAND'	0.223	0.007
'INSK'	'IMPD'	'FUM'	0.236	0.004
'INSK'	'IMPD'	'MDH'	0.236	0.004
'INSK'	'GUAD'	'FUM'	0.223	0.007
'INSK'	'GUAD'	'MDH'	0.223	0.007
'INSK'	'MDH'	'PUNP2'	0.223	0.007
'INSK'	'MDH'	'IMPD'	0.236	0.004
'INSK'	'MDH'	'GUAD'	0.223	0.007
'INSK'	'MDH'	'HXAND'	0.223	0.007
'INSK'	'HXAND'	'FUM'	0.223	0.007
'INSK'	'HXAND'	'MDH'	0.223	0.007
'ADPT'	'ADNSE'	'FUM'	0.223	0.007
'ADPT'	'ADNSE'	'MDH'	0.223	0.007
'ADPT'	'FUM'	'ADNSE'	0.223	0.007
'ADPT'	'FUM'	'HCYSMT'	0.224	0.007
'ADPT'	'MDH'	'ADNSE'	0.223	0.007
'ADPT'	'MDH'	'HCYSMT'	0.224	0.007
'ADPT'	'HCYSMT'	'FUM'	0.224	0.007
'ADPT'	'HCYSMT'	'MDH'	0.224	0.007
'ADNSE'	'ADPT'	'FUM'	0.223	0.007
'ADNSE'	'ADPT'	'MDH'	0.223	0.007
'ADNSE'	'PUNP1'	'FUM'	0.223	0.007
'ADNSE'	'PUNP1'	'MDH'	0.223	0.007
'ADNSE'	'FUM'	'ADPT'	0.223	0.007
'ADNSE'	'FUM'	'PUNP1'	0.223	0.007
'ADNSE'	'MDH'	'ADPT'	0.223	0.007
'ADNSE'	'MDH'	'PUNP1'	0.223	0.007
'PUNP1'	'ADNSE'	'FUM'	0.223	0.007
'PUNP1'	'ADNSE'	'MDH'	0.223	0.007
'PUNP1'	'FUM'	'ADNSE'	0.223	0.007
'PUNP1'	'FUM'	'HCYSMT'	0.224	0.007
'PUNP1'	'MDH'	'ADNSE'	0.223	0.007
'PUNP1'	'MDH'	'HCYSMT'	0.224	0.007
'PUNP1'	'HCYSMT'	'FUM'	0.224	0.007
'PUNP1'	'HCYSMT'	'MDH'	0.224	0.007
'PUNP2'	'INSK'	'FUM'	0.223	0.007
'PUNP2'	'INSK'	'MDH'	0.223	0.007
'PUNP2'	'FUM'	'INSK'	0.223	0.007
'PUNP2'	'MDH'	'INSK'	0.223	0.007
'FUM'	'INSK'	'PUNP2'	0.223	0.007
'FUM'	'INSK'	'IMPD'	0.236	0.004
'FUM'	'INSK'	'GUAD'	0.223	0.007
'FUM'	'INSK'	'HXAND'	0.223	0.007
'FUM'	'ADPT'	'ADNSE'	0.223	0.007
'FUM'	'ADPT'	'HCYSMT'	0.224	0.007
'FUM'	'ADNSE'	'ADPT'	0.223	0.007
'FUM'	'ADNSE'	'PUNP1'	0.223	0.007

KO Reactions			EPS	Biomass
'FUM'	'PUNP1'	'ADNSE'	0.223	0.007
'FUM'	'PUNP1'	'HCYSMT'	0.224	0.007
'FUM'	'PUNP2'	'INSK'	0.223	0.007
'FUM'	'IMPD'	'INSK'	0.236	0.004
'FUM'	'GUAD'	'INSK'	0.223	0.007
'FUM'	'GUAPRT'	'UPPRT'	0.223	0.007
'FUM'	'GUAPRT'	'URIP'	0.223	0.007
'FUM'	'HCYSMT'	'ADPT'	0.224	0.007
'FUM'	'HCYSMT'	'PUNP1'	0.224	0.007
'FUM'	'UPPRT'	'GUAPRT'	0.223	0.007
'FUM'	'URIP'	'GUAPRT'	0.223	0.007
'FUM'	'HXAND'	'INSK'	0.223	0.007
'IMPD'	'INSK'	'FUM'	0.236	0.004
'IMPD'	'INSK'	'MDH'	0.236	0.004
'IMPD'	'FUM'	'INSK'	0.236	0.004
'IMPD'	'MDH'	'INSK'	0.236	0.004
'GUAD'	'INSK'	'FUM'	0.223	0.007
'GUAD'	'INSK'	'MDH'	0.223	0.007
'GUAD'	'FUM'	'INSK'	0.223	0.007
'GUAD'	'MDH'	'INSK'	0.223	0.007
'GUAPRT'	'FUM'	'UPPRT'	0.223	0.007
'GUAPRT'	'FUM'	'URIP'	0.223	0.007
'GUAPRT'	'MDH'	'UPPRT'	0.223	0.007
'GUAPRT'	'MDH'	'URIP'	0.223	0.007
'GUAPRT'	'UPPRT'	'FUM'	0.223	0.007
'GUAPRT'	'UPPRT'	'MDH'	0.223	0.007
'GUAPRT'	'URIP'	'FUM'	0.223	0.007
'GUAPRT'	'URIP'	'MDH'	0.223	0.007
'MDH'	'INSK'	'PUNP2'	0.223	0.007
'MDH'	'INSK'	'IMPD'	0.236	0.004
'MDH'	'INSK'	'GUAD'	0.223	0.007
'MDH'	'INSK'	'HXAND'	0.223	0.007
'MDH'	'ADPT'	'ADNSE'	0.223	0.007
'MDH'	'ADPT'	'HCYSMT'	0.224	0.007
'MDH'	'ADNSE'	'ADPT'	0.223	0.007
'MDH'	'ADNSE'	'PUNP1'	0.223	0.007
'MDH'	'PUNP1'	'ADNSE'	0.223	0.007
'MDH'	'PUNP1'	'HCYSMT'	0.224	0.007
'MDH'	'PUNP2'	'INSK'	0.223	0.007
'MDH'	'IMPD'	'INSK'	0.236	0.004
'MDH'	'GUAD'	'INSK'	0.223	0.007
'MDH'	'GUAPRT'	'UPPRT'	0.223	0.007
'MDH'	'GUAPRT'	'URIP'	0.223	0.007
'MDH'	'HCYSMT'	'ADPT'	0.224	0.007
'MDH'	'HCYSMT'	'PUNP1'	0.224	0.007
'MDH'	'UPPRT'	'GUAPRT'	0.223	0.007
'MDH'	'URIP'	'GUAPRT'	0.223	0.007
'MDH'	'HXAND'	'INSK'	0.223	0.007
'HCYSMT'	'ADPT'	'FUM'	0.224	0.007
'HCYSMT'	'ADPT'	'MDH'	0.224	0.007
'HCYSMT'	'PUNP1'	'FUM'	0.224	0.007
'HCYSMT'	'PUNP1'	'MDH'	0.224	0.007
'HCYSMT'	'FUM'	'ADPT'	0.224	0.007
'HCYSMT'	'FUM'	'PUNP1'	0.224	0.007
'HCYSMT'	'MDH'	'ADPT'	0.224	0.007
'HCYSMT'	'MDH'	'PUNP1'	0.224	0.007
'UPPRT'	'FUM'	'GUAPRT'	0.223	0.007
'UPPRT'	'GUAPRT'	'FUM'	0.223	0.007
'UPPRT'	'GUAPRT'	'MDH'	0.223	0.007
'UPPRT'	'MDH'	'GUAPRT'	0.223	0.007
'URIP'	'FUM'	'GUAPRT'	0.223	0.007
'URIP'	'GUAPRT'	'FUM'	0.223	0.007
'URIP'	'GUAPRT'	'MDH'	0.223	0.007
'URIP'	'MDH'	'GUAPRT'	0.223	0.007
'HXAND'	'INSK'	'FUM'	0.223	0.007
'HXAND'	'INSK'	'MDH'	0.223	0.007
'HXAND'	'FUM'	'INSK'	0.223	0.007
'HXAND'	'MDH'	'INSK'	0.223	0.007

Appendix 2

Figure A2-1 Predicted Pathways Figures

Compounds are represented within circles, yellow corresponds to *E. coli* native metabolites according to iJO1366, intermediate compounds are colored in brown and the target compound in blue. Reactions are represented by arrows. The color specifies whether the reaction is predicted as promiscuous (red) or known (green). Pathways were highlighted whether they were identified as already experimentally implemented in *E. coli* (purple) or reported as computationally generated (green) by previous work. Furthermore, for each target compound and set of pathways prediction, two different tables were defined. The first table specifies each predicted reaction ID, substrate TC / Promiscuity Score, EC number, Species and ΔrG'. The second table specifies each predicted pathway ID, reaction chain according to the reaction IDs, the metabolism precursor, the pathway index and the productivity under aerobic and aerobic conditions using glucose, xylose and glycerol as a substrate. For the productivity calculations, each carbon source was set to 120 C-mmol gDW-1h-1, O2 was 20 mmol gDW-1h-1 when specified, a minimal growth rate to sustain growth was set to 0.1 h-1 (as set by the amount of flux necessary through the BOFcore) and FBA was used to maximize the flux through each of the exchange reaction in the model for the targeted products. In this table experimental, computational and the maximum productivity pathways for each condition are colored in purple, green and orange respectively.
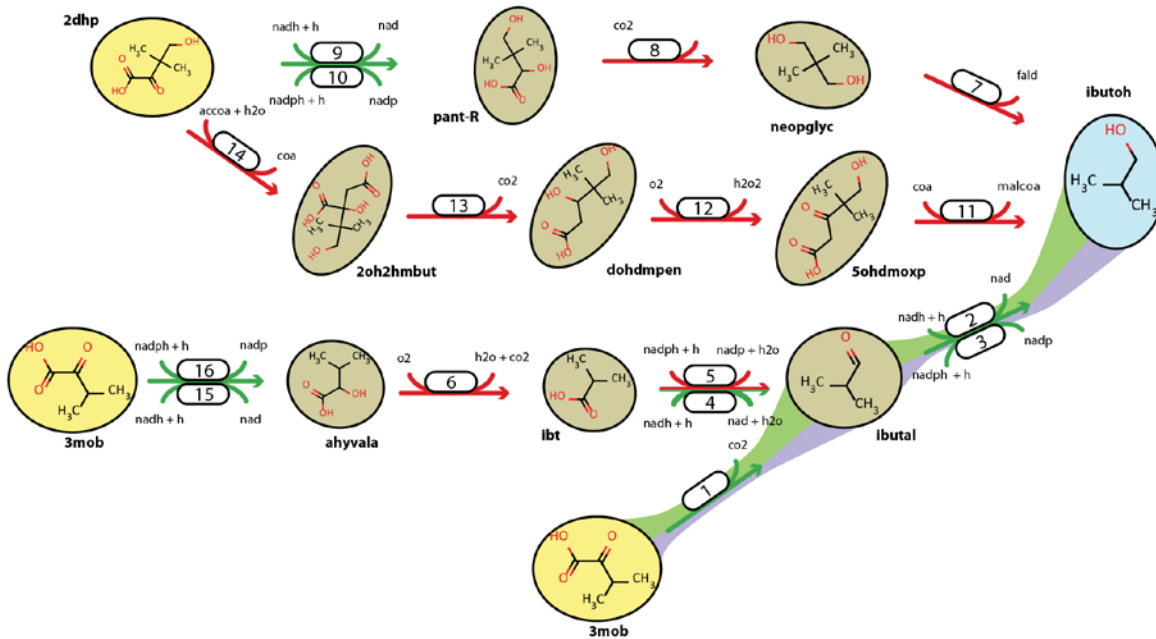
Legend:
- *E. coli* Native Precursor
- *E. coli* Non Native Intermediate
- Target Compound
- *E. coli* Experimentally Implemented Pathways
- *E. coli* Computationally Generated Pathways
- *E. coli* Maximum Productivity Pathways
- Promiscuous Reaction
- Known Reaction (BRENDA)

isobutanol

Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0/0	4.1.1.1	Lactococcus lactis	-47.4
2	0/0	1.1.1.1	Sulfolobus solfataricus	-53.2
3	0/0	1.1.1.2	Thermococcus litoralis	-73.1
4	0/0	1.2.1.19	Rattus norvegicus	6.4
5	0.19/0.88	1.2.1.22	Escherichia coli*	-13.4
6	0.34/0.77	1.13.12.4	Mycobacterium phlei	-462
7	0.57/1.07	4.1.2.46	Linum usitatissimum	4.9
8	0.24/1.12	4.1.1.73	Pseudomonas sp	-43.7
9	0/0	1.1.1.169	Salmonnella enetrica	-56.5
10	0/0	1.1.1.169	Saccharomyces cerevisiae	-76.3
11	0.44/1.13	2.3.1.29	Escherichia coli	-31.6
12	0.35/1.00	1.1.3.15	Mus musculus	-82.9
13	0.09/1.17	4.1.1.6	Aspergillus terreus	-36.3
14	0.15/0.6	2.3.3.13	Salmonella enterica	-32.8
15	0/0	1.1.1.27	Cupriavidus necator	-56.6
16	0/0	1.1.1.169	Escherichia coli*	-76.5

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1 exp/comp	2, 1	3mob	1.00	19.27	19.27	19.93	19.28	19.28	22.19
2 exp/comp	3, 1	3mob	1.00	19.27	18.19	16.42	19.28	19.21	21.88
3	7, 8, 9	2dhp	0.33	15.73	14.32	16.63	18.73	18.51	21.29
4	7, 8, 10	2dhp	0.33	14.08	12.73	13.60	18.39	18.17	21.01
5	11, 12, 13, 14	2dhp	0.00	13.56	11.57	10.56	18.59	18.38	21.15
6	2, 4, 6, 15	3mob	0.50	0.00	0.00	0.00	16.53	16.53	19.38
7	2, 4, 6, 16	3mob	0.75	0.00	0.00	0.00	16.53	16.53	19.38
8	2, 5, 6, 15	3mob	0.50	0.00	0.00	0.00	16.53	16.53	19.38
9	2, 5, 6, 16	3mob	0.50	0.00	0.00	0.00	16.53	16.53	19.38
10	3, 4, 6, 15	3mob	0.75	0.00	0.00	0.00	16.53	16.53	19.38
11	3, 4, 6, 16	3mob	0.75	0.00	0.00	0.00	16.53	16.53	19.38
12	3, 5, 6, 15	3mob	0.50	0.00	0.00	0.00	16.53	16.53	19.38
13	3, 5, 6, 16	3mob	0.50	0.00	0.00	0.00	16.53	16.40	19.38
	Max. Exp & Comp			19.27	19.27	19.93	19.28	19.28	22.19
	Max. GEM-Path			15.73	14.32	16.63	18.73	18.51	21.29
	Exp. & Comp .v/s GEM-Path			-18%	-26%	-17%	-3%	-4%	-4%

2-keto-valeric acid

Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0/0	1.2.4.1	Zymomonas mobilis	22.3

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	1	btcoa	1.00	21.02	20.93	13.22	21.03	21.01	23.18
	Max. Exp & Comp			0.00	0.00	0.00	0.00	0.00	0.00
	Max. GEM-Path			21.02	20.93	13.22	21.03	21.01	23.18

1,3-propanediol
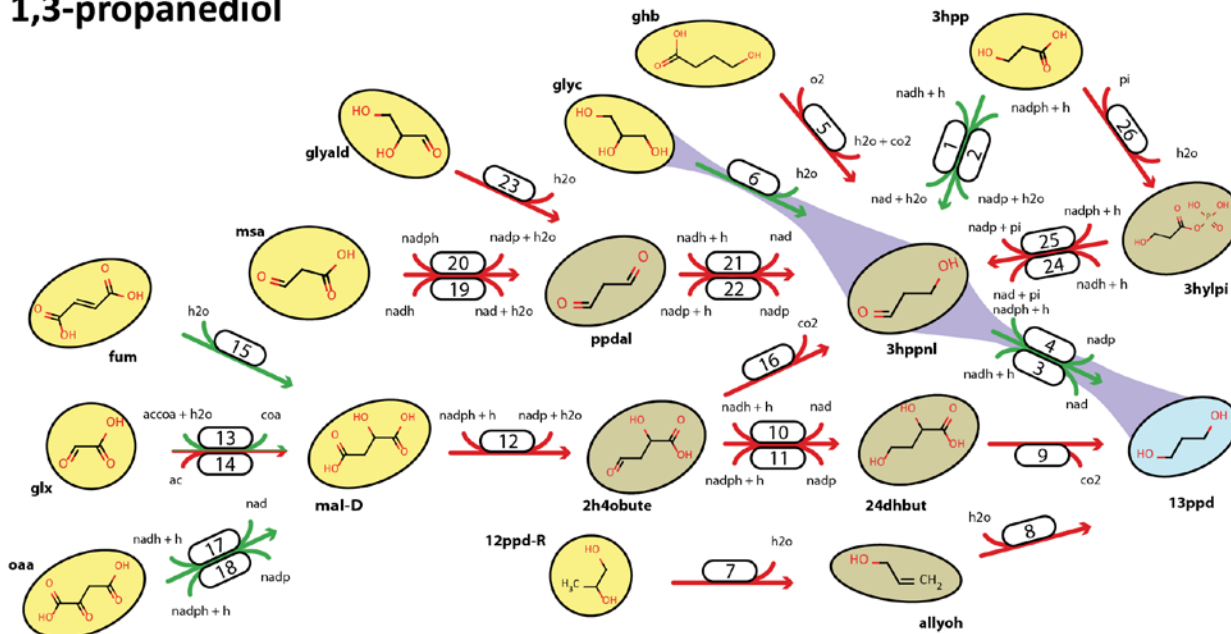
Predicted Reactions Specification

Reaction	Subs. TC/Enz. Prom. Coef	Enz. Prom. Coef	E.C. number	Specie	ΔG$_{rxn}$
1	0/0	0	1.2.1.3	Escherichia coli*	16.8
2	0/0	0	1.2.1.B6	Escherichia coli*	-3
3	0/0	0	1.1.1.202	Klebsiella pneumoniae	-53.75
4	0/0	0	1.1.1.202	Escherichia coli*	-73.6
5	0.5/0.91	0.91	1.13.12.4	Mycobacterium phlei	-361.8
6	0/0	0	4.2.1.28	Klebsiella pneumoniae	-28.3
7	0.09/0.88	0.88	4.2.1.28	Klebsiella pneumoniae	13.1
8	0.5/1.15	1.15	4.2.1.28	Klebsiella oxytoca	-5.5
9	0.12/0.76	0.76	4.1.1.73	Pseudomonas sp	-44.2
10	0.12/0.74	0.74	1.1.1.39	Brevundimonas diminuta	-53.8
11	0.29/1.11	1.11	1.1.1.10	Homo sapiens	-73.7
12	0.12/0.94	0.94	1.2.1.272	Enterococcus faecalis	-12.9
13	0/0	0	2.3.3.9	Pseudomonas putida	-32.9
14	0.36/1.14	1.14	4.1.2.18	Pseudomonas sp.	18.7
15	0/0	0	4.2.1.2	Escherichia coli*	2.7
16	0.12/0.69	0.69	4.1.1.73	Pseudomonas sp.	-44.2
17	0/0	0	1.1.1.28	Lactobacillus pentosus	-61.07
18	0/0	0	1.1.1.299	...othermobacter thermautotr...	-80.9
19	0.059/1	1	1.2.1.15	Pseudomonas aeruginosa	-30.3
20	0.059/0.698	0.698	1.2.1.16	Euglena gracilis	-50.2
21	0.125/0.83	0.83	1.1.1.1	Saccharomyces cerevisiae	-17.65
22	0.125/1.1035	1.1	1.1.1.91	Neurospora crassa	-37.5
23	0.3125/0.8	0.8	4.2.1.28	Klebsiella oxytoca	-82.5
24	0.14/0.76	0.76	1.2.1.12	Homo sapiens	-70.3
25	0.19/1.1	1.1	1.2.1.13	Mus musculus	-90.1
26	0.25/1.14	1.14	3.6.1.1	Streptomyces aureofaciens	-160.3

Predicted Pathways Specification

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	1, 3	3hpp	1.00	0.00	0.00	0.00	19.99	19.81	23.49
2	2, 3	3hpp	1.00	0.00	0.00	0.00	19.81	19.64	23.38
3	1, 4	3hpp	1.00	0.00	0.00	0.00	19.81	19.64	23.38
4	2, 4	3hpp	1.00	0.00	0.00	0.00	19.55	19.43	23.06
5	5, 3	ghb	0.50	0.00	0.00	0.00	21.91	21.76	24.60
6	5, 4	ghb	0.50	0.00	0.00	0.00	21.63	21.52	24.54
7[exp]	6, 3	glyc	1.00	18.97	16.57	31.50	26.87	26.68	33.53
8[exp]	6, 4	glyc	1.00	15.81	13.81	28.69	26.41	26.28	33.20
9	7, 8	12ppd-R	0.50	21.83	20.24	14.73	27.32	27.08	31.71
10	9, 10, 12, 13	glx	0.25	8.90	7.49	4.04	24.43	24.21	27.44
11	9, 11, 12, 13	glx	0.25	7.96	6.70	4.04	24.00	23.79	26.95
12	9, 10, 12, 14	glx	0.00	10.31	9.09	4.85	25.10	24.88	28.21
13	9, 11, 12, 14	glx	0.00	9.45	7.95	4.85	24.65	24.43	27.69
14	9, 10, 12, 15	fum	0.25	22.82	20.72	24.56	27.68	27.50	32.60
15	9, 11, 12, 15	fum	0.25	18.97	16.57	21.58	27.19	27.08	32.22
16	3, 16, 12, 17	oaa	0.50	22.82	20.72	24.56	27.68	27.50	32.60
17	3, 16, 12, 18	oaa	0.50	18.97	16.57	21.58	27.19	27.08	32.22
18	4, 16, 12, 17	oaa	0.50	18.97	16.57	21.58	27.19	27.08	32.22
19	4, 16, 12, 18	oaa	0.50	15.81	13.81	16.79	26.60	26.54	31.41
20	3, 16, 12, 13	glx	0.50	8.90	7.49	4.04	24.43	24.21	27.44
21	4, 16, 12, 13	glx	0.50	7.96	6.70	4.04	24.00	23.79	26.95
22	3, 16, 12, 14	glx	0.25	10.31	9.09	4.85	25.10	24.88	28.21
23	4, 16, 12, 14	glx	0.25	9.45	7.95	4.85	24.65	24.43	27.69
24	3, 16, 12, 15	fum	0.50	22.82	20.72	24.56	27.68	27.50	32.60
25	4, 16, 12, 15	fum	0.50	18.97	16.57	21.58	27.19	27.08	32.22
26	3, 21, 19	msa	0.33	0.00	0.00	0.00	19.99	19.81	23.49
27	3, 21, 20	msa	0.33	0.00	0.00	0.00	19.99	19.81	23.49
28	3, 22, 19	msa	0.33	0.00	0.00	0.00	19.99	19.81	23.49
29	3, 22, 20	msa	0.33	0.00	0.00	0.00	19.81	19.64	23.38
30	4, 21, 19	msa	0.33	0.00	0.00	0.00	19.99	19.81	23.49
31	4, 21, 20	msa	0.33	0.00	0.00	0.00	19.81	19.64	23.38
32	4, 22, 19	msa	0.33	0.00	0.00	0.00	19.81	19.64	23.38
33	4, 22, 20	msa	0.33	0.00	0.00	0.00	19.55	19.43	23.06
34	3, 21, 23	glyald	0.33	18.97	16.57	31.50	26.87	26.68	33.53
35	3, 22, 23	glyald	0.33	15.81	13.81	28.69	26.41	26.28	33.20
36	4, 21, 23	glyald	0.33	15.81	13.81	28.69	26.41	26.28	33.20
37	4, 22, 23	glyald	0.33	13.55	11.57	21.58	25.92	25.86	32.28
38	3, 24, 26	3hpp	0.33	0.00	0.00	0.00	19.99	19.81	23.49
39	3, 25, 26	3hpp	0.33	0.00	0.00	0.00	19.81	19.64	23.38
40	4, 24, 26	3hpp	0.33	0.00	0.00	0.00	19.81	19.64	23.38
41	4, 25, 26	3hpp	0.33	0.00	0.00	0.00	19.55	19.43	23.06
		Max. Exp & Comp		18.97	16.57	31.50	26.87	26.68	33.53
		Max. GEM-Path		22.82	20.72	31.50	27.68	27.50	33.53
		Exp. & Comp .v/s GEM-Path		20%	25%	0%	3%	3%	0%

1,4-butanediol
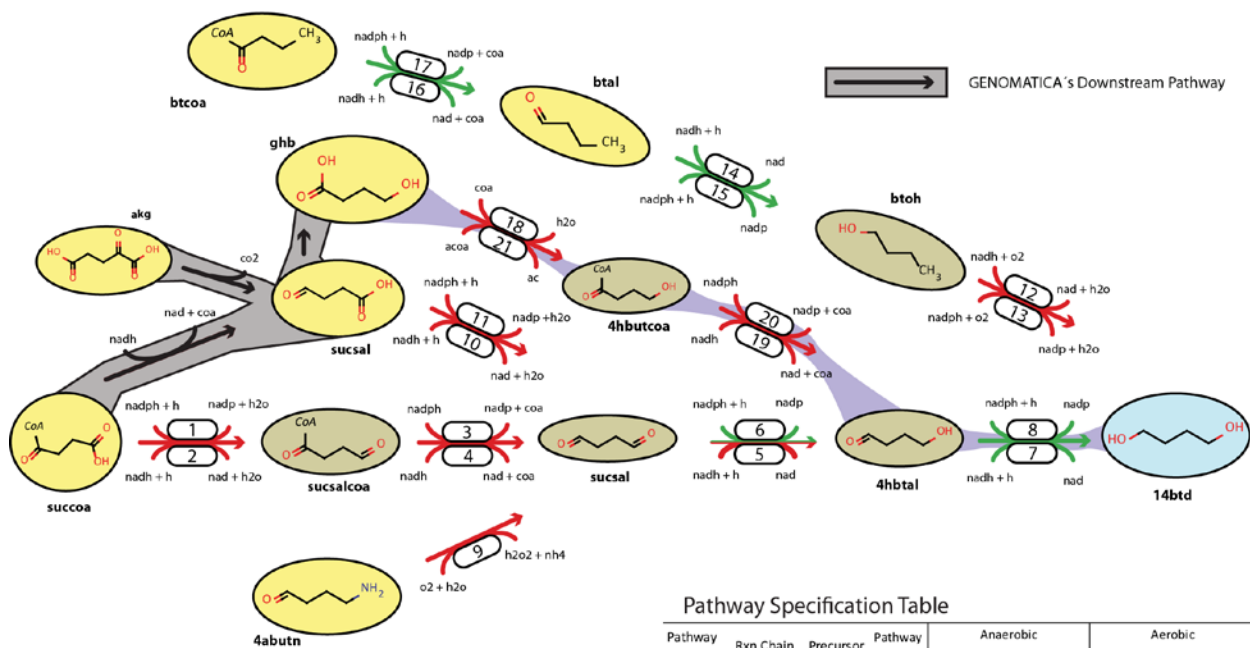
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0.026/083	1.2.1.76	Clostridium kluyveri	-3.6
2	0.026/0.95	1.2.1.10	Propionibacteriumsp.	16.2
3	0.069/1.18	1.2.1.10	Escherichia coli	-60.8
4	0.069/0.83	1.2.1.76	Clostridium kluyveri	-40.9
5	0.1/0.82	1.1.1.1	Saccharomyces cerevisiae	-53.7
6	0/0	1.1.1.71	Acinetobacter calcoaceticus	-73.6
7	0/0	1.1.1.165	Fusarium merismoides	-53.7
8	0/0	1.1.1.2	Saccharomyces cerevisiae	-73.6
9	0.4/0.86	1.4.3.21	Bos Taurus	-119
10	0.09/1.05	1.2.1.21	Escherichia coli	6.2
11	0.09/1.01	1.2.1.16	Escherichia coli	-13.6
12	0.88/1.12	1.14.12.11	Pseudomonas putida	-324.6
13	0.61/1.04	1.14.13.22	Acinetobacter sp.	-344.5
14	0/0	1.1.1.1	Saccharomyces cerevisiae	-52.7
15	0/0	1.1.1.1	Cucumis melo	-72.6
16	0/0	1.2.1.10	Escherichia coli*	-45.9
17	0/0	1.2.1.76	Clostridium acetobutylicum	-65.7
18	0.09/1.09	3.1.2.2	Sus scrofa	48
19	0.005/1.18	1.2.1.10	Escherichia coli	-42
20	0.005/0.83	1.2.1.76	Clostridium kluyveri	-61.9
21	0.15/1.103	2.8.3.1	Clostridium propionicum	-3.4

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	10, 5, 7	sucsal	0.33	19.21	16.81	9.38	20.60	20.42	23.09
2	11, 5, 7	sucsal	0.33	18.91	15.91	9.38	20.47	20.32	23.09
3	10, 6, 7	sucsal	0.67	18.91	15.91	9.38	20.47	20.32	23.09
4	11, 6, 7	sucsal	0.67	15.13	12.73	9.38	20.16	20.00	22.86
5	10, 5, 8	sucsal	0.33	18.91	15.91	9.38	20.47	20.32	23.09
6	11, 4, 8	sucsal	0.33	15.13	12.73	9.38	20.16	20.00	22.86
7	10, 6, 8	sucsal	0.67	15.13	12.73	9.38	20.16	20.00	22.86
8	11, 6, 8	sucsal	0.67	12.61	10.61	9.38	19.85	19.68	22.52
9	9, 5, 7	4abutn	0.33	8.90	7.49	4.93	19.07	18.85	21.71
10	9, 6, 7	4abutn	0.67	7.96	6.70	4.93	18.72	18.51	21.40
11	9, 5, 8	4abutn	0.33	7.96	6.70	4.93	18.72	18.51	21.40
12	9, 6, 8	4abutn	0.67	7.21	6.06	4.93	18.38	18.17	21.11
13	2, 3, 5, 7	succoa	0.25	21.04	19.92	23.46	21.04	20.83	24.45
14	1, 3, 5, 7	succoa	0.50	20.99	19.53	23.46	21.03	20.78	24.45
15	2, 4, 5, 7	succoa	0.25	20.99	19.53	23.46	21.03	20.78	24.45
16	1, 4, 5, 7	succoa	0.50	19.47	18.27	23.46	20.75	20.54	24.45
17	2, 3, 6, 7	succoa	0.50	20.99	19.53	23.46	21.03	20.78	24.45
18	1, 3, 6, 7	succoa	0.75	19.47	18.27	23.46	20.75	20.54	24.45
19	2, 4, 6, 7	succoa	0.50	19.47	18.27	23.46	20.75	20.54	24.45
20	1, 4, 6, 7	succoa	0.75	18.15	17.16	21.97	20.47	20.31	24.16
21	2, 3, 5, 8	succoa	0.25	20.99	19.53	23.46	21.03	20.78	24.45
22	1, 3, 5, 8	succoa	0.50	19.47	18.27	23.46	20.75	20.54	24.45
23	2, 4, 5, 8	succoa	0.25	19.47	18.27	23.46	20.75	20.54	24.45
24	1, 4, 5, 8	succoa	0.50	18.15	17.16	21.97	20.47	20.31	24.16
25	2, 3, 6, 8	succoa	0.50	19.47	18.27	23.46	20.75	20.54	24.45
26	1, 3, 6, 8	succoa	0.75	18.15	17.16	21.97	20.47	20.31	24.16
27	2, 4, 6, 8	succoa	0.50	18.15	17.16	21.97	20.47	20.31	24.16
28	1, 4, 6, 8	succoa	0.75	16.61	15.08	19.89	20.20	20.08	23.83
29	16, 14, 12	btcoa	0.67	0.00	0.00	0.00	17.80	17.80	20.88
30	16, 14, 13	btcoa	0.67	0.00	0.00	0.00	17.80	17.80	20.88
31	16, 15, 12	btcoa	0.67	0.00	0.00	0.00	17.80	17.80	20.88
32	16, 15, 13	btcoa	0.67	0.00	0.00	0.00	17.80	17.80	20.88
33	17, 14, 12	btcoa	0.67	0.00	0.00	0.00	17.80	17.80	20.88
34	17, 14, 13	btcoa	0.67	0.00	0.00	0.00	17.80	17.80	20.88
35	17, 15, 12	btcoa	0.67	0.00	0.00	0.00	17.80	17.80	20.88
36	17, 15, 13	btcoa	0.67	0.00	0.00	0.00	17.80	17.69	20.67
37	18, 19, 7	ghb	0.33	19.21	16.81	9.38	20.60	20.42	23.09
38	18, 20, 7	ghb	0.33	18.91	15.91	9.38	20.47	20.32	23.09
39	18, 19, 8	ghb	0.33	18.91	15.91	9.38	20.47	20.32	23.09
40	18, 20, 8	ghb	0.33	15.13	12.73	9.38	20.16	20.00	22.86
41 exp/comp	21, 19, 7	ghb	0.33	14.78	12.85	6.76	20.08	19.90	22.50
42 exp/comp	21, 20, 7	ghb	0.33	13.76	11.57	6.76	20.00	19.84	22.50
43 exp/comp	21, 19, 8	ghb	0.33	13.76	11.57	6.76	20.00	19.84	22.50
44 exp/comp	21, 20, 8	ghb	0.33	11.64	9.79	6.76	19.70	19.52	22.36
		Max. Exp & Comp		14.78	12.85	6.76	20.08	19.90	22.50
		Max. GEM-Path		21.04	19.92	23.46	21.04	20.83	24.45
		Exp. & Comp .v/s GEM-Path		42%	55%	247%	5%	5%	9%

129
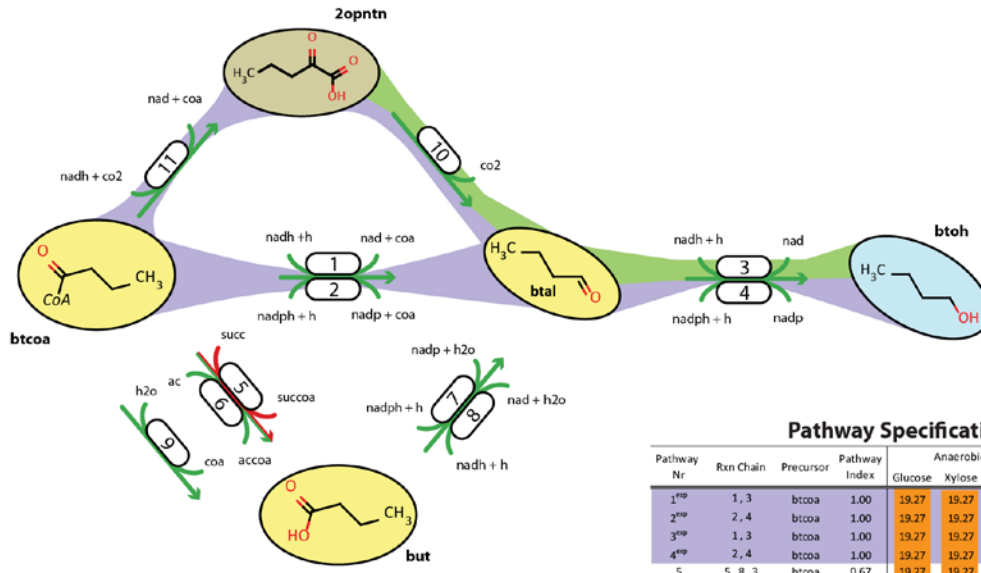
1-butanol

Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0/0	1.2.1.10	Escherichia coli*	-45.8
2	0/0	1.2.1.76	Clostridium acetobutylicum	-65.7
3	0/0	1.1.1.1	Saccharomyces cerevisiae	-52.7
4	0/0	1.1.1.1	Cucumis melo	-72.6
5	0.001/0.95	2.8.3.5	Homo sapiens	-9.7
6	0/0	2.8.3.3	Escherichia col^	-2.6
7	0/0	1.2.1.8	Escherichia coli*	-11.5
8	0/0	1.2.1.19	Escherichia coli^	8.4
9	0/0	3.1.2.20	Rattus norvegicus	-54.2
10	0/0	4.1.1.1	Lactococcus lactis	-47.8
11	0/0	1.2.4.1	Zymomonas mobilis	22.3

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1[exp]	1 , 3	btcoa	1.00	19.27	19.27	22.60	19.28	19.28	22.62
2[exp]	2 , 4	btcoa	1.00	19.27	19.27	22.27	19.28	19.28	22.52
3[exp]	1 , 3	btcoa	1.00	19.27	19.27	22.27	19.28	19.28	22.52
4[exp]	2 , 4	btcoa	1.00	19.27	19.27	21.15	19.28	19.28	22.19
5	5 , 8 , 3	btcoa	0.67	19.27	19.27	22.60	19.28	19.28	22.62
6	6 , 8 , 3	btcoa	1.00	19.27	19.27	22.60	19.28	19.28	22.62
7	5 , 7 , 3	btcoa	0.67	19.27	19.27	22.60	19.28	19.28	22.62
8	6 , 7 , 3	btcoa	1.00	19.27	19.27	22.60	19.28	19.28	22.62
9	5 , 8 , 4	btcoa	0.67	19.27	19.27	22.60	19.28	19.28	22.62
10	6 , 8 , 4	btcoa	1.00	19.27	19.27	22.60	19.28	19.28	22.62
11	5 , 7 , 4	btcoa	0.67	19.27	19.27	22.60	19.28	19.28	22.62
12	6 , 7 , 4	btcoa	1.00	19.27	19.27	22.60	19.28	19.28	22.62
13	9 , 8 , 3	btcoa	1.00	19.27	19.27	22.60	19.28	19.28	22.62
14	9 , 7 , 3	btcoa	1.00	19.27	19.27	22.27	19.28	19.28	22.52
15	9 , 8 , 4	btcoa	1.00	19.27	19.27	22.27	19.28	19.28	22.52
16	9 , 7 , 4	btcoa	1.00	19.27	19.27	21.15	19.28	19.28	22.19
17[exp/comp]	11 , 10 , 3	btcoa	1.00	19.27	19.27	21.93	19.28	19.28	22.50
18[exp/comp]	11 , 10 , 4	btcoa	1.00	19.27	19.27	21.93	19.28	19.28	22.50
	Max. Exp & Comp			19.27	19.27	22.60	19.28	19.28	22.62
	Max. GEM-Path			19.27	19.27	22.60	19.28	19.28	22.62
	Exp. & Comp .v/s GEM-Path			0%	0%	0%	0%	0%	0%

4-hydroxybutyrate: (native *e. coli*)
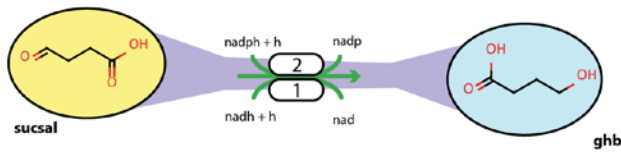
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0/0	1.1.1.21	Homo sapiens	-53.7
2	0/0	1.1.1.19	Mesocricetus auratus	-73.6

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1[exp]	1	sucsal	1.00	13.40	11.68	5.49	23.26	23.18	26.28
2[exp]	2	sucsal	1.00	11.87	10.37	5.49	22.88	22.79	26.00
	Max. Exp & Comp			13.40	11.68	5.49	23.26	23.18	26.28
	Max. GEM-Path			0.00	0.00	0.00	0.00	0.00	0.00
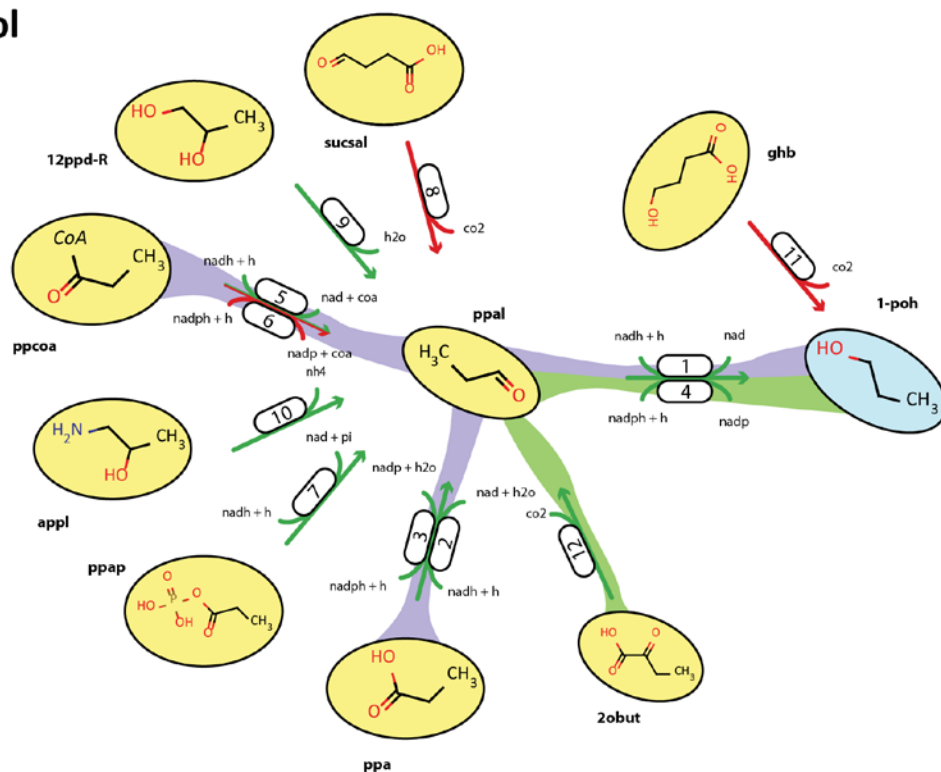
130

1-propanol

Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	Enz. Prom. Coef	E.C. number	Specie	ΔG$_{rxn}$
1	0/0	0	1.1.1.1	Saccharomyces erevisiae	-53.6
2	0/0	0	1.2.1.3	Escherichia coli*	6.2
3	0/0	0	1.2.1.22	Escherichia coli^	-13.7
4	0/0	0	1.1.1.91	Neurospora crassa	-73.5
5	0/0	0	1.2.1.10	Escherichia coli*	-30.1
6	0.005/0.83	0.83	1.2.1.76	Clostridium kluyveri	-49.9
7	0/0	0	1.2.1.12	Homo sapiens	-70.2
8	0.1/0.92	0.92	4.1.1.54	Rattus norvegicus	-52.1
9	0/0	0	4.2.1.28	Klebsiella sp	-40.1
10	0/0	0	4.3.1.7	Salmonella enterica	-84.2
11	0.09/0.92	0.92	4.1.1.54	Rattus norvegicus	-51.9
12	0/0	0	4.1.1.1	Zymomonas mobilis	-47.9

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1[exp]	2 , 1	ppa	1.00	25.71	25.71	30.16	25.71	25.71	30.16
2[exp]	3 , 1	ppa	1.00	25.71	25.71	30.16	25.71	25.71	30.16
3[exp]	2 , 4	ppa	1.00	25.71	25.71	30.16	25.71	25.71	30.16
4[exp]	3 , 4	ppa	0.50	25.71	25.71	30.16	25.71	25.71	30.16
5[exp]	5 , 1	ppcoa	1.00	25.70	23.82	28.06	25.71	25.36	29.77
6[exp]	6 , 1	ppcoa	0.50	23.96	22.65	28.06	25.39	25.17	29.77
7[exp]	5 , 4	ppcoa	1.00	23.96	22.65	28.06	25.39	25.17	29.77
8[exp]	6 , 4	ppcoa	0.50	22.00	20.74	26.99	24.97	24.82	29.56
9	7 , 1	ppap	1.00	25.70	23.82	28.06	25.71	25.36	29.77
10	7 , 4	ppap	1.00	23.96	22.65	28.06	25.39	25.17	29.77
11	8 , 1	sucsal	0.50	13.56	11.68	6.32	23.26	23.18	26.28
12	8 , 4	sucsal	0.50	11.87	10.37	6.32	22.88	22.79	26.00
13	9 , 1	12ppd-R	1.00	21.33	19.75	21.73	24.48	24.35	28.67
14	9 , 4	12ppd-R	1.00	19.62	18.26	21.73	24.38	24.15	28.67
15	10 , 1	appl	1.00	14.79	12.74	10.53	23.56	23.35	27.47
16	10 , 4	appl	1.00	12.61	10.61	10.22	23.07	22.86	26.90
17	11	ghb	0.00	13.56	11.68	6.32	23.26	23.18	26.28
18[comp]	12 , 4	2obut	1.00	9.45	7.95	9.21	21.88	21.80	25.82
19[comp]	12 , 1	2obut	1.00	10.54	9.09	9.21	22.20	22.07	26.23
		Max. Exp & Comp		25.71	25.71	30.16	25.71	25.71	30.16
		Max. GEM-Path		25.70	23.82	28.06	25.71	25.36	29.77
		Exp. & Comp .v/s GEM-Path		0%	-7%	-7%	0%	-1%	-1%

131

3-hydroxypropanoate (native *e. coli*)
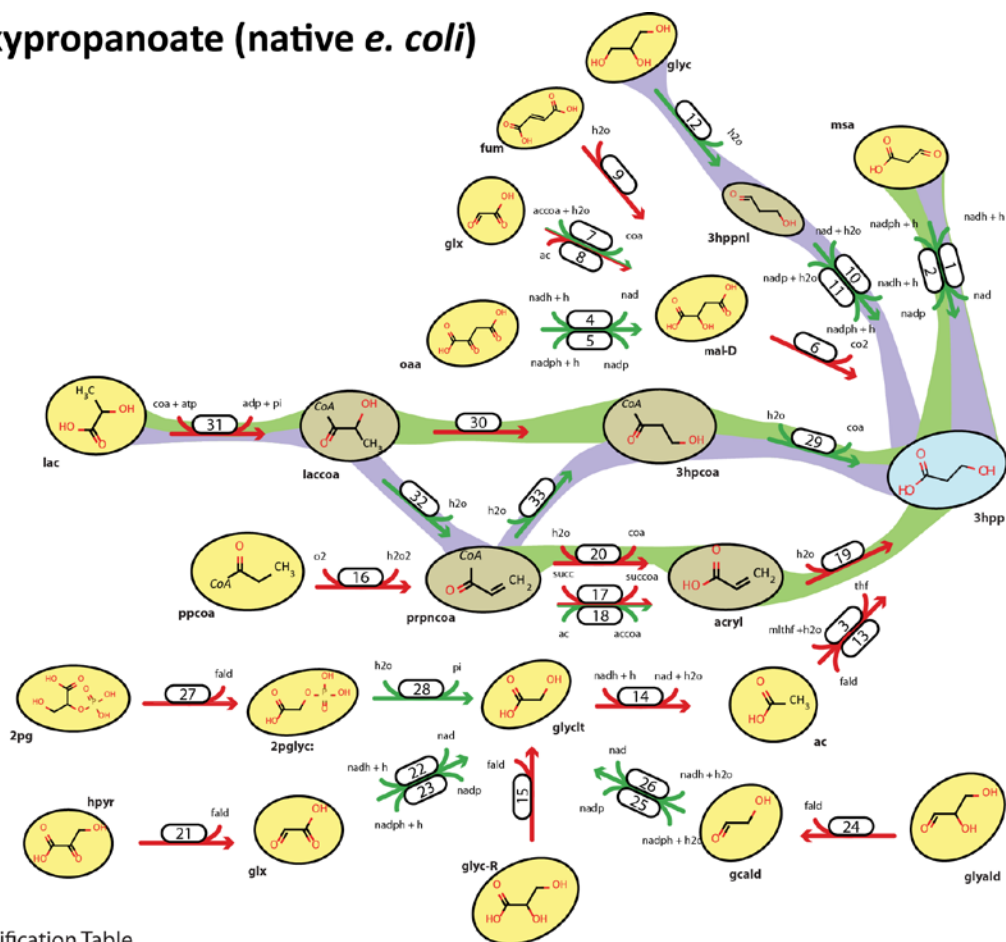
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	Δ G$_{rxn}$
1	0/0	1.1.1.59	Escherichia coli*	-54
2	0/0	1.1.1.276	Escherichia coli	-74.1
3	0.4/1.02	2.1.2.1	Geobacillus sp.	14
4	0/0	1.1.1.28	Lactobacillus delbrueckii	-61.1
5	0/0	1.1.1.299	Methanothermobacter sp.	-80.9
6	~0/0.7	4.1.1.73	Pseudomonas sp.	-43.5
7	0/0	2.3.3.9	Escherichia coli*	-32.9
8	0.36/1.14	4.1.2.18	Sinorhizobium meliloti	18.8
9	~0/0.87	4.2.1.2	Sus crofa	2.7
10	0/0	1.2.1.3	Escherichia coli*	-6.2
11	0/0	1.2.1.B6	Escherichia coli*	13.6
12	0/0	4.2.1.28	Klebsiella pneumoniae	-28.3
13	0.22/0.86	4.1.2.36	Rattus novergicus	14.2
14	0.25/1.08	1.17.1.4	Triticum aestivum	-91.6
15	0.1/1.08	4.1.2.36	Rattus novergicus	-26.4
16	0.0046/0.24	1.3.3.6	Mus musculus	-18.6
17	0.013/0.918	2.8.3.8	Acetobacter aceti	1.75
18	0/0	2.8.3.12	Acidaminococcus sp.	8.8
19	0.39/1.03	4.2.1.2	Sus crofa	-2.1
20	0.0091/0.8	3.1.2.2	Sus scrofa	-42.8
21	0.42/1.02	4.1.2.12	Escherichia coli	-9.7
22	0/0	1.1.1.26	Cyanophora paradoxa	-79
23	0/0	1.1.1.215	Acetobacter pasteurianus	-98.9
24	0.25/0.98	4.1.2.36	Rattus norvegicus	-6.5
25	0/0	1.2.1.21	Salmonella enterica	-2.6
26	0/0	1.2.1.21	Escherichia coli**	-22.5
27	0.5/1.09	4.1.2.17	Escherichia coli	-23.2
28	0/0	3.1.3.18	Escherichia coli**	-195
29	0/0	3.1.2.4	Homo sapiens	-46.3
30	0.015/1.1	5.4.99.13	Streptomyces sp.	7.6
31	0.32/1.09	6.2.1.13	Giardia intestinalis	-5
32	0/0	4.2.1.4	Clostridium propionicum	6.3
33	0/0	4.2.1.116	Metallosphaera sedula	1.3

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic Glucose	Anaerobic Xylose	Anaerobic Glycerol	Aerobic Glucose	Aerobic Xylose	Aerobic Glycerol
1[exp/comp]	1	msa	1.00	0.00	0.00	0.00	24.06	23.83	28.14
2[exp/comp]	2	msa	1.00	0.00	0.00	0.00	23.69	23.52	28.02
3	3	ac	0.00	37.82	31.82	21.84	38.51	38.06	38.64
4	6, 4	oaa	0.50	21.30	18.18	9.38	35.89	35.72	38.64
5	6, 5	oaa	0.50	16.81	14.14	9.38	34.86	34.78	38.64
6	6, 7	glx	0.50	7.26	6.58	3.09	28.45	28.25	32.11
7	6, 8	glx	0.00	7.93	7.19	3.55	29.36	29.16	33.17
8	9, 6	fum	0.00	21.30	18.18	9.38	35.89	35.72	38.64
9[exp]	10, 12	glyc	1.00	13.75	11.57	12.64	33.70	33.62	38.64
10[exp]	11, 12	glyc	1.00	16.81	14.14	12.64	34.49	34.31	38.64
11	13, 14, 15	glyc-R	0.00	6.27	5.67	2.67	28.75	28.54	32.45
12	19, 17, 16	ppcoa	0.00	38.57	38.57	19.43	38.57	38.57	38.64
13	19, 18, 16	ppcoa	0.33	38.57	38.57	19.43	38.57	38.57	38.64
14[comp]	19, 20, 16	ppcoa	0.00	24.37	21.23	10.79	36.61	36.46	38.64
15	13, 14, 22, 21	hpyr	0.25	8.18	7.42	3.73	29.67	29.48	33.53
16	13, 14, 23, 21	hpyr	0.25	8.18	7.42	3.73	29.67	29.48	33.53
17	13, 14, 26, 24	glyald	0.25	13.75	11.57	12.64	33.70	33.62	38.64
18	13, 14, 25, 24	glyald	0.25	16.81	14.14	12.64	34.62	34.39	38.64
19	13, 14, 28, 27	2pg	0.25	24.35	21.21	10.77	36.66	36.49	38.64
20[comp]	29, 30, 31	lac-L	0.33	10.09	8.49	5.29	31.47	31.36	37.37
21[exp]	29, 33, 32, 31	lac-L	0.75	10.09	8.49	5.29	31.47	31.36	37.37
22	29, 33, 16	ppcoa	0.33	24.37	21.23	10.79	36.61	36.46	38.64
23[comp]	19, 20, 32, 31	ppcoa	0.25	12.61	10.61	6.76	33.15	33.07	38.64
24	19, 17, 32, 31	ppcoa	0.25	12.61	10.61	6.76	33.15	33.07	38.64
25	19, 18, 32, 31	ppcoa	0.50	10.09	8.49	5.29	31.47	31.36	37.37
Max. Exp & Comp				24.37	21.23	12.64	36.61	36.46	38.64
Max. GEM-Path				38.57	38.57	21.84	38.57	38.57	38.64
Exp. & Comp .v/s GEM-Path				58%	82%	73%	5%	6%	0%

2-keto-isovaleric acid (native *e. coli*)
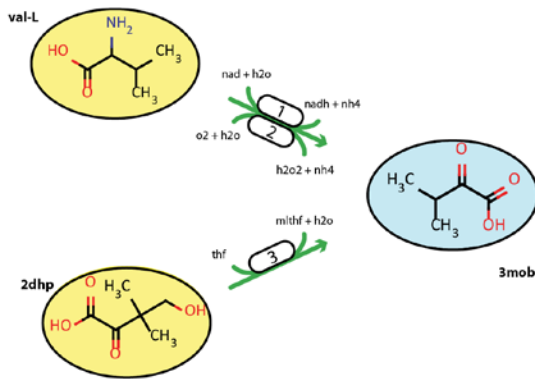
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	Δ G$_{rxn}$
1	0/0	1.4.1.1	Bacillus subtilis	-5.3
2	0/0	1.4.3.19	Rhodococcus opacus	-95.1
3	0/0	2.1.2.11	Escherichia oli^	5.6

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	1	val-L	1.00	17.26	15.08	10.19	20.53	20.36	23.16
2	2	val-L	1.00	15.82	13.82	9.54	20.38	20.21	22.99
3	3	2dhp	1.00	18.98	16.58	10.95	20.68	20.51	23.18
		Max. Exp & Comp		0.00	0.00	0.00	0.00	0.00	0.00
		Max. GEM-Path		18.98	16.58	10.95	20.68	20.51	23.18

2-phenylethanol
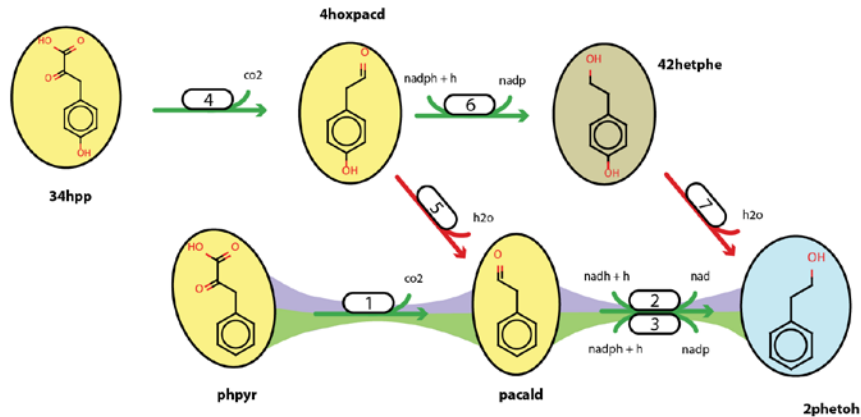
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	Δ G$_{rxn}$
1	0/0	4.1.1.1	Zygosaccharomyces bisporus	-53.8
2	0/0	1.1.1.1	Thermus sp.	-53.4
3	0/0	1.1.1.191	Phycomyces blakesleeanus	-73.2
4	0/0	4.1.1.72	Lactococcus lactis	-47.7
5	0.24/0.98	4.2.1.84	Nocardia sp.	4.2
6	0/0	1.1.1.2	Rattus norvegicus	-73.7
7	0.39/0.89	4.2.1.87	Pseudomonas aeruginosa	-47.7

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1$^{exp\comp}$	1, 2	phe-L	1.00	5.47	4.60	2.95	10.55	10.51	12.15
2$^{exp\comp}$	1, 3	phe-L	1.00	5.10	4.29	2.95	10.48	10.45	12.14
3	2, 5, 6	34hpp	0.67	5.10	4.29	2.65	10.99	10.96	12.51
4	3, 5, 6	34hpp	0.67	4.78	4.02	2.65	10.87	10.85	12.44
5	8, 10, 11	34hpp	0.67	4.78	4.02	2.65	10.87	10.85	12.44
		Max. Exp & Comp		5.47	4.60	2.95	10.55	10.51	12.15
		Max. GEM-Path		5.10	4.29	2.65	10.99	10.96	12.51
		Exp. & Comp .v/s GEM-Path		-7%	-7%	-10%	4%	4%	3%

2-methyl-1-butanol
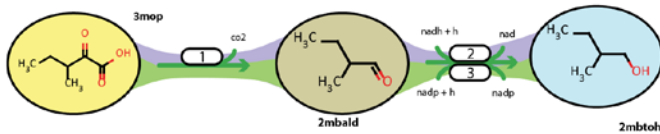
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	Δ G$_{rxn}$
1	0/0	4.1.1.1	Zymobacter palmae	-61
2	0/0	1.1.1.1	Saccharomyces cerevisiae	-52.9
3	0/0	1.1.1.2	Synechocystis sp	-72.7

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1$^{exp\comp}$	1, 2	3mop	1.00	8.62	7.49	8.65	14.34	14.19	16.84
2$^{exp\comp}$	1, 3	3mop	1.00	7.91	6.70	7.95	14.14	14.01	16.60
		Max. Exp & Comp		8.62	7.49	8.65	14.34	14.19	16.84
		Max. GEM-Path		0.00	0.00	0.00	0.00	0.00	0.00

3-hydroxybutyrate
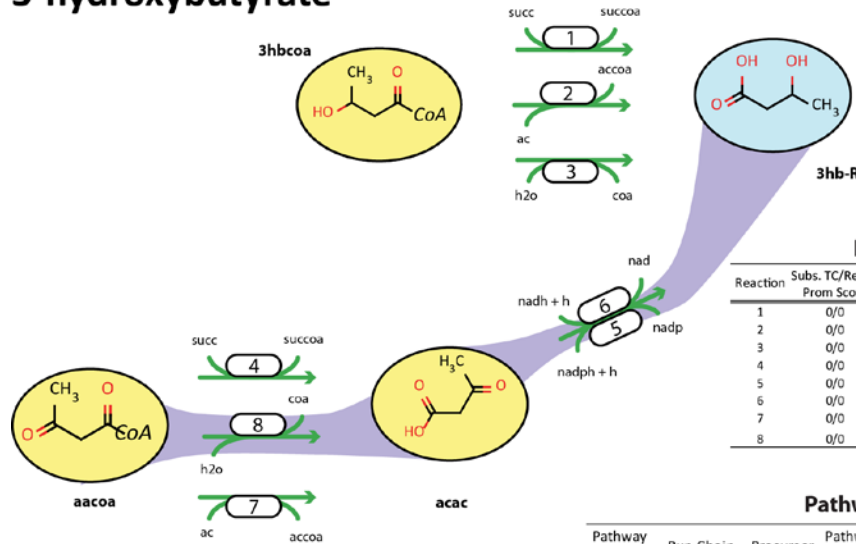
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0/0	2.8.3.5	Homo sapiens	-3.8
2	0/0	2.8.3.9	Escherichia coli*	3.3
3	0/0	3.1.2.2	Sus crofa	-48.4
4	0/0	2.8.3.5	Bacillus subtilis	-14.4
5	0/0	1.1.1.276	Escherichia coli	-66.8
6	0/0	1.1.1.30	Pseudomonas putida	-46.9
7	0/0	2.8.3.9	Escherichia coli	-0.4
8	0/0	3.1.2.11	Homo sapiens	3.9

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	1	3hbcoa	1.00	21.97	20.40	13.10	24.87	24.65	27.95
2	2	3hbcoa	1.00	21.97	20.40	13.10	24.87	24.65	27.95
3	3	3hbcoa	1.00	18.08	16.40	8.81	24.21	24.00	27.19
4	4, 6	aacoa	1.00	21.97	20.40	13.10	24.87	24.65	27.95
5	7, 6	aacoa	1.00	21.97	20.40	13.10	24.87	24.65	27.95
6	4, 5	aacoa	1.00	19.43	17.64	11.13	24.43	24.21	27.44
7	7, 5	aacoa	1.00	19.43	17.64	11.13	24.43	24.21	27.44
8[exp]	8, 5	aacoa	1.00	18.08	16.40	8.81	24.21	24.00	27.19
9[exp]	8, 6	aacoa	1.00	15.86	14.38	8.81	23.79	23.58	26.71
	Max. Exp & Comp			18.08	16.40	8.81	24.21	24.00	27.19
	Max. GEM-Path			21.97	20.40	13.10	24.87	24.65	27.95
	Exp. & Comp .v/s GEM-Path			22%	24%	49%	3%	3%	3%

Isopropanol
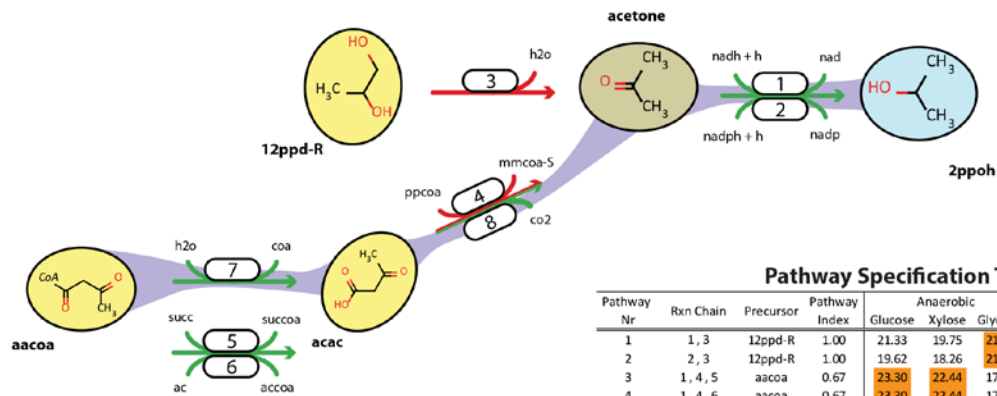
Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	1, 3	12ppd-R	1.00	21.33	19.75	21.73	24.48	24.35	28.67
2	2, 3	12ppd-R	1.00	19.62	18.26	21.73	24.38	24.15	28.67
3	1, 4, 5	aacoa	0.67	23.30	22.44	17.39	25.10	24.88	28.21
4	1, 4, 6	aacoa	0.67	23.30	22.44	17.39	25.10	24.88	28.21
5	2, 4, 5	aacoa	0.67	21.68	20.64	14.08	24.65	24.43	27.69
6	2, 4, 6	aacoa	0.67	21.68	20.64	14.08	24.65	24.43	27.69
7	1, 4, 7	aacoa	0.67	20.78	19.78	12.85	24.43	24.21	27.44
8	2, 4, 7	aacoa	0.67	18.98	16.58	10.95	24.00	23.79	26.95
9	1, 8, 5	aacoa	1.00	22.50	21.57	15.56	24.87	24.65	27.95
10	1, 8, 6	aacoa	1.00	22.50	21.57	15.56	24.87	24.65	27.95
11	2, 8, 5	aacoa	1.00	20.78	19.78	12.85	24.43	24.21	27.44
12	2, 8, 6	aacoa	1.00	20.78	19.78	12.85	24.43	24.21	27.44
13[exp]	1, 8, 7	aacoa	1.00	19.94	18.43	11.19	24.21	24.00	27.19
14[exp]	2, 8, 7	aacoa	1.00	17.26	15.08	10.19	23.79	23.58	26.71
	Max. Exp & Comp			19.94	18.43	11.19	24.21	24.00	27.19
	Max. GEM-Path			23.30	22.44	21.73	25.10	24.88	28.67
	Exp. & Comp .v/s GEM-Path			17%	22%	94%	4%	4%	5%

Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0/0	1.1.1.1	Saccharomyces cerevisiae	-47.6
2	0/0	1.1.1.2	Lactobacillus brevis	-67.5
3	0.09 / 0.88	4.2.1.28	Klebsiella pneumoniae	-53.4
4	0.3/1.12	2.1.3.1	Canis lupus	-2.11
5	0/0	2.8.3.5	Bacillus subtilis	-14.4
6	0/0	2.8.3.9	Escherichia coli^	-0.4
7	0/0	3.1.2.11	Homo sapiens	3.9
8	0/0	4.1.1.4	Pseudomonas putida	-50.9

3-hydroxyvalerate
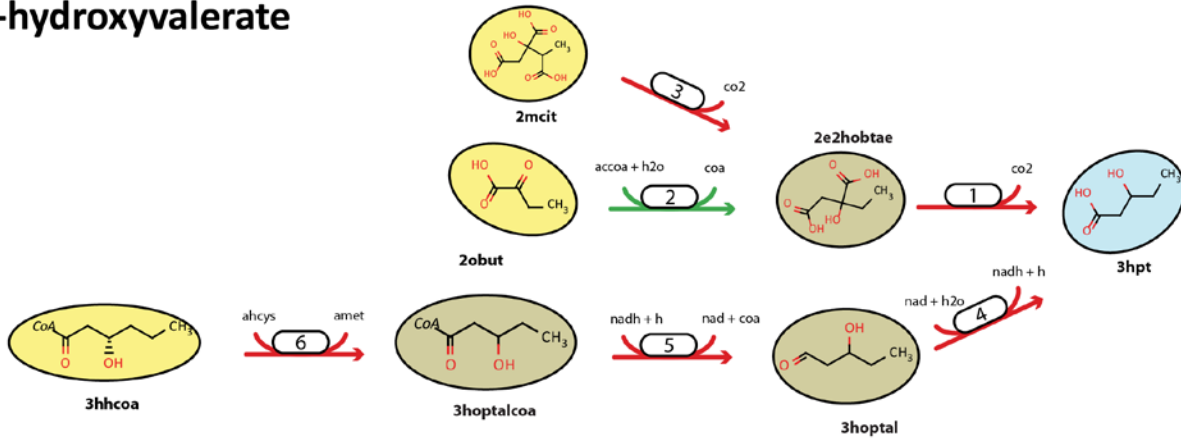
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	Enz. Prom. Coef	E.C. number	Specie	Δ G$_{rxn}$
1	0.1/1.1	1.1	4.1.1.6	Aspergillus terreus	-59.2
2	0/0	0	2.3.3.13	Saccharomyces cerevisiae	-16.9
3	0.06/1.18	1.18	4.1.1.6	Aspergillus terreus	-64.9
4	0.3/0.87	0.87	1.2.1.20	Pseudomonas putida	10.8
5	0.007/0.94	0.94	1.2.1.10	Clostridium kluyveri	-44.4
6	0.13/0.93	0.93	2.1.1.104	Petroselinum crispus	-49.5

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	1 , 2	2obut	0.50	7.56	6.36	4.93	17.13	16.95	19.99
2	1 , 3	2mcit	0.00	10.66	9.10	5.30	17.95	17.86	20.44
3	4 , 5 , 6	3hhcoa	0.00	1.85	1.56	1.20	4.25	4.21	5.00
		Max. Exp & Comp		0.00	0.00	0.00	0.00	0.00	0.00
		Max. GEM-Path		10.66	9.10	5.30	17.95	17.86	20.44

2-keto-butanoic acid (native *e. coli*)
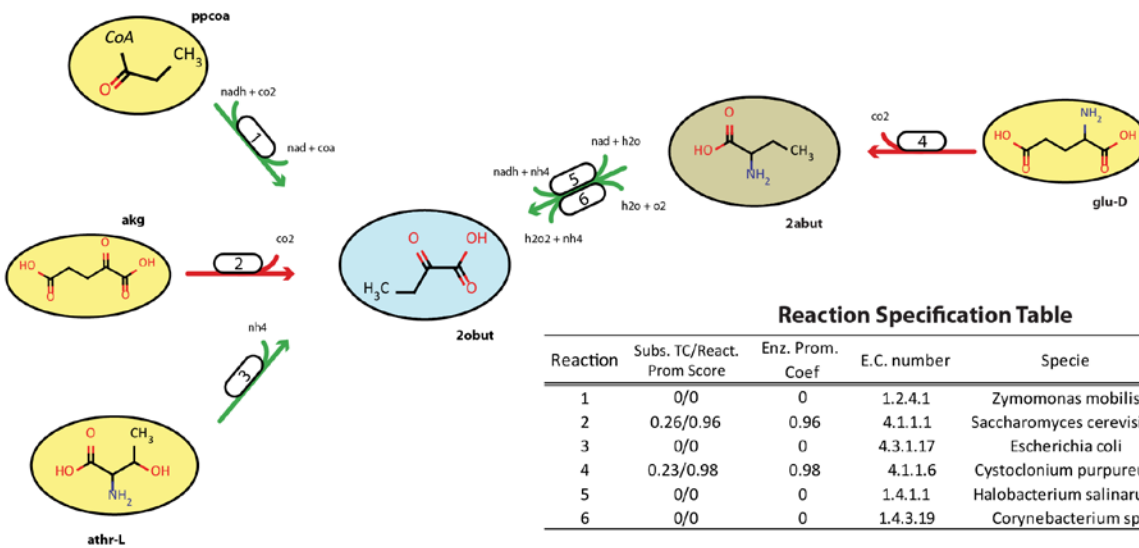
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	Enz. Prom. Coef	E.C. number	Specie	Δ G$_{rxn}$
1	0/0	0	1.2.4.1	Zymomonas mobilis	-214
2	0.26/0.96	0.96	4.1.1.1	Saccharomyces cerevisiae	-48.1
3	0/0	0	4.3.1.17	Escherichia coli	-80.8
4	0.23/0.98	0.98	4.1.1.6	Cystoclonium purpureum	-51.4
5	0/0	0	1.4.1.1	Halobacterium salinarum	-26.7
6	0/0	0	1.4.3.19	Corynebacterium sp	-116

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	1	ppcoa	1.00	25.10	23.36	18.21	28.16	27.78	28.98
2	2	akg	1.00	10.32	9.18	4.28	24.50	24.43	27.65
3	3	athr-L	1.00	9.99	8.49	6.61	24.61	24.40	28.71
4	4 , 5	glu-D	1.00	9.92	8.73	4.54	24.28	24.21	27.44
5	4 , 6	glu-D	1.00	9.49	8.29	4.28	24.07	24.00	27.22
		Max. Exp & Comp		0.00	0.00	0.00	0.00	0.00	0.00
		Max. GEM-Path		25.10	23.36	18.21	28.16	27.78	28.98

3-methyl-1-butanol
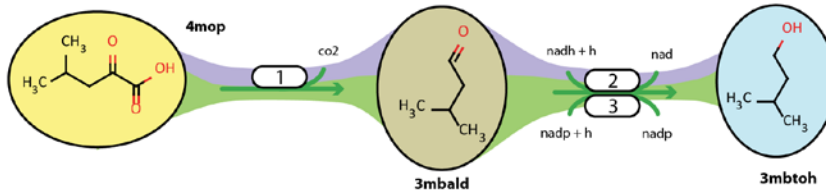
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0/0	4.1.1.1	Sacharomyces cerevisiae	-47.5
2	0/0	1.1.1.1	Saccharomyces cerevisiae	-53.3
3	0/0	1.1.1.2	Sacharomyces cerevisiae	-73.2

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1[exp/comp]	1, 2	4mop	1.00	11.87	10.37	7.58	14.79	14.66	16.70
2[exp/comp]	1, 3	4mop	1.00	10.55	9.21	6.87	14.63	14.50	16.53
		Max. Exp & Comp		11.87	10.37	7.58	14.79	14.66	16.70
		Max. GEM-Path		0.00	0.00	0.00	0.00	0.00	0.00

2,3-propanediol
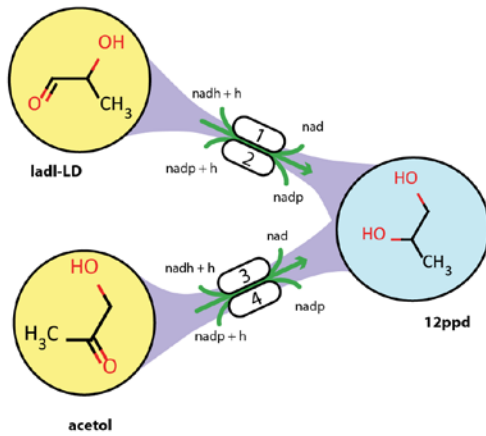
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0/0	1.1.1.77	Escherichia coli	-54
2	0/0	1.1.1.2	Sus scrofa	-73.9
3	0/0	1.1.1.1	Methylobacterium extorquens	-47.1
4	0/0	1.1.1.156	Schizosaccharomyces pombe	-67

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1[exp]	1	ladl-LD	1.00	21.83	20.24	14.73	27.32	27.08	31.71
2[exp]	2	ladl-LD	1.00	18.97	16.57	14.73	26.87	26.68	31.71
3[exp]	3	acetol	1.00	18.97	16.57	14.73	26.87	26.68	31.71
4[exp]	4	acetol	1.00	15.81	13.81	14.73	26.41	26.29	31.28
		Max. Exp & Comp		21.83	20.24	14.73	27.32	27.08	31.71
		Max. GEM-Path		0.00	0.00	0.00	0.00	0.00	0.00

acrylamide
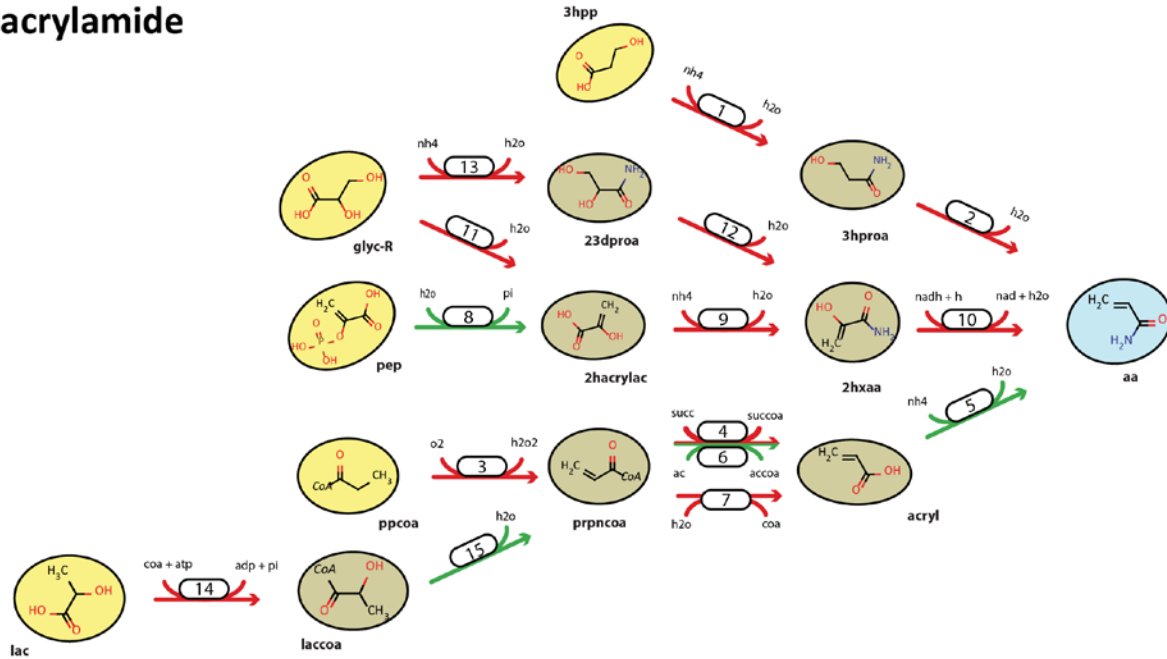
Reaction Specification Table

Reaction	Subs. TC/Enz. Prom. Coef	Enz. Prom. Coef	E.C. number	Specie	ΔG_{rxn}
1	0.1/0.94	0.94	3.5.1.4	Pseudonocardia sp	-62
2	0.2/1.01	1.013	4.2.1.84	Pseudomonas chlororaphis	-0.65
3	0.004/0.25	0.25	1.3.3.6	Arabidopsis thaliana	-18.6
4	0.0138/0.92	0.92	2.8.3.8	Acetobacter aceti	1.7
5	0/0	0	3.5.1.4	Pseudomonas aeruginosa	24.6
6	0/0	0	2.8.3.1	Clostridium propionicum	8.8
7	0.01/0.8	0.8	3.1.2.2	Sus scrofa	-42.8
8	0/0	0	3.1.3.2	Aspergillus niger	-37.3
9	0.29/1.1	1.1	3.5.1.4	Pseudomonas aeruginosa	-3.4
10	0.72/1.15	1.15	1.17.1.4	Triticum aestivum	-54.1
11	0.36/0.89	0.89	4.2.1.2	Escherichia coli	-12.5
12	0.18/1.13	1.13	4.2.1.84	Brevibacterium sp.	-40
13	0.09/0.99	0.99	3.5.1.4	Brevibacterium sp.	24.4
14	0.32/1.09	1.09	6.2.1.13	Giardia intestinalis	-5
15	0/0	0	4.2.1.4	Clostridium propionicum	6.3

Pathway Specification Table

Pathway	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	1, 2	3hpp	0.00	0.00	0.00	0.00	23.69	23.52	28.02
2	3, 4, 5	ppcoa	0.33	38.57	38.57	19.43	38.57	38.57	38.64
3	3, 6, 5	ppcoa	0.67	38.57	38.57	19.43	38.57	38.57	38.64
4	3, 7, 5	ppcoa	0.33	24.37	21.23	10.79	36.61	36.46	38.64
5	8, 9, 10	pep	0.33	24.35	21.21	10.77	36.60	36.46	38.64
6	11, 9, 10	glyc-R	0.00	5.70	4.90	2.33	27.60	27.39	31.11
7	13, 12, 10	glyc-R	0.00	5.70	4.90	2.33	27.60	27.39	31.11
8	5, 4, 15, 14	lac	0.5	12.61	10.61	6.76	33.15	33.07	38.64
9	5, 6, 15, 14	lac	0.75	12.61	10.61	6.76	33.15	33.07	38.64
10	5, 4, 15, 14	lac	0.50	10.09	8.49	5.29	31.47	31.36	37.37
	Max. Exp & Comp			0.00	0.00	0.00	0.00	0.00	0.00
	Max. GEM-Path			38.57	38.57	19.43	38.57	38.57	38.64

2,3-butanediol
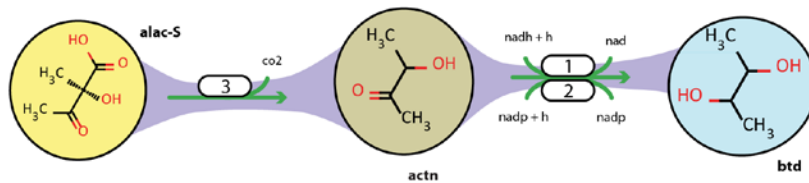
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0/0	1.1.1.4	Saccharomyces cerevisiae	-47.3
2	0/0	1.1.1.162	Gallus gallus	-67.2
3	0/0	4.1.1.5	Streptococcus thermophilus	-37.2

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1[exp]	1, 3	alac-S	1.00	20.78	19.78	15.16	20.99	20.81	23.68
2[exp]	2, 3	alac-S	1.00	18.98	16.58	12.58	20.68	20.51	23.33
	Max. Exp & Comp			20.78	19.78	15.16	20.99	20.81	23.68
	Max. GEM-Path			0.00	0.00	0.00	0.00	0.00	0.00

acrylic acid
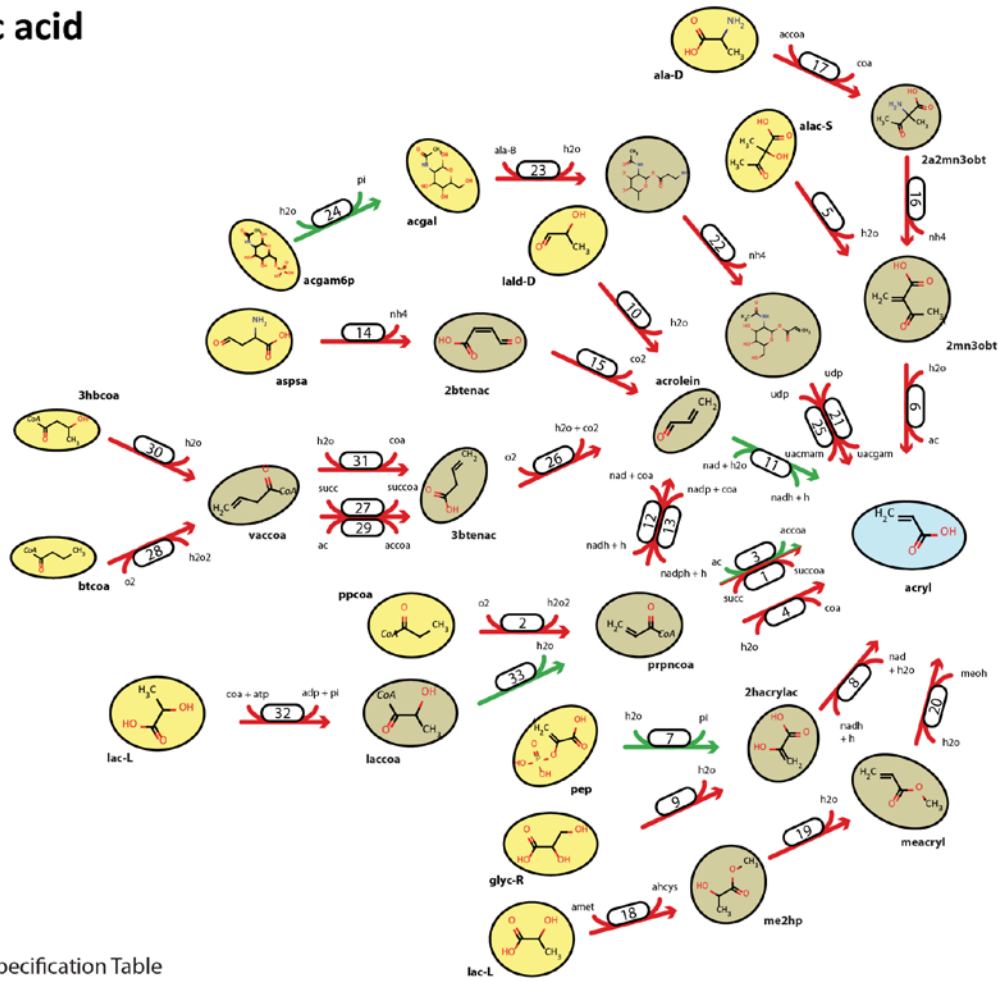
Reaction Specification Table

Reaction	Subs. TC/React. Prom Score	E.C. number	Specie	ΔG_{rxn}
1	0.0138/0.91	2.8.3.8	Acetobacter aceti	1.7
2	0.005/0.24	1.3.3.6	Mus musculus	-18.6
3	0/0	2.8.3.1	Clostridium propionicum	8.8
4	0.009/0.8	3.1.2.2	Mycobacterium tuberculosis	-42.8
5	0.3/1.17	4.2.1.9	Salmonella enterica	4.2
6	0.5/1.01	3.7.1.2	Bos Taurus	-62.4
7	0/0	3.1.3.2	Aspergillus niger	-37.3
8	0.55/1.14	1.17.1.4	Triticum aestivum	89.1
9	0.36/0.89	4.2.1.2	Escherichia coli	-12.5
10	0.53/0.9	4.2.1.28	Salmonella enterica	7.14
11	0/0	1.2.1.22	Methanocaldococcus jannaschi	-3.9
12	0.013/1.18	1.2.1.10	Escherichia coli	-38.9
13	0.025/0.83	1.2.1.76	Clostridium kluyveri	-58.8
14	0.13/0.71	4.3.1.1	Escherichia coli	-41.2
15	0.62/1.14	4.1.1.1	Saccharomyces cerevisiae	-47.7
16	0.35/0.88	4.3.1.17	Mus musculus	-40.4
17	0.08/1.16	2.3.1.30	Salmonella enterica	-160
18	0.53/1.08	2.1.1.9	Rattus norvegicus	-24.3
19	0.52/1.08	4.2.1.2	Sus scrofa	7.3
20	0.31/0.99	3.1.1.1	Pseudomonas fluorescens	-11
21	0.21/1.17	2.4.1.135	Homo sapiens	-52.6
22	0.7/1.1	4.3.1.14	Clostridium sp.	-44.6
23	0.1/0.82	3.1.1.72	Streptomyces lividans	24.8
24	0/0	3.1.3.29	Homo sapiens	-179.9
25	0.21/1.17	2.4.1.135	Homo sapiens	-58.1
26	0.48/1.12	1.13.12.4	Mycobacterium smegmatis	-367.5
27	0.0182/0.89	2.8.3.5	Homo sapiens	-3.4
28	0.01/0.56	1.3.3.6	Rattus norvegicus	-25
29	0.0023/1	2.8.3.14	Clostridium aminovalericum	3.7
30	0.007/0.55	4.2.1.119	Aeromonas caviae	12.8
31	0.01/0.87	3.1.2.1	Pisum sativum	-47.9
32	0.32/1.09	6.2.1.13	Giardia intestinalis	-5
33	0/0	4.2.1.4	Clostridium propionicum	6.3

Pathway Specification Table

Pathway Nr	Rxn Chain	Precursor	Pathway Index	Anaerobic			Aerobic		
				Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
1	1 , 2	ppcoa	0.00	38.57	38.57	19.43	38.57	38.57	38.64
2	3 , 2	ppcoa	0.50	38.57	38.57	19.43	38.57	38.57	38.64
3	4 , 2	ppcoa	0.00	24.37	21.23	10.79	36.61	36.46	38.64
4	5 , 6	alac-S	0.00	23.73	20.73	12.64	36.46	36.04	38.64
5	7 , 8	pep	0.50	24.35	21.21	10.77	36.60	36.46	38.64
6	9 , 8	glyc-R	0.00	5.70	4.90	2.33	27.60	27.39	31.11
7	10 , 11	lald-D	0.50	16.81	14.14	7.46	34.54	34.37	38.64
8	2 , 12 , 11	ppcoa	0.33	24.37	21.23	10.79	36.61	36.46	38.64
9	2 , 13 , 11	ppcoa	0.33	18.93	15.93	10.79	35.47	35.39	38.64
10	14 , 15 , 11	aspsa	0.00	10.81	9.09	6.76	32.09	32.02	38.10
11	17 , 16 , 6	ala-D	0.33	21.61	18.18	12.64	36.45	36.02	38.64
12	18 , 19 , 20	glyc-R	0.00	6.77	5.70	3.49	11.25	11.14	12.88
13	21 , 22 , 23 , 24	acgal	0.25	6.17	5.20	3.27	26.89	26.84	31.92
14	25 , 22 , 23 , 24	acgal	0.25	6.17	5.20	3.27	26.89	26.84	31.92
15	11 , 26 , 27 , 28	btcoa	0.50	0.00	0.00	0.00	26.55	26.33	29.88
16	11 , 26 , 29 , 28	btcoa	0.50	0.00	0.00	0.00	26.55	26.33	29.88
17	11 , 26 , 27 , 30	3hbcoa	0.50	0.00	0.00	0.00	26.55	26.33	29.88
18	11 , 26 , 29 , 30	3hbcoa	0.50	0.00	0.00	0.00	26.55	26.33	29.88
19	11 , 26 , 31 , 28	btcoa	0.50	0.00	0.00	0.00	25.81	25.58	29.02
20	11 , 26 , 31 , 30	3hbcoa	0.50	0.00	0.00	0.00	25.81	25.58	29.02
21	1 , 33 , 32	lac	0.33	12.61	10.61	6.76	33.15	33.07	38.64
22	3 , 33 , 32	lac	0.67	12.61	10.61	6.76	33.15	33.07	38.64
23	4 , 33 , 32	lac	0.33	10.09	8.49	5.29	31.47	31.36	37.37
	Max. Exp & Comp			0.00	0.00	0.00	0.00	0.00	0.00
	Max. GEM-Path			38.57	38.57	19.43	38.57	38.57	38.64

Table A2-1. Pathway Prediction Tools Comparison

	BROs Retrosynthetic Based Approach	Thermodynamic Analysis	Prioritization	Toxicity Analysis	GEMs Integration	Chemoinformatic Metabolome Search	Knock Out for Growth Coupled Designs	EC Number Reaction Association	Gene - Reaction Association	Promiscuous Analysis	Enzyme performance
GEM-Path	✓	✓	✓	✗	✓	✓	✓: RobustKnock[29]	✓: BRENDA[22]	✗	✓: BRENDA[22]	✗
RetroPath[38]	✓	✓	✓	✓	✗	✗	✗	✓: KEGG[14]	✓: KEGG[14]	✓: KEGG[14]	✓: BRENDA[22]
BNICE[39]	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
System of Cho et al.[7]	✓	✓	✓	✗	✗	✗	✗	✓: KEGG[14]	✓: KEGG[14]	✓: KEGG[14]	✗
Biopathway[35]	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
OptStrain[40]	✗	✗	✗	✗	✓	✗	✓: OptKnock[5]	✓: KEGG[14]	✓: KEGG[14]	✗	✗
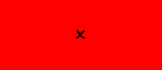
✓	Present in the Pathway prediction Algorithm
✗	Not Present in the Pathway prediction Algorithm

Table A2-2. High Yield growth-coupled designs of synthetic pathways specifications

Design identifier	Product ID	Substrate	Oxygenation	Reaction KO abbreviation	Reaction KI identifier					
					Reaction Formula	Reaction ID	Delta Gr	Species	EC Number	Prom Idex
1	3hpp	Glycerol	Aerobic	GLYK ; GLYCDx ; ALCD19	glyc -> 3hppnl + h2o	1.1	-28.3	*Klebsiella pneumoniae*	4.2.1.28	0/0
					3hppnl + nad + h2o -> 3hpp+ nadh + h	1.2	-6.2	*Escherichia coli*	1.2.1.3	0/0
2	3hpp	Glycerol	Aerobic	GLYK ; GLYCDx ; ALCD19	1.1 && 3hppnl + nadp + h2o -> 3hpp + nadph + h	2.1	13.6	*Escherichia coli*	1.2.1.B6	0/0
3	aa	D-Xylose	ECOM	PYK ; F6PA ; ACKr	ppcoa + o2 -> prpncoa + h2o2	3.1	-18.6	*Mus musculus*	1.3.3.6	0.005/0.25
					prpncoa + succ -> acryl + succoa	3.2	1.7	*Acetobacter aceti*	2.8.3.8	0.014/0.91
					acryl + nh4 -> aa+ h2o	3.3	24.6	*Pseudomonas aeruginosa*	3.5.1.4	0/0
4	acryl	D-Xylose	ECOM	PYK ; DHAPT ; ACKr	3.1 + 3.2					
5	acryl	D-Xylose	ECOM	PYK ; GLYK ; F6PA ; ACKr	3.1 + 3.2					
6	3hpp	D-Xylose	ECOM	PYK ; DHAPT ; ACKr	3.1 + 3.2 && acryl + h2o -> 3hpp	6.1	-2.1	*Sus scrofa*	4.2.1.2	0.39/1.02
7	3hpp	D-Xylose	ECOM	PYK ; F6PA ; ACKr	3.1 + 3.2 + 8.1					
8	aa	Glycerol	ECOM	PYK ; GLYCDx ; ACKr	3.1 +3.2 +3.3					
9	acryl	Glycerol	ECOM	PYK ; GLYCDx ; ACKr	3.1 + 3.2					
10	3hpp	Glycerol	ECOM	PYK ; GLYCDx ; ACKr	3.1 + 3.2 + 6.1					
11	btoh	Glycerol	Anaerobic	PFL ; G3PD2	btcoa + succ -> but + succoa	11.1	-9.69	*Homo Sapiens*	2.8.3.5	0.001/0.95
					but + nadph + h -> btal + nadp + h2o	11.2	-11.5	*Acetobacter aceti*	1.2.1.4	0/0
					btal + nadh + h -> btoh + nad	11.3	-52.7	*Saccharomyces cerevisiae*	1.1.1.1	0/0
12	btoh	Glycerol	Anaerobic	PFL ; G3PD2	11.1 && but + nadh + h -> btal + nad + h2o	12.1	8.4	*Escherichia coli*	1.2.1.19	0/0
					btal + nadph + h -> btoh + nadp	12.2	-72.6	*Cucumis melo*	1.1.1.1	0/0
13	btoh	Glycerol	Anaerobic	PFL ; GLYK	12.1 + 12.2 && btcoa + ac -> but + accoa	13.1	-2.6	*Escherichia coli*	2.8.3.3	0/0
14	btoh	Glycerol	Anaerobic	PYK ; PFL ; G3PD2	11.1 + 11.2 + 11.3					
15	btoh	Glycerol	Anaerobic	PYK ; PFL ; G3PD2	11.1 + 12.1 + 12.2					
16	13ppd	Glycerol	Anaerobic	GLYK ; ALCD2x	1.1 && 3hppnl + nadh + h -> 13ppd + nad	16.1	-53.7	*Klebsiella pneumoniae*	1.1.1.202	0/0
17	13ppd	Glycerol	Anaerobic	PYK ; ACALD	16.1 && glyald -> ppdal + h2o	17.1	-82.5	*Klebsiella oxytoca*	4.2.1.28	0.3/0.8
					ppdal + nadh + h -> 3hppnl + nad	17.2	-17.7	*Saccharomyces cerevisiae*	1.1.1.1	0.13/0.8
18	13ppd	Glycerol	Anaerobic	PYK ; GLYCDx ; ATPS4rpp	1.1 + 16.1					
19	13ppd	Glycerol	Anaerobic	PYK ; FLDR2 ; DHAPT	1.1 && 3hppnl + nadph + h -> 13ppd + nadp	19.1	-73.6	*Escherichia coli*	1.1.1.202	0/0
20	13ppd	Glycerol	Anaerobic	PDH ; G3PD2 ; ACALD	17.1 + 17.2 + 16.1					

Table A2-3. Predicted reactions and 3rd EC Numbers NC-IUBMB frequency specifications

3rd EC Number	Frequency	Nomenclature Committee of the International Union of Biochemistry and Molecular Biology
1.1.1	49	Oxidoreductases, acting on CH-OH group of donors, with NAD or NADP acceptor
1.2.1	39	Oxidoreductases, acting on aldehyde or oxo group of donors, with NAD or NADP acceptor
4.2.1	22	Lyases, Carbon-Oxygeb Lyases, Hydro-Lyases
4.1.1	21	Lyases, Carbon-Carbon Lyases, Carboxy-Lyases
2.8.3	17	Transferases, Transferring Sulfur-Containing Groups, CoA-transferases
3.1.2	10	Hydrolases, Acting on Ester Bonds, Thioester Hydrolases
4.1.2	8	Lyases, Carbon-Carbon Lyases, Aldehyde-Lyases
4.3.1	5	Lyases, Carbon-Nitrogen Lyases, Ammonia-Lyases
3.1.3	4	Hydrolases, Acting on Ester Bonds, Phosphoric Monoester Hydrolases
1.3.3	4	Oxidoreductases, acting on the CH-CH group of donors, with oxygen as acceptor
2.3.3	4	Transferases, Acyltransferases, Acyl groups converted into alkyl on transfer
3.5.1	4	Hydrolases, Acting on Carbon-Nitrogen Bonds, other than Peptide Bonds, In Linear Amides
1.4.3	3	Oxidoreductases, acting on the CH-NH2 group of donors, with oxygen acceptor
1.13.12	3	Oxidoreductases, Acting on single donors with incorporation of molecular oxygen (oxygenases), With incorporation of one atom of oxygen (internal monooxygenases or internal mixed function oxidases)
1.17.1	3	Oxidoreductases, Acting on CH or CH2 groups, With NAD+ or NADP+ as acceptor
6.2.1	3	Ligases,Forming Carbon-Sulfur Bonds, Acid-Thiol Ligases
1.2.4	3	Oxidoreductases, acting on aldehyde or oxo group of donors, with a disulfide as acceptor
2.1.2	2	Transferases, Transferring One-Carbon Groups,Hydroxymethyl-, Formyl- and Related Transferases
2.3.1	2	Transferases, Acyltransferases, Transferring groups other than amino-acyl groups
2.4.1	2	Transferases, Glycosyltransferases, Hexosyltransferases
2.1.1	2	Transferases, Transferring One-Carbon Groups, Methyltransferases
1.4.1	2	Oxidoreductases, acting on the CH-NH2 group of donors, with NAD or NADP acceptor
3.1.1	2	Hydrolases, Acting on Ester Bonds, Carboxylic Ester Hydrolases
1.14.12	1	Oxidoreductases, Acting on paired donors, with incorporation or reduction of molecular oxygen, With NADH or NADPH as one donor, and incorporation of two atoms of oxygen into one donor
1.14.13	1	Oxidoreductases, Acting on paired donors, with incorporation or reduction of molecular oxygen, With NADH or NADPH as one donor, and incorporation of one atom of oxygen
2.1.3	1	Transferases, Transferring One-Carbon Groups, Carboxy- and Carbamoyltransferases
3.6.1	1	Hydrolases, Acting on Acid Anhydrides, In Phosphorus-Containing Anhydrides
3.7.1	1	Hydrolases, Acting on Carbon-Carbon Bonds, In Ketonic Substances
1.1.3	1	Oxidoreductases, acting on CH-OH group of donors, with oxygen as acceptor
5.4.99	1	Isomerases, Intramolecular Transferases, Thansfering Other Groups

Table A2-4. Knock Out Reactions frequency specified by Oxygenation, Substrate and Yield conditions

KO Reaction	Frequency	Substrate			Oxygenation			Yield Interval				
		glc	xyl	glyc	aerobic	ECOM	anaerobic	0,0 - 0,2	0,2 - 0,4	0,4 - 0,6	0,6 - 0,8	0,8 - 1,0
ORNDC	15	5	6	4	12	0	3	12	0	3	0	0
FUM	72	34	27	11	57	3	12	44	16	11	1	0
GND	19	12	6	1	13	5	1	14	4	1	0	0
PYK	328	26	174	128	97	68	163	70	95	132	23	8
DHAPT	44	8	15	21	12	7	25	9	0	27	6	2
ACKr	107	28	56	23	32	45	30	28	24	42	5	8
PGCD	65	25	14	26	45	7	13	33	19	13	0	0
RPI	22	12	0	10	11	7	4	15	4	3	0	0
SUCOAS	38	10	8	20	9	11	18	31	0	7	0	0
AKGDH	21	9	8	4	9	8	4	21	0	0	0	0
ACCOAL	26	7	7	12	7	7	12	21	0	5	0	0
TALA	24	4	18	2	6	9	9	10	3	8	3	0
ACACT1r	37	7	23	7	4	9	24	8	13	16	0	0
G6PDH2r	38	27	10	1	3	20	15	9	23	5	1	0
POR5	18	3	3	12	3	4	11	4	7	4	3	0
PFL	528	156	127	245	284	26	218	172	172	148	36	0
PDH	105	45	35	25	6	80	19	17	71	15	2	0
TPI	116	62	24	30	64	13	39	45	32	18	21	0
ATPS4rpp	534	171	168	195	206	139	189	154	239	118	23	0
TKT1	14	2	10	2	12	0	2	4	10	0	0	0
PPCSCT	12	3	1	8	2	4	6	10	0	2	0	0
ASPO3	1	1	0	0	1	0	0	1	0	0	0	0
CYTBO3_4pp	24	4	8	12	24	0	0	10	13	1	0	0
THD2pp	44	14	9	21	15	0	29	9	22	3	10	0
PGM	9	5	3	1	8	1	0	5	2	1	1	0
GAPD	8	6	0	2	7	1	0	6	1	1	0	0
TKT2	70	18	12	40	42	2	26	27	27	16	0	0
ENO	11	3	3	5	8	3	0	7	3	1	0	0
PPS	15	2	1	12	15	0	0	4	3	0	8	0
GHMT2r	12	1	1	10	3	2	7	5	6	1	0	0
PPC	72	18	12	42	19	3	50	16	22	24	10	0
LDH_D	56	25	20	11	3	6	47	8	17	26	5	0
ACALD	94	20	23	51	14	26	54	12	30	36	16	0
PGI	124	117	6	1	4	27	93	27	86	11	0	0
GLYCL	9	1	1	7	8	0	1	8	1	0	0	0
MALS	5	0	5	0	5	0	0	0	5	0	0	0
RPE	117	0	117	0	68	9	40	58	41	16	2	0
F6PA	19	8	11	0	1	6	12	2	4	7	3	3
CYTBDpp	1	0	1	0	1	0	0	0	1	0	0	0
CYTBD2pp	1	0	1	0	1	0	0	0	1	0	0	0
MDH2	1	0	0	1	1	0	0	1	0	0	0	0
IPDDI	9	1	3	5	2	0	7	3	4	2	0	0
GLYK	85	0	1	84	21	18	46	22	11	25	24	3
GLYCDx	81	0	0	81	31	19	31	0	35	37	4	5
ALCD19	3	0	0	3	2	0	1	0	0	0	1	2
GLUDy	129	26	37	66	2	112	15	12	64	33	20	0
FLDR2	23	1	6	16	5	5	13	6	7	8	2	0
ALCD2x	110	26	24	60	10	24	76	12	45	33	20	0
PPKr	51	0	11	40	8	1	42	9	24	18	0	0
MDH	173	39	74	60	1	148	24	16	80	60	17	0
ASPO6	5	1	4	0	0	5	0	4	1	0	0	0
ALDD2y	1	0	1	0	0	1	0	0	0	1	0	0
G3PD2	26	0	0	26	0	1	25	1	0	12	13	0
PYAM5PO	1	0	0	1	0	1	0	0	0	1	0	0
ASPT	1	0	0	1	0	1	0	0	0	1	0	0
RNTR3c2	1	0	0	1	0	1	0	1	0	0	0	0
ATHRDHr	1	0	0	1	0	1	0	1	0	0	0	0
HEX7	10	10	0	0	0	0	10	0	8	2	0	0
XYLI2	19	18	1	0	0	0	19	0	17	2	0	0
EDA	4	2	2	0	0	0	4	0	4	0	0	0
HEX1	14	14	0	0	0	0	14	0	14	0	0	0
PFK	2	1	0	1	0	0	2	0	1	0	1	0
CBPS	18	4	8	6	0	0	18	1	10	7	0	0
CBMKr	20	4	10	6	0	0	20	1	12	7	0	0
FRD2	2	0	2	0	0	0	2	0	1	1	0	0
FRD3	1	0	1	0	0	0	1	0	0	1	0	0
DRPA	1	0	1	0	0	0	1	1	0	0	0	0
NADH18pp	3	0	2	1	0	0	3	0	1	2	0	0
NADH17pp	1	0	1	0	0	0	1	0	1	0	0	0
MGSA	8	0	8	0	0	0	8	6	2	0	0	0
AGDC	1	0	1	0	0	0	1	0	0	1	0	0

ME2	6	0	4	2	0	0	6	0	0	4	2	0
CYTK1	1	0	1	0	0	0	1	0	0	1	0	0
G6PDA	1	0	0	1	0	0	1	0	0	1	0	0
THRD	6	0	0	6	0	0	6	0	6	0	0	0
THRAi	6	0	0	6	0	0	6	0	6	0	0	0
PPM	1	0	0	1	0	0	1	0	0	1	0	0
DMPPS	1	0	0	1	0	0	1	1	0	0	0	0
UPPRT	1	0	0	1	0	0	1	0	0	1	0	0
R15BPK	1	0	0	1	0	0	1	0	0	1	0	0
ALATA_D2	2	0	0	2	0	0	2	0	2	0	0	0
PRPPS	1	0	0	1	0	0	1	0	0	1	0	0
DURIK1	1	0	0	1	0	0	1	0	0	1	0	0
NDPK3	4	0	0	4	0	0	4	4	0	0	0	0

Table A2-5. Pathways Strain Design Yield average for different oxygenation/substrate/KO combinations

Target / Pathway	Overall	oxygenation			substrate			Knock Outs		
		aerobic	ECOM	anaerobic	glc	xyl	glyc	2 KO	3 KO	4 KO
acrylamide	40	40	40	40	40	40	40	40	40	40
2	'0.33 / 22'	'0.21 / 6'	'0.38 / 8'	'0.37 / 8'	'0.24 / 8'	'0.48 / 6'	'0.30 / 8'	'0.25 / 9'	'0.47 / 9'	'0.20 / 4'
3	'0.29 / 22'	'0.12 / 7'	'0.34 / 7'	'0.39 / 8'	'0.21 / 9'	'0.42 / 8'	'0.22 / 5'	'0.21 / 9'	'0.31 / 7'	'0.37 / 6'
4	'0.18 / 25'	'0.08 / 7'	'0.19 / 9'	'0.24 / 9'	'0.14 / 9'	'0.26 / 9'	'0.11 / 7'	'0.14 / 8'	'0.19 / 9'	'0.20 / 8'
8	'0.01 / 8'	'0.01 / 3'	'0.01 / 2'	'0.00 / 3'	'0.01 / 3'	'0.01 / 2'	'0.00 / 3'	'0.00 / 1'	'0.01 / 6'	'0.00 / 1'
acrilic acid	43	39	43	30	43	43	43	30	43	39
1	'0.36 / 23'	'0.21 / 6'	'0.43 / 9'	'0.39 / 8'	'0.24 / 8'	'0.54 / 7'	'0.31 / 8'	'0.25 / 9'	'0.49 / 9'	'0.32 / 5'
2	'0.30 / 23'	'0.12 / 7'	'0.34 / 7'	'0.40 / 9'	'0.21 / 9'	'0.39 / 8'	'0.29 / 6'	'0.22 / 9'	'0.35 / 8'	'0.32 / 6'
3	'0.18 / 25'	'0.08 / 7'	'0.19 / 9'	'0.24 / 9'	'0.14 / 9'	'0.26 / 9'	'0.11 / 7'	'0.14 / 8'	'0.19 / 9'	'0.20 / 8'
4	'0.10 / 9'	'0.10 / 3'	'0.13 / 4'	'0.06 / 2'	'0.12 / 3'	'0.04 / 3'	'0.14 / 3'	'0 / 0'	'0.11 / 8'	'0.10 / 1'
8	'0.18 / 25'	'0.08 / 7'	'0.19 / 9'	'0.24 / 9'	'0.14 / 9'	'0.26 / 9'	'0.11 / 7'	'0.14 / 8'	'0.19 / 9'	'0.20 / 8'
9	'0.16 / 19'	'0.04 / 4'	'0.16 / 7'	'0.23 / 8'	'0.13 / 7'	'0.21 / 5'	'0.16 / 7'	'0.15 / 6'	'0.18 / 9'	'0.14 / 4'
15	'0.32 / 9'	'0.18 / 1'	'0.34 / 8'	'0 / 0'	'0.26 / 4'	'0.31 / 2'	'0.29 / 3'	'0.26 / 2'	'0.27 / 3'	'0.30 / 4'
16	'0.29 / 3'	'0 / 0'	'0.29 / 3'	'0 / 0'	'0.19 / 1'	'0.20 / 1'	'0.19 / 1'	'0 / 0'	'0.29 / 3'	'0 / 0'
17	'0.36 / 9'	'0.18 / 1'	'0.38 / 8'	'0 / 0'	'0.32 / 4'	'0.31 / 2'	'0.30 / 3'	'0.29 / 2'	'0.35 / 3'	'0.31 / 4'
21	'0.01 / 7'	'0.01 / 2'	'0.01 / 3'	'0.00 / 2'	'0.01 / 2'	'0.01 / 3'	'0.01 / 2'	'0 / 0'	'0.01 / 4'	'0.01 / 3'
3-hydroxypropanoate	28	28	20	20	20	20	28	24	28	24
12	'0.37 / 24'	'0.22 / 6'	'0.43 / 9'	'0.40 / 9'	'0.24 / 8'	'0.54 / 8'	'0.31 / 8'	'0.25 / 9'	'0.48 / 9'	'0.35 / 6'
13	'0.29 / 21'	'0.13 / 6'	'0.28 / 6'	'0.41 / 9'	'0.22 / 8'	'0.35 / 7'	'0.30 / 6'	'0.24 / 8'	'0.36 / 8'	'0.25 / 5'
14	'0.18 / 25'	'0.08 / 7'	'0.19 / 9'	'0.24 / 9'	'0.14 / 9'	'0.26 / 9'	'0.11 / 7'	'0.14 / 8'	'0.19 / 9'	'0.20 / 8'
22	'0.17 / 25'	'0.07 / 7'	'0.19 / 9'	'0.24 / 9'	'0.13 / 9'	'0.26 / 9'	'0.11 / 7'	'0.14 / 8'	'0.19 / 9'	'0.20 / 8'
23	'0.01 / 11'	'0.01 / 4'	'0.01 / 3'	'0.00 / 4'	'0.01 / 4'	'0.01 / 3'	'0.01 / 4'	'0.00 / 1'	'0.01 / 8'	'0.01 / 2'
9	'0.47 / 1'	'0.47 / 1'	'0 / 0'	'0 / 0'	'0 / 0'	'0 / 0'	'0.47 / 1'	'0 / 0'	'0.47 / 1'	'0 / 0'
10	'0.45 / 3'	'0.45 / 3'	'0 / 0'	'0 / 0'	'0 / 0'	'0 / 0'	'0.45 / 3'	'0.24 / 1'	'0.46 / 1'	'0.20 / 1'
1-propanol	63	63	21	63	63	58	63	63	63	42
1	'0.37 / 11'	'0.14 / 2'	'0.31 / 1'	'0.43 / 8'	'0.25 / 4'	'0.43 / 3'	'0.39 / 4'	'0.31 / 5'	'0.33 / 2'	'0.41 / 4'
2	'0.36 / 13'	'0.18 / 4'	'0 / 0'	'0.44 / 9'	'0.26 / 4'	'0.35 / 5'	'0.39 / 4'	'0.30 / 6'	'0.36 / 4'	'0.36 / 3'
3	'0.31 / 11'	'0.14 / 2'	'0 / 0'	'0.35 / 9'	'0.24 / 4'	'0.20 / 3'	'0.39 / 4'	'0.23 / 5'	'0.39 / 3'	'0.26 / 3'
4	'0.29 / 7'	'0.14 / 2'	'0 / 0'	'0.35 / 5'	'0.22 / 3'	'0 / 0'	'0.32 / 4'	'0.15 / 3'	'0.33 / 2'	'0.31 / 2'
5	'0.22 / 12'	'0.11 / 6'	'0 / 0'	'0.33 / 6'	'0.15 / 4'	'0.19 / 4'	'0.29 / 4'	'0.22 / 6'	'0.23 / 6'	'0 / 0'
6	'0.26 / 10'	'0.15 / 4'	'0 / 0'	'0.33 / 6'	'0.21 / 2'	'0.19 / 4'	'0.29 / 4'	'0.25 / 5'	'0.26 / 5'	'0 / 0'
7	'0.20 / 8'	'0.15 / 4'	'0 / 0'	'0.25 / 4'	'0.14 / 1'	'0.06 / 3'	'0.29 / 4'	'0.18 / 5'	'0.22 / 3'	'0 / 0'
8	'0.24 / 10'	'0.15 / 4'	'0 / 0'	'0.30 / 6'	'0.19 / 2'	'0.18 / 4'	'0.28 / 4'	'0.22 / 5'	'0.25 / 5'	'0 / 0'
9	'0.23 / 16'	'0.11 / 6'	'0.19 / 2'	'0.32 / 8'	'0.16 / 5'	'0.21 / 6'	'0.31 / 5'	'0.22 / 6'	'0.23 / 6'	'0.23 / 4'
10	'0.21 / 9'	'0.12 / 3'	'0 / 0'	'0.25 / 6'	'0.07 / 1'	'0.09 / 4'	'0.33 / 4'	'0.16 / 5'	'0.20 / 2'	'0.22 / 2'
13	'0.22 / 19'	'0.09 / 9'	'0.25 / 2'	'0.37 / 8'	'0.22 / 6'	'0.19 / 6'	'0.28 / 7'	'0.20 / 6'	'0.22 / 6'	'0.26 / 7'
14	'0.20 / 17'	'0.11 / 9'	'0.24 / 1'	'0.30 / 7'	'0.14 / 5'	'0.18 / 5'	'0.26 / 7'	'0.13 / 6'	'0.22 / 6'	'0.25 / 5'
isopropanol	14	14	7	14	14	14	14	14	14	14
1	'0.23 / 18'	'0.10 / 9'	'0.24 / 1'	'0.37 / 8'	'0.19 / 5'	'0.21 / 6'	'0.29 / 7'	'0.20 / 6'	'0.22 / 6'	'0.27 / 6'
2	'0.19 / 16'	'0.10 / 9'	'0 / 0'	'0.30 / 7'	'0.14 / 5'	'0.17 / 5'	'0.23 / 6'	'0.13 / 6'	'0.22 / 6'	'0.19 / 4'
1-butanol	100	100	72	100	100	100	100	100	100	89
1	'0.31 / 11'	'0.17 / 4'	'0.13 / 1'	'0.43 / 6'	'0.26 / 4'	'0.30 / 3'	'0.31 / 4'	'0.28 / 4'	'0.33 / 7'	'0 / 0'
2	'0.36 / 13'	'0.11 / 2'	'0.30 / 4'	'0.46 / 7'	'0.29 / 3'	'0.30 / 3'	'0.39 / 7'	'0.29 / 4'	'0.35 / 7'	'0.41 / 2'
3	'0.31 / 10'	'0.11 / 2'	'0.29 / 2'	'0.39 / 6'	'0.21 / 3'	'0.18 / 1'	'0.40 / 6'	'0.29 / 4'	'0.25 / 4'	'0.41 / 2'
4	'0.32 / 7'	'0.13 / 2'	'0 / 0'	'0.39 / 5'	'0.10 / 1'	'0.18 / 1'	'0.36 / 5'	'0.24 / 2'	'0.29 / 4'	'0.28 / 1'
5	'0.42 / 10'	'0.43 / 3'	'0 / 0'	'0.41 / 7'	'0.35 / 3'	'0.17 / 2'	'0.51 / 5'	'0.44 / 4'	'0.39 / 4'	'0.29 / 2'
6	'0.36 / 16'	'0.33 / 7'	'0.12 / 1'	'0.41 / 8'	'0.29 / 5'	'0.32 / 5'	'0.42 / 6'	'0.37 / 6'	'0.36 / 7'	'0.28 / 3'
7	'0.43 / 11'	'0.43 / 3'	'0 / 0'	'0.43 / 8'	'0.38 / 2'	'0.22 / 3'	'0.51 / 6'	'0.47 / 4'	'0.39 / 4'	'0.32 / 3'
8	'0.40 / 16'	'0.33 / 7'	'0.33 / 2'	'0.50 / 7'	'0.29 / 4'	'0.42 / 6'	'0.45 / 6'	'0.38 / 6'	'0.39 / 7'	'0.47 / 3'
9	'0.43 / 11'	'0.43 / 3'	'0 / 0'	'0.43 / 8'	'0.37 / 2'	'0.22 / 3'	'0.51 / 6'	'0.47 / 4'	'0.39 / 4'	'0.32 / 3'
10	'0.43 / 17'	'0.33 / 7'	'0.42 / 3'	'0.54 / 7'	'0.34 / 5'	'0.44 / 7'	'0.52 / 5'	'0.38 / 6'	'0.46 / 8'	'0.47 / 3'
11	'0.31 / 11'	'0.43 / 3'	'0.33 / 1'	'0.25 / 7'	'0.15 / 2'	'0.09 / 3'	'0.52 / 6'	'0.29 / 4'	'0.36 / 5'	'0.26 / 2'
12	'0.38 / 17'	'0.33 / 7'	'0.33 / 1'	'0.43 / 9'	'0.27 / 5'	'0.33 / 5'	'0.50 / 7'	'0.27 / 6'	'0.46 / 7'	'0.41 / 4'
13	'0.33 / 9'	'0.14 / 2'	'0.13 / 1'	'0.43 / 6'	'0.28 / 3'	'0.33 / 2'	'0.31 / 4'	'0.28 / 4'	'0.38 / 5'	'0 / 0'
14	'0.35 / 10'	'0.14 / 2'	'0.31 / 1'	'0.41 / 7'	'0.21 / 2'	'0.27 / 2'	'0.41 / 6'	'0.29 / 4'	'0.32 / 4'	'0.41 / 2'
15	'0.32 / 9'	'0.11 / 2'	'0.29 / 2'	'0.42 / 5'	'0.20 / 2'	'0.18 / 1'	'0.40 / 6'	'0.29 / 4'	'0.26 / 3'	'0.41 / 2'
16	'0.32 / 7'	'0.13 / 2'	'0 / 0'	'0.39 / 5'	'0.10 / 1'	'0.18 / 1'	'0.36 / 5'	'0.24 / 2'	'0.29 / 4'	'0.28 / 1'
17	'0.36 / 13'	'0.14 / 2'	'0.28 / 3'	'0.45 / 8'	'0.28 / 3'	'0.32 / 4'	'0.41 / 6'	'0.29 / 4'	'0.36 / 6'	'0.40 / 3'
18	'0.36 / 13'	'0.14 / 2'	'0.28 / 3'	'0.45 / 8'	'0.28 / 3'	'0.32 / 4'	'0.41 / 6'	'0.29 / 4'	'0.36 / 6'	'0.40 / 3'
isobutanol	77	8	69	15	46	46	62	8	77	8
1	'0.37 / 2'	'0 / 0'	'0 / 0'	'0.37 / 2'	'0.27 / 1'	'0.29 / 1'	'0 / 0'	'0 / 0'	'0.37 / 2'	'0 / 0'
5	'0.01 / 6'	'0.00 / 1'	'0.04 / 2'	'0.00 / 3'	'0.01 / 1'	'0.00 / 3'	'0.04 / 2'	'0.00 / 1'	'0.00 / 1'	'0.03 / 4'
6	'0.32 / 3'	'0 / 0'	'0.32 / 3'	'0 / 0'	'0.18 / 1'	'0.18 / 1'	'0.29 / 1'	'0 / 0'	'0.32 / 3'	'0 / 0'
7	'0.27 / 3'	'0 / 0'	'0.27 / 3'	'0 / 0'	'0.13 / 1'	'0.13 / 1'	'0.28 / 1'	'0 / 0'	'0.27 / 3'	'0 / 0'
8	'0.17 / 2'	'0 / 0'	'0.17 / 2'	'0 / 0'	'0.13 / 1'	'0.13 / 1'	'0 / 0'	'0 / 0'	'0.17 / 2'	'0 / 0'
10	'0.27 / 3'	'0 / 0'	'0.27 / 3'	'0 / 0'	'0.13 / 1'	'0.13 / 1'	'0.28 / 1'	'0 / 0'	'0.27 / 3'	'0 / 0'
9	'0.18 / 1'	'0 / 0'	'0.18 / 1'	'0 / 0'	'0 / 0'	'0 / 0'	'0.18 / 1'	'0 / 0'	'0.18 / 1'	'0 / 0'
11	'0.18 / 1'	'0 / 0'	'0.18 / 1'	'0 / 0'	'0 / 0'	'0 / 0'	'0.18 / 1'	'0 / 0'	'0.18 / 1'	'0 / 0'
12	'0.18 / 1'	'0 / 0'	'0.18 / 1'	'0 / 0'	'0 / 0'	'0 / 0'	'0.18 / 1'	'0 / 0'	'0.18 / 1'	'0 / 0'
13	'0.13 / 1'	'0 / 0'	'0.13 / 1'	'0 / 0'	'0 / 0'	'0 / 0'	'0.13 / 1'	'0 / 0'	'0.13 / 1'	'0 / 0'
1,3-propanediol	34	34	7	34	10	34	34	34	34	34
16	'0.21 / 10'	'0.16 / 4'	'0 / 0'	'0.25 / 6'	'0.07 / 1'	'0.16 / 4'	'0.27 / 5'	'0.14 / 3'	'0.24 / 3'	'0.20 / 4'
17	'0.21 / 8'	'0.16 / 4'	'0 / 0'	'0.26 / 4'	'0.07 / 1'	'0.15 / 2'	'0.26 / 5'	'0.17 / 2'	'0.23 / 3'	'0.17 / 3'
18	'0.21 / 8'	'0.16 / 4'	'0 / 0'	'0.26 / 4'	'0.07 / 1'	'0.15 / 2'	'0.26 / 5'	'0.17 / 2'	'0.23 / 3'	'0.17 / 3'
19	'0.18 / 7'	'0.16 / 4'	'0 / 0'	'0.20 / 3'	'0.07 / 1'	'0.11 / 2'	'0.22 / 4'	'0.10 / 2'	'0.17 / 3'	'0.20 / 2'
7	'0.35 / 7'	'0.26 / 2'	'0 / 0'	'0.39 / 5'	'0 / 0'	'0.14 / 2'	'0.45 / 5'	'0.35 / 2'	'0.31 / 3'	'0.28 / 2'
8	'0.31 / 7'	'0.26 / 2'	'0.25 / 1'	'0.35 / 4'	'0 / 0'	'0.09 / 1'	'0.41 / 6'	'0.28 / 2'	'0.40 / 3'	'0.20 / 2'
14	'0.22 / 9'	'0.16 / 3'	'0 / 0'	'0.25 / 6'	'0 / 0'	'0.16 / 4'	'0.27 / 5'	'0.14 / 3'	'0.24 / 3'	'0.21 / 3'
15	'0.22 / 9'	'0.16 / 3'	'0 / 0'	'0.27 / 6'	'0 / 0'	'0.15 / 2'	'0.26 / 5'	'0.17 / 2'	'0.23 / 3'	'0.18 / 3'
24	'0.22 / 9'	'0.16 / 3'	'0 / 0'	'0.25 / 6'	'0 / 0'	'0.16 / 4'	'0.27 / 5'	'0.14 / 3'	'0.24 / 3'	'0.21 / 3'
25	'0.22 / 7'	'0.16 / 3'	'0 / 0'	'0.27 / 4'	'0 / 0'	'0.15 / 2'	'0.26 / 5'	'0.17 / 2'	'0.23 / 3'	'0.18 / 2'

34	'0.35 / 7'	'0.31 / 2'	'0 / 0'	'0.36 / 5'	'0 / 0'	'0.14 / 2'	'0.45 / 5'	'0.35 / 2'	'0.32 / 3'	'0.26 / 2'
35	'0.30 / 7'	'0.26 / 2'	'0.25 / 1'	'0.34 / 4'	'0 / 0'	'0.09 / 1'	'0.40 / 6'	'0.28 / 2'	'0.38 / 3'	'0.20 / 2'
36	'0.30 / 7'	'0.26 / 2'	'0.25 / 1'	'0.34 / 4'	'0 / 0'	'0.09 / 1'	'0.40 / 6'	'0.28 / 2'	'0.38 / 3'	'0.20 / 2'
37	'0.22 / 6'	'0.21 / 2'	'0 / 0'	'0.22 / 4'	'0 / 0'	'0.06 / 1'	'0.27 / 5'	'0.16 / 2'	'0.27 / 2'	'0.14 / 2'
2,3-propanediol	0	0	0	0	0	0	0	0	0	0
1,4-butanediol	55	36	55	36	52	52	55	39	55	52
13	'0.29 / 25'	'0.18 / 9'	'0.32 / 7'	'0.39 / 9'	'0.22 / 8'	'0.30 / 9'	'0.38 / 8'	'0.27 / 7'	'0.30 / 9'	'0.32 / 9'
14	'0.29 / 23'	'0.19 / 9'	'0.29 / 5'	'0.40 / 9'	'0.18 / 7'	'0.30 / 8'	'0.38 / 8'	'0.26 / 6'	'0.29 / 9'	'0.32 / 8'
15	'0.29 / 23'	'0.19 / 9'	'0.29 / 5'	'0.40 / 9'	'0.18 / 7'	'0.30 / 8'	'0.38 / 8'	'0.26 / 6'	'0.29 / 9'	'0.32 / 8'
16	'0.29 / 21'	'0.19 / 9'	'0.30 / 3'	'0.38 / 9'	'0.21 / 6'	'0.28 / 7'	'0.35 / 8'	'0.26 / 6'	'0.32 / 8'	'0.27 / 7'
17	'0.29 / 23'	'0.19 / 9'	'0.29 / 5'	'0.40 / 9'	'0.18 / 7'	'0.30 / 8'	'0.38 / 8'	'0.26 / 6'	'0.28 / 9'	'0.32 / 8'
18	'0.29 / 21'	'0.19 / 9'	'0.30 / 3'	'0.39 / 9'	'0.21 / 6'	'0.29 / 7'	'0.35 / 8'	'0.26 / 6'	'0.32 / 8'	'0.27 / 7'
19	'0.29 / 21'	'0.19 / 9'	'0.30 / 3'	'0.39 / 9'	'0.21 / 6'	'0.29 / 7'	'0.35 / 8'	'0.26 / 6'	'0.32 / 8'	'0.27 / 7'
20	'0.28 / 20'	'0.18 / 9'	'0.13 / 2'	'0.41 / 9'	'0.21 / 6'	'0.24 / 7'	'0.38 / 7'	'0.26 / 6'	'0.30 / 6'	'0.28 / 8'
21	'0.29 / 23'	'0.18 / 9'	'0.29 / 5'	'0.40 / 9'	'0.18 / 7'	'0.30 / 8'	'0.38 / 8'	'0.26 / 6'	'0.29 / 9'	'0.32 / 8'
22	'0.29 / 21'	'0.19 / 9'	'0.30 / 3'	'0.38 / 9'	'0.20 / 6'	'0.29 / 7'	'0.35 / 8'	'0.26 / 6'	'0.32 / 8'	'0.27 / 7'
23	'0.29 / 21'	'0.19 / 9'	'0.30 / 3'	'0.39 / 9'	'0.21 / 6'	'0.29 / 7'	'0.35 / 8'	'0.26 / 6'	'0.32 / 8'	'0.27 / 7'
24	'0.28 / 20'	'0.19 / 9'	'0.13 / 2'	'0.41 / 9'	'0.21 / 6'	'0.24 / 7'	'0.38 / 7'	'0.26 / 6'	'0.30 / 6'	'0.28 / 8'
25	'0.29 / 21'	'0.19 / 9'	'0.30 / 3'	'0.39 / 9'	'0.21 / 6'	'0.29 / 7'	'0.35 / 8'	'0.26 / 6'	'0.32 / 8'	'0.27 / 7'
26	'0.28 / 20'	'0.19 / 9'	'0.13 / 2'	'0.41 / 9'	'0.21 / 6'	'0.24 / 7'	'0.38 / 7'	'0.26 / 6'	'0.30 / 6'	'0.28 / 8'
27	'0.28 / 20'	'0.19 / 9'	'0.13 / 2'	'0.41 / 9'	'0.21 / 6'	'0.24 / 7'	'0.38 / 7'	'0.26 / 6'	'0.30 / 6'	'0.28 / 8'
28	'0.27 / 18'	'0.18 / 9'	'0.20 / 1'	'0.37 / 8'	'0.13 / 5'	'0.27 / 6'	'0.36 / 7'	'0.22 / 6'	'0.28 / 6'	'0.29 / 6'
29	'0.36 / 9'	'0 / 0'	'0.36 / 9'	'0 / 0'	'0.27 / 3'	'0.28 / 3'	'0.34 / 3'	'0.14 / 3'	'0.37 / 3'	'0.39 / 3'
30	'0.32 / 4'	'0 / 0'	'0.32 / 4'	'0 / 0'	'0.10 / 1'	'0.11 / 1'	'0.40 / 2'	'0 / 0'	'0.25 / 3'	'0.30 / 1'
31	'0.32 / 4'	'0 / 0'	'0.32 / 4'	'0 / 0'	'0.10 / 1'	'0.11 / 1'	'0.40 / 2'	'0 / 0'	'0.25 / 3'	'0.30 / 1'
32	'0.25 / 4'	'0 / 0'	'0.25 / 4'	'0 / 0'	'0.10 / 1'	'0.19 / 2'	'0.24 / 1'	'0 / 0'	'0.23 / 3'	'0.18 / 1'
33	'0.32 / 4'	'0 / 0'	'0.32 / 4'	'0 / 0'	'0.10 / 1'	'0.11 / 1'	'0.40 / 2'	'0 / 0'	'0.25 / 3'	'0.30 / 1'
34	'0.25 / 4'	'0 / 0'	'0.25 / 4'	'0 / 0'	'0.10 / 1'	'0.19 / 2'	'0.24 / 1'	'0 / 0'	'0.23 / 3'	'0.18 / 1'
35	'0.25 / 4'	'0 / 0'	'0.25 / 4'	'0 / 0'	'0.10 / 1'	'0.19 / 2'	'0.24 / 1'	'0 / 0'	'0.23 / 3'	'0.18 / 1'
36	'0.19 / 1'	'0 / 0'	'0.19 / 1'	'0 / 0'	'0 / 0'	'0 / 0'	'0.19 / 1'	'0 / 0'	'0.19 / 1'	'0 / 0'
2,3-butanediol	50	0	0	50	50	50	0	0	50	0
1	'0.26 / 2'	'0 / 0'	'0 / 0'	'0.26 / 2'	'0.18 / 1'	'0.20 / 1'	'0 / 0'	'0 / 0'	'0.26 / 2'	'0 / 0'
3-methyl-1-butanol	0	0	0	0	0	0	0	0	0	0
2-methyl-1-butanol	0	0	0	0	0	0	0	0	0	0
2-keto-isovaleric acid	33	0	33	0	0	0	33	0	0	33
2	'0.27 / 1'	'0 / 0'	'0.27 / 1'	'0 / 0'	'0 / 0'	'0 / 0'	'0.27 / 1'	'0 / 0'	'0 / 0'	'0.27 / 1'
3-hydroxyvalerate	33	0	33	0	0	0	33	33	0	33
3	'0.00 / 2'	'0 / 0'	'0.00 / 2'	'0 / 0'	'0 / 0'	'0 / 0'	'0.00 / 2'	'0.00 / 1'	'0 / 0'	'0.00 / 1'
2-keto-butanoic acid	20	20	20	20	20	20	20	20	20	20
1	'0.31 / 17'	'0.11 / 5'	'0.23 / 3'	'0.45 / 9'	'0.31 / 4'	'0.34 / 6'	'0.27 / 7'	'0.28 / 4'	'0.29 / 7'	'0.33 / 6'
3-hydroxybutyrate	11	11	11	11	11	11	11	11	11	11
1	'0.34 / 17'	'0.34 / 5'	'0.47 / 8'	'0.08 / 4'	'0.37 / 7'	'0.29 / 4'	'0.34 / 6'	'0.26 / 6'	'0.37 / 6'	'0.39 / 5'
4-hydroxybutyrate	0	0	0	0	0	0	0	0	0	0
2-phenylethanol	0	0	0	0	0	0	0	0	0	0

Table A2-6. Metabolites abbreviation names

abbreviation	Compound Name	abbreviation	Compound Name
acryl	acrylic acid	glyald	D-glyceraldehyde
btd	2,3-butandiol	glyc	glycerol
btoh	1-butanol	glyclt	glycolate
ghb	4-hydroxybutyrate	glyc-R	(R)-glycerate
ibutoh	isobutanol	glx	glyoxylate
1poh	1-propanol	hpyr	hydroxypyruvate
12ppd	2,3-propanediol	ibt	isobutyric acid
13ppd	1,3-propanediol	prpncoa	Propenoyl-CoA
14btd	1,4-butanediol	lac	lactate
2mbtoh	2-methyl-1-butanol	laccoa	Lactoyl-CoA
2obut	2-oxobutanoate	ladl-LD	lactaldehyde
2ppoh	isoporpanol	lald-D	D-lactaldehyde
2phetoh	2-Phenylethanol	lald-L	L-lactaldehyde
3hb-R	(R)-3-hydroxybutyrate	mal-D	D-malate
3hpt	3-hydroxyvalerate	meoh	methanol
3hpp	3-hydroxypropanoate	mlthf	5,10-methylenetetrahydrofolate
3mob	alpha-ketoisovalerate	mmcoa-S	(S)-methylmalonyl-CoA
3mop	ketoisoleucine	msa	3-Oxopropanoate
4mop	2-oxoisocaproate	oaa	oxaloacetate
12ppd-R	D-1,2-propanediol	pacald	phenylacetaldehyde
2dhp	2-dehydropantoate	pant-R	pantoate
2abut	(S)-2-aminobutanoate	pep	phosphoenolpyruvate
2mbald	2-methylbutanal	phpyr	phenylpyruvate
2mcit	2-methylcitrate	ppal	propanal
2oh2hmbut	2-hydroxy-2-(1-hydroxy-2-methylpropan-2-yl) butanedioic acid	ppap	propanoyl phosphate
		ppcoa	propanoyl-CoA
2opntn	2-oxopentanoic acid	succoa	succinyl-CoA
2pg	D-glycerate 2-phosphate	sucsal	succinic semialdehyde
2pglyc	2-phosphoglycolate	thf	tetrahydrofolate
3hbcoa	(S)-3-hydroxybutanoyl-CoA	vaccoa	Vinylacetyl-CoA
3hhcoa	(S)-3-hydroxyhexanoyl-CoA	val-L	L-valine
3hpcoa	3-hydroxypropanoyl-CoA	acrolein***	acrolein
3hppnl	3-Hydroxypropanal	ahyvala***	2-hydroxyisovaleric acid
3mbald	3-methylbutanal	allyoh***	2-propenol
34hpp	4-hydroxyphenylpyruvate	ibutal***	isobutyraldehyde
4abutn	4-aminobutanal	meacryl***	methyl acrylate
4hbutcoa	4-hydroxybutyryl-CoA	me2hp***	methyl 2-hydroxypropanoate
4hoxpacd	4-Hydroxyphenylacetaldehyde	neopglyc***	neopentyl glycol
5ohdmoxp	5-hydroxy-4,4-dimethyl-3-oxopentanoic acid	ppdal***	propanedial
aacoa	acetoacetyl-CoA	sucsal***	succindialdehyde
ac	acetate	sucsalcoa***	succindialdehyde-Coa
acac	acetoacetate	2a2mn3obt***	2-amino-2-methyl-3-oxobutanoic acid
acetoin	acetoin	2btenac***	2-Butenoicacid
acetone	acetone	2hacrylac***	2-hydroxyacrylic acid
acetol	acetol	2e2hobtae***	2-ethyl-2-hydroxybutanedioic acid
acgal	N-acetyl-D-galactosamine	2hacrylac***	2-hydroxyacrylic acid
acgam6p	N-Acetyl-D-glucosamine 6-phosphate	2hxaa***	2-hydroxyacrylamide
ahcys	S-adenosyl-L-homocysteine	2h4obute***	2-hydroxy-4-oxobutanoic acid
akg	2-oxoglutarate	2mn3obt***	2-methylidene-3-oxobutanoic acid
alac-S	(S)-2-acetolactate	24dhbut***	2,4-dihydroxybutyrate
ala-D	D-alanine	23dproa***	2,3-dihydroxypropanamide
amet	S-adenosyl-L-methionine	3btenac***	3-butenoic acid
appl	1-Aminopropan-2-ol	3hproa***	3-hydroxypropanamide
aspsa	L-aspartate 4-semialdehyde	3hoptal***	3-hydroxypentanal
athr-L	L-allothreonine	3hoptalcoa***	3-hydroxy pentanoil coa
btal	butanal	3hylpi***	(3-hydroxypropanoyl)oxyphosphonic acid
btcoa	butanoyl-CoA	3mbtoh***	2-methyl-1-butanol
but	butanoic acid	4hbtal***	4-hydroxybutanal
dohdmpen	3,5-dihydroxy-4,4-dimethylpentanoic acid	42hetphe***	4-(2-Hydroxyethyl)phenol
fald	formaldehyde		
fum	fumarate		

Metabolite names and the corresponding abbreviation were tabulated. The first 20 metabolites correspond to the target compounds. The rest are *E. coli* native and non-native intermediates. Abbreviations were taken from the BIGG database(Schellenberger et al., 2010). In case a specific metabolite wasn´t in the database, an abbreviation was assigned arbitrarily (***).

Bibliography

1. ALMAAS, E., B. KOVACS, T. VICSEK, Z. N. OLTVAI, et al. Global organization of metabolic fluxes in the bacterium Escherichia coli. Nature, Feb 26 2004, 427(6977), 839-843.

2. ALTARAS, N. E. AND D. C. CAMERON Metabolic engineering of a 1,2-propanediol pathway in Escherichia coli. Appl Environ Microbiol, Mar 1999, 65(3), 1180-1185.

3. ARITA, M. Metabolic reconstruction using shortest paths. Simulation Practice and Theory, 2000, 8(1), 109-125.

4. ARUNDEL, A. AND D. SAWAYA The bioeconomy to 2030: Designing a policy agenda 2009.

5. ASSARY, R. S. AND L. J. BROADBELT 2-Keto Acids to branched-chain alcohols as biofuels: Application of reaction network analysis and high-level quantum chemical methods to understand thermodynamic landscapes. Computational and Theoretical Chemistry, 2011.

6. ATSUMI, S., A. F. CANN, M. R. CONNOR, C. R. SHEN, et al. Metabolic engineering of Escherichia coli for 1-butanol production. Metabolic engineering, 2008, 10(6), 305-311.

7. ATSUMI, S. AND J. C. LIAO Metabolic engineering for advanced biofuels production from Escherichia coli. Curr Opin Biotechnol, Oct 2008, 19(5), 414-419.

8. ATSUMI, S., T. Y. WU, E. M. ECKL, S. D. HAWKINS, et al. Engineering the isobutanol biosynthetic pathway in Escherichia coli by comparison of three aldehyde reductase/alcohol dehydrogenase genes. Appl Microbiol Biotechnol, Jan 2010, 85(3), 651-657.

9. BAR-EVEN, A., E. NOOR, A. FLAMHOLZ, J. M. BUESCHER, et al. Hydrophobicity and charge shape cellular metabolite concentrations. PLoS Comput Biol, Oct 2011, 7(10), e1002166.

10. BARRETO, M., E. JEDLICKI AND D. S. HOLMES Identification of a gene cluster for the formation of extracellular polysaccharide precursors in the chemolithoautotroph Acidithiobacillus ferrooxidans. Appl Environ Microbiol, Jun 2005, 71(6), 2902-2909.

11. BAYER, T. S., D. M. WIDMAIER, K. TEMME, E. A. MIRSKY, et al. Synthesis of methyl halides from biomass using engineered microbes. J Am Chem Soc, May 13 2009, 131(18), 6508-6515.

12. BEARD, S., A. PARADELA, J. P. ALBAR AND C. A. JEREZ Growth of Acidithiobacillus Ferrooxidans ATCC 23270 in Thiosulfate Under Oxygen-Limiting Conditions Generates Extracellular Sulfur Globules by Means of a Secreted Tetrathionate Hydrolase. Front Microbiol, 2011, 2, 79.

13. BENNETT, B. D., E. H. KIMBALL, M. GAO, R. OSTERHOUT, et al. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol, Aug 2009, 5(8), 593-599.

14. BOLTON, E. E., Y. WANG, P. A. THIESSEN AND S. H. BRYANT PubChem: integrated platform of small molecules and biological activities. Annual reports in computational chemistry, 2008, 4, 217-241.

15. BOON, M. Theoretical and experimental methods in the modelling of bio-oxidation kinetics of sulphide minerals 1996.

16. BORDBAR, A., J. M. MONK, Z. A. KING AND B. O. PALSSON Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet, Feb 2014a, 15(2), 107-120.

17. BORDBAR, A., J. M. MONK, Z. A. KING AND B. O. PALSSON Constraint-based models predict metabolic and associated cellular functions. Nature Reviews Genetics, 2014b, 15(2), 107-120.

18. BRAMUCCI, M. G., D. FLINT, E. S. MILLER, V. NAGARAJAN, et al. Method for the production of 1-butanol. In.: U.S. Patent Application 2/110,503, 2008.

19. BRANDL, H. Microbial leaching of metals. Biotechnology Set, Second Edition, 2008, 191-224.

20. BRUSCELLA, P., C. APPIA-AYME, G. LEVICAN, J. RATOUCHNIAK, et al. Differential expression of two bc1 complexes in the strict acidophilic chemolithoautotrophic bacterium Acidithiobacillus ferrooxidans suggests a model for their respective roles in iron or sulfur oxidation. Microbiology, Jan 2007, 153(Pt 1), 102-110.

21. BUNKE, H. Graph matching: Theoretical foundations, algorithms, and applications. In *Proc. Vision Interface.* 2000, vol. 2000, p. 82-88.

22. BURGARD, A. P., P. PHARKYA AND C. D. MARANAS Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng, Dec 20 2003, 84(6), 647-657.

23. CAMPODONICO, M. A., B. A. ANDREWS, J. A. ASENJO, B. O. PALSSON, et al. Generation of an atlas for commodity chemical production in Escherichia coli and a novel pathway prediction algorithm, GEM-Path. Metab Eng, Jul 28 2014, 25C, 140-158.

24. CANN, A. F. AND J. C. LIAO Production of 2-methyl-1-butanol in engineered Escherichia coli. Appl Microbiol Biotechnol, Nov 2008, 81(1), 89-98.

25. CAO, Y., T. JIANG AND T. GIRKE A maximum common substructure-based algorithm for searching and predicting drug-like compounds. Bioinformatics, Jul 1 2008, 24(13), i366-374.

26. CARBAJOSA, S., M. MALKI, R. CAILLARD, M. F. LOPEZ, et al. Electrochemical growth of Acidithiobacillus ferrooxidans on a graphite electrode for obtaining a biocathode for direct electrocatalytic reduction of oxygen. Biosens Bioelectron, Oct 15 2010, 26(2), 877-880.

27. CARBONELL, P., A. G. PLANSON, D. FICHERA AND J. L. FAULON A retrosynthetic biology approach to metabolic pathway design for therapeutic production. BMC Syst Biol, 2011, 5, 122.

28. CASPI, R., T. ALTMAN, K. DREHER, C. A. FULCHER, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res, Jan 2012, 40(Database issue), D742-753.

29. CHI, A., L. VALENZUELA, S. BEARD, A. J. MACKEY, et al. Periplasmic proteins of the extremophile Acidithiobacillus ferrooxidans: a high throughput proteomics analysis. Mol Cell Proteomics, Dec 2007, 6(12), 2239-2251.

30. CHO, A., H. YUN, J. H. PARK, S. Y. LEE, et al. Prediction of novel synthetic pathways for the production of desired chemicals. BMC Syst Biol, 2010, 4, 35.

31. CIA, C. I. A. The world factbook. Retrieved November, 2011, 23, 2011.

32. CONNOR, M. R., A. F. CANN AND J. C. LIAO 3-Methyl-1-butanol production in Escherichia coli: random mutagenesis and two-phase fermentation. Appl Microbiol Biotechnol, Apr 2010, 86(4), 1155-1164.

33. COX, J. C., D. G. NICHOLLS AND W. J. INGLEDEW Transmembrane electrical potential and transmembrane pH gradient in the acidophile Thiobacillus ferro-oxidans. Biochem J, Jan 15 1979, 178(1), 195-200.

34. CURRAN, K. A. AND H. S. ALPER Expanding the chemical palate of cells by combining systems biology and metabolic engineering. Metab Eng, Jul 2012, 14(4), 289-297.

35. DALBY, A., J. G. NOURSE, W. D. HOUNSHELL, A. K. GUSHURST, et al. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. Journal of chemical information and computer sciences, 1992, 32(3), 244-255.

36. DALE, J. M., L. POPESCU AND P. D. KARP Machine learning methods for metabolic pathway prediction. BMC Bioinformatics, 2010, 11, 15.

37. DE GROOT, M. J., R. J. VAN BERLO, W. A. VAN WINDEN, P. J. VERHEIJEN, et al. Metabolite and reaction inference based on enzyme specificities. Bioinformatics, Nov 15 2009, 25(22), 2975-2982.

38. DREWKE, C. AND M. CIRIACY Overexpression, purification and properties of alcohol dehydrogenase IV from Saccharomyces cerevisiae. Biochim Biophys Acta, May 6 1988, 950(1), 54-60.

39. DROBNER, E., H. HUBER AND K. O. STETTER Thiobacillus ferrooxidans, a facultative hydrogen oxidizer. Appl Environ Microbiol, Sep 1990, 56(9), 2922-2923.

40. ECCLESTON, M. AND D. P. KELLY Oxidation kinetics and chemostat growth kinetics of Thiobacillus ferrooxidans on tetrathionate and thiosulfate. J Bacteriol, Jun 1978, 134(3), 718-727.

41. EDWARDS, C. *Microbiology of extreme environments.* Edtion ed.: Open University Press, 1990. ISBN 0335158927.

42. ELBEHTI, A., G. BRASSEUR AND D. LEMESLE-MEUNIER First evidence for existence of an uphill electron transfer through the bc(1) and NADH-Q oxidoreductase complexes of the acidophilic obligate chemolithotrophic ferrous

ion-oxidizing bacterium Thiobacillus ferrooxidans. J Bacteriol, Jun 2000, 182(12), 3602-3606.

43. ESPARZA, M., J. P. CARDENAS, B. BOWIEN, E. JEDLICKI, et al. Genes and pathways for CO2 fixation in the obligate, chemolithoautotrophic acidophile, Acidithiobacillus ferrooxidans, carbon fixation in A. ferrooxidans. BMC Microbiol, 2010, 10, 229.

44. FALCO, L., C. POGLIANI, G. CURUTCHET AND E. DONATI A comparison of bioleaching of covellite using pure cultures of Acidithiobacillus ferrooxidans and Acidithiobacillus thiooxidans or a mixed culture of Leptospirillum ferrooxidans and Acidithiobacillus thiooxidans. Hydrometallurgy, 2003, 71(1-2), 31-36.

45. FEIST, A. M., C. S. HENRY, J. L. REED, M. KRUMMENACKER, et al. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol, 2007, 3, 121.

46. FEIST, A. M., M. J. HERRGARD, I. THIELE, J. L. REED, et al. Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol, Feb 2009, 7(2), 129-143.

47. FEIST, A. M., J. C. SCHOLTEN, B. O. PALSSON, F. J. BROCKMAN, et al. Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. Mol Syst Biol, 2006, 2, 2006 0004.

48. FEIST, A. M., D. C. ZIELINSKI, J. D. ORTH, J. SCHELLENBERGER, et al. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metab Eng, May 2010, 12(3), 173-186.

49. FERGUSON, S. J. AND W. J. INGLEDEW Energetic problems faced by micro-organisms growing or surviving on parsimonious energy sources and at acidic pH: I. Acidithiobacillus ferrooxidans as a paradigm. Biochim Biophys Acta, Dec 2008, 1777(12), 1471-1479.

50. FISCHER-KOWALSKI, M. AND M. SWILLING *Decoupling: natural resource use and environmental impacts from economic growth.* Edtion ed.: United Nations Environment Programme, 2011. ISBN 928073167X.

51. FISCHER, C. R., D. KLEIN-MARCUSCHAMER AND G. STEPHANOPOULOS Selection and optimization of microbial hosts for biofuels production. Metab Eng, Nov 2008, 10(6), 295-304.

52. FONG, S. S., A. P. BURGARD, C. D. HERRING, E. M. KNIGHT, et al. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnology and bioengineering, Sep 5 2005, 91(5), 643-648.

53. FRANKS, D. M., D. V. BOGER, C. M. CÔTE AND D. R. MULLIGAN Sustainable development principles for the disposal of mining and mineral processing wastes. Resources Policy, 2011, 36(2), 114-122.

54. FUJII, T., K. ITO, S. KATSUMA, R. NAKANO, et al. Molecular and functional characterization of an acetyl-CoA acetyltransferase from the adzuki bean borer moth Ostrinia scapulalis (Lepidoptera: Crambidae). Insect Biochem Mol Biol, Jan 2010, 40(1), 74-78.

55. FURUYOSHI, S., Y. NAWA, N. KAWABATA, H. TANAKA, et al. Purification and characterization of a new NAD(+)-dependent enzyme, L-tartrate decarboxylase, from Pseudomonas sp. group Ve-2. J Biochem, Oct 1991, 110(4), 520-525.

56. GALE, N. L. AND J. V. BECK Evidence for the Calvin cycle and hexose monophosphate pathway in Thiobacillus ferrooxidans. J Bacteriol, Oct 1967, 94(4), 1052-1059.

57. GEHRKE, T., J. TELEGDI, D. THIERRY AND W. SAND Importance of Extracellular Polymeric Substances from Thiobacillus ferrooxidans for Bioleaching. Appl Environ Microbiol, Jul 1 1998, 64(7), 2743-2747.

58. GREENE, N., P. N. JUDSON, J. J. LANGOWSKI AND C. A. MARCHANT Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. SAR QSAR Environ Res, 1999, 10(2-3), 299-314.

59. GULEVICH, A. Y., A. Y. SKOROKHODOVA, A. V. SUKHOZHENKO, R. S. SHAKULOV, et al. Metabolic engineering of Escherichia coli for 1-butanol biosynthesis through the inverted aerobic fatty acid beta-oxidation pathway. Biotechnol Lett, Mar 2012, 34(3), 463-469.

60. HANAHAN, D. AND R. A. WEINBERG Hallmarks of cancer: the next generation. Cell, Mar 4 2011, 144(5), 646-674.

61. HANAI, T., S. ATSUMI AND J. C. LIAO Engineered synthetic pathway for isopropanol production in Escherichia coli. Appl Environ Microbiol, Dec 2007, 73(24), 7814-7818.

62. HARNEIT, K., A. GÖKSEL, D. KOCK, J.-H. KLOCK, et al. Adhesion to metal sulfide surfaces by cells of Acidithiobacillus ferrooxidans, Acidithiobacillus thiooxidans and Leptospirillum ferrooxidans. Hydrometallurgy, 2006, 83(1), 245-254.

63. HATZIMANIKATIS, V., C. LI, J. A. IONITA, C. S. HENRY, et al. Exploring the diversity of complex metabolic networks. Bioinformatics, Apr 15 2005, 21(8), 1603-1609.

64. HEATH, A. P., G. N. BENNETT AND L. E. KAVRAKI Finding metabolic pathways using atom tracking. Bioinformatics, Jun 15 2010, 26(12), 1548-1555.

65. HENRY, C. S., L. J. BROADBELT AND V. HATZIMANIKATIS Thermodynamics-based metabolic flux analysis. Biophys J, Mar 1 2007, 92(5), 1792-1805.

66. HENRY, C. S., L. J. BROADBELT AND V. HATZIMANIKATIS Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. Biotechnol Bioeng, Jun 15 2010, 106(3), 462-473.

67. HOLD, C., B. A. ANDREWS AND J. A. ASENJO A stoichiometric model of Acidithiobacillus ferrooxidans ATCC 23270 for metabolic flux analysis. Biotechnol Bioeng, Apr 1 2009, 102(5), 1448-1459.

68. HOLLIDAY, G. L., C. ANDREINI, J. D. FISCHER, S. A. RAHMAN, et al. MACiE: exploring the diversity of biochemical reactions. Nucleic Acids Res, Jan 2012, 40(Database issue), D783-789.

69. HOU, B. K., L. P. WACKETT AND L. B. ELLIS Microbial pathway prediction: a functional group approach. J Chem Inf Comput Sci, May-Jun 2003, 43(3), 1051-1057.

70. HU, Q. N., Z. DENG, H. HU, D. S. CAO, et al. RxnFinder: biochemical reaction search engines using molecular structures, molecular fragments and reaction similarity. Bioinformatics, Sep 1 2011, 27(17), 2465-2467.

71. HU, Q. N., H. ZHU, X. LI, M. ZHANG, et al. Assignment of EC numbers to enzymatic reactions with reaction difference fingerprints. PLoS One, 2012, 7(12), e52901.

72. HWANG, J. Y., J. PARK, J. H. SEO, M. CHA, et al. Simultaneous synthesis of 2-phenylethanol and L-homophenylalanine using aromatic transaminase with yeast Ehrlich pathway. Biotechnol Bioeng, Apr 1 2009, 102(5), 1323-1329.

73. IBARRA, R. U., J. S. EDWARDS AND B. O. PALSSON Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature, Nov 14 2002, 420(6912), 186-189.

74. INGLEDEW, W. Thiobacillus ferrooxidans. The bioenergetics of an acidophilic chemolithotroph. Biochimica et biophysica acta, 1982, 683(2), 89-117.

75. ISHIGAKI, T., A. NAKANISHI, M. TATEDA, M. IKE, et al. Bioleaching of metal from municipal waste incineration fly ash using a mixed culture of sulfur-oxidizing and iron-oxidizing bacteria. Chemosphere, Aug 2005, 60(8), 1087-1094.

76. ISHIKURA, Y., S. TSUZUKI, O. TAKAHASHI, C. TOKUDA, et al. Recognition site for the side chain of 2-ketoacid substrate in d-lactate dehydrogenase. J Biochem, Dec 2005, 138(6), 741-749.

77. JAMES, C. A., D. WEININGER AND J. DELANY Daylight theory manual. Daylight chemical information systems, 2004, 3951.

78. JANG, Y. S., J. M. PARK, S. CHOI, Y. J. CHOI, et al. Engineering of microorganisms for the production of biofuels and perspectives based on systems metabolic engineering approaches. Biotechnol Adv, Sep-Oct 2012, 30(5), 989-1000.

79. JANKOWSKI, M. D., C. S. HENRY, L. J. BROADBELT AND V. HATZIMANIKATIS Group contribution method for thermodynamic analysis of complex metabolic networks. Biophys J, Aug 2008, 95(3), 1487-1499.

80. JI, X. J., H. HUANG AND P. K. OUYANG Microbial 2,3-butanediol production: a state-of-the-art review. Biotechnol Adv, May-Jun 2011, 29(3), 351-364.

81. JO, J. E., S. MOHAN RAJ, C. RATHNASINGH, E. SELVAKUMAR, et al. Cloning, expression, and characterization of an aldehyde dehydrogenase from Escherichia

coli K-12 that utilizes 3-Hydroxypropionaldehyde as a substrate. Appl Microbiol Biotechnol, Nov 2008, 81(1), 51-60.

82. JOJIMA, T., M. INUI AND H. YUKAWA Production of isopropanol by metabolically engineered Escherichia coli. Applied microbiology and biotechnology, 2008, 77(6), 1219-1224.

83. KAJIURA, H., K. MORI, N. SHIBATA AND T. TORAYA Molecular basis for specificities of reactivating factors for adenosylcobalamin-dependent diol and glycerol dehydratases. FEBS J, Nov 2007, 274(21), 5556-5566.

84. KANAO, T., K. KAMIMURA AND T. SUGIO Identification of a gene encoding a tetrathionate hydrolase in Acidithiobacillus ferrooxidans. J Biotechnol, Oct 15 2007, 132(1), 16-22.

85. KANEHISA, M., S. GOTO, M. HATTORI, K. F. AOKI-KINOSHITA, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res, Jan 1 2006, 34(Database issue), D354-357.

86. KARP, P. D., S. M. PALEY, M. KRUMMENACKER, M. LATENDRESSE, et al. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief Bioinform, Jan 2010, 11(1), 40-79.

87. KEASLING, J. D. Synthetic biology and the development of tools for metabolic engineering. Metab Eng, May 2012, 14(3), 189-195.

88. KIKUMOTO, M., S. NOGAMI, T. KANAO, J. TAKADA, et al. Tetrathionate-forming thiosulfate dehydrogenase from the acidophilic, chemolithoautotrophic bacterium Acidithiobacillus ferrooxidans. Appl Environ Microbiol, Jan 2013, 79(1), 113-120.

89. KIM, T. Y., S. B. SOHN, H. U. KIM AND S. Y. LEE Strategies for systems-level metabolic engineering. Biotechnol J, May 2008, 3(5), 612-623.

90. KING, Z. A. AND A. M. FEIST Optimizing Cofactor Specificity of Oxidoreductase Enzymes for the Generation of Microbial Production Strains—OptSwap. Industrial Biotechnology, 2013, 9(4), 236-246.

91. KOMA, D., H. YAMANAKA, K. MORIYOSHI, T. OHMOTO, et al. Production of aromatic compounds by metabolically engineered Escherichia coli with an expanded shikimate pathway. Appl Environ Microbiol, Sep 2012, 78(17), 6203-6216.

92. KUENEN, J. Growth yields and "maintenance energy requirement" in Thiobacillus species under energy limitation. Archives of Microbiology, 1979, 122(2), 183-188.

93. LAFFEND, L. A., V. NAGARAJAN AND C. E. NAKAMURA. Bioconversion of a fermentable carbon source to 1, 3-propanediol by a single microorganism. In.: U.S. Patent No. 5,686,276, 1997.

94. LAN, E. I. AND J. C. LIAO ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. Proc Natl Acad Sci U S A, Apr 17 2012, 109(16), 6018-6023.

95. LATINO, D. A. AND J. AIRES-DE-SOUSA Assignment of EC numbers to enzymatic reactions with MOLMAP reaction descriptors and random forests. J Chem Inf Model, Jul 2009, 49(7), 1839-1846.

96. LEE, J. W., D. NA, J. M. PARK, J. LEE, et al. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat Chem Biol, Jun 2012, 8(6), 536-546.

97. LEE, S. Y. AND J. H. PARK. Enhanced butanol producing microorganisms and method for preparing butanol using the same. In.: WO Patent WO/2008/072,921, 2008.

98. LERMAN, J. A., D. R. HYDUKE, H. LATIF, V. A. PORTNOY, et al. In silico method for modelling metabolism and gene product expression at genome scale. Nat Commun, 2012, 3, 929.

99. LETOUZE, E., C. MARTINELLI, C. LORIOT, N. BURNICHON, et al. SDH mutations establish a hypermethylator phenotype in paraganglioma. Cancer Cell, Jun 10 2013, 23(6), 739-752.

100. LEVICAN, G., P. BRUSCELLA, M. GUACUNANO, C. INOSTROZA, et al. Characterization of the petI and res operons of Acidithiobacillus ferrooxidans. J Bacteriol, Mar 2002, 184(5), 1498-1501.

101. LEWIS, N. E. AND A. M. ABDEL-HALEEM The evolution of genome-scale models of cancer metabolism. Front Physiol, 2013, 4, 237.

102. LI, X., Y. LIU, G. ZENG, Y. NIU, et al. Direct current stimulation of Thiobacillus ferrooxidans bacterial metabolism in a bioelectrical reactor without cation-specific membrane. Bioresour Technol, Aug 2010, 101(15), 6035-6038.

103. LU, C., P. S. WARD, G. S. KAPOOR, D. ROHLE, et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. Nature, Mar 22 2012a, 483(7390), 474-478.

104. LU, M., S. LEE, B. KIM, C. PARK, et al. Identification of factors regulating Escherichia coli 2,3-butanediol production by continuous culture and metabolic flux analysis. J Microbiol Biotechnol, May 2012b, 22(5), 659-667.

105. LU, S., M. A. EITEMAN AND E. ALTMAN Effect of CO2 on succinate production in dual-phase Escherichia coli fermentations. J Biotechnol, Sep 10 2009, 143(3), 213-223.

106. LUN, D. S., G. ROCKWELL, N. J. GUIDO, M. BAYM, et al. Large-scale identification of genetic design strategies using local search. Mol Syst Biol, 2009, 5, 296.

107. LYNCH, M. D. Compositions and methods for 3-hydroxypropionate bio-production from biomass. In.: US Patent 8,048,624, 2011, 2011.

108. MA, F. AND M. A. HANNA Biodiesel production: a review. Bioresource Technology, 1999, 70(1), 1-15.

109. MAHADEVAN, R. AND C. H. SCHILLING The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng, Oct 2003, 5(4), 264-276.

110. MAYER, H., J. H. KRAUSS, T. URBANIK-SYPNIEWSKA, V. PUVANESARAJAH, et al. Lipid A with 2,3-diamino-2,3-dideoxy-glucose in lipopolysaccharides from slow-growing members of Rhizobiaceae and from "Pseudomonas carboxydovorans". Arch Microbiol, 1989, 151(2), 111-116.

111. MCCLOSKEY, D., J. A. GANGOITI, Z. A. KING, R. K. NAVIAUX, et al. A model-driven quantitative metabolomics analysis of aerobic and anaerobic metabolism in E. coli K-12 MG1655 that is biochemically and thermodynamically consistent. Biotechnol Bioeng, Apr 2014, 111(4), 803-815.

112. MCCLOSKEY, D., B. O. PALSSON AND A. M. FEIST Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. Mol Syst Biol, 2013, 9, 661.

113. MCSHAN, D. C., S. RAO AND I. SHAH PathMiner: predicting metabolic pathways by heuristic search. Bioinformatics, Sep 1 2003, 19(13), 1692-1698.

114. MEDEMA, M. H., R. VAN RAAPHORST, E. TAKANO AND R. BREITLING Computational tools for the synthetic design of biochemical pathways. Nat Rev Microbiol, Mar 2012, 10(3), 191-202.

115. MU, F., C. J. UNKEFER, P. J. UNKEFER AND W. S. HLAVACEK Prediction of metabolic reactions based on atomic and molecular properties of small-molecule compounds. Bioinformatics, Jun 1 2011, 27(11), 1537-1545.

116. MYKYTCZUK, N. C., J. T. TREVORS, G. D. FERRONI AND L. G. LEDUC Cytoplasmic membrane fluidity and fatty acid composition of Acidithiobacillus ferrooxidans in response to pH stress. Extremophiles, Sep 2010, 14(5), 427-441.

117. NAGARAJAN, V. AND C. E. NAKAMURA. Production of 1, 3-propanediol from glycerol by recombinant bacteria expressing recombinant diol dehydratase. In.: US Patent 5,821,092, 1998.

118. NAM, H., M. CAMPODONICO, A. BORDBAR, D. R. HYDUKE, et al. A systems approach to predict oncometabolites via context-specific genome-scale metabolic networks. PLoS Comput Biol, Sep 2014, 10(9), e1003837.

119. NAM, H., N. E. LEWIS, J. A. LERMAN, D. H. LEE, et al. Network context and selection in the evolution to enzyme specificity. Science, Aug 31 2012, 337(6098), 1101-1104.

120. O'BRIEN, E. J., J. A. LERMAN, R. L. CHANG, D. R. HYDUKE, et al. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol, 2013, 9, 693.

121. ORTH, J. D., T. M. CONRAD, J. NA, J. A. LERMAN, et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. Mol Syst Biol, 2011, 7, 535.

122. ORTH, J. D., I. THIELE AND B. Ø. PALSSON What is flux balance analysis? Nature biotechnology, 2010, 28(3), 245-248.

123. OSORIO, H., S. MANGOLD, Y. DENIS, I. NANCUCHEO, et al. Anaerobic sulfur metabolism coupled to dissimilatory iron reduction in the extremophile Acidithiobacillus ferrooxidans. Appl Environ Microbiol, Apr 2013, 79(7), 2172-2181.

124. PALSSON, B. AND K. ZENGLER The challenges of integrating multi-omic data sets. Nat Chem Biol, Nov 2010, 6(11), 787-789.

125. PASTER, M., J. L. PELLEGRINO, T. M. CAROLE, I. ENERGETICS, et al. *Industrial Bioproducts: Today and Tomorrow*. Edtion ed.: Energetics, Incorporated, 2003.

126. PATIL, K. R., I. ROCHA, J. FORSTER AND J. NIELSEN Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics, 2005, 6, 308.

127. PERALTA-YAHYA, P. P., F. ZHANG, S. B. DEL CARDAYRE AND J. D. KEASLING Microbial engineering for the production of advanced biofuels. Nature, Aug 16 2012, 488(7411), 320-328.

128. PERLACK, R. D. AND B. J. STOKES *US billion-ton update: biomass supply for a bioenergy and bioproducts industry*. Edtion ed.: Oak Ridge National Laboratory, 2011.

129. PETERSON, J. D., L. A. UMAYAM, T. DICKINSON, E. K. HICKEY, et al. The Comprehensive Microbial Resource. Nucleic Acids Res, Jan 1 2001, 29(1), 123-125.

130. PHARKYA, P. Microorganisms and methods for the co-production of isopropanol eith primary alcohols, diols and acids. In.: WO Patent WO/2011/031,897, 2011.

131. PHARKYA, P., A. P. BURGARD AND C. D. MARANAS OptStrain: a computational framework for redesign of microbial production systems. Genome Res, Nov 2004, 14(11), 2367-2376.

132. PORTNOY, V. A., D. BEZDAN AND K. ZENGLER Adaptive laboratory evolution--harnessing the power of biology for metabolic engineering. Curr Opin Biotechnol, Aug 2011, 22(4), 590-594.

133. PORTNOY, V. A., M. J. HERRGARD AND B. O. PALSSON Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient Escherichia coli strain. Appl Environ Microbiol, Dec 2008, 74(24), 7561-7569.

134. PRONK, J. T., W. M. MEIJER, W. HAZEU, J. P. VAN DIJKEN, et al. Growth of Thiobacillus ferrooxidans on Formic Acid. Appl Environ Microbiol, Jul 1991, 57(7), 2057-2062.

135. QIU, M.-Q., S.-Y. XIONG, W.-M. ZHANG AND G.-X. WANG A comparison of bioleaching of chalcopyrite using pure culture or a mixed culture. Minerals Engineering, 2005, 18(9), 987-990.

136. QUATRINI, R., C. APPIA-AYME, Y. DENIS, E. JEDLICKI, et al. Extending the models for iron and sulfur oxidation in the extreme acidophile Acidithiobacillus ferrooxidans. BMC Genomics, 2009, 10, 394.

137. RAWLINGS, D. E. Heavy metal mining using microbes 1. Annual Reviews in Microbiology, 2002, 56(1), 65-91.

138. RAWLINGS, D. E., D. DEW AND C. DU PLESSIS Biomineralization of metal-containing ores and concentrates. TRENDS in Biotechnology, 2003, 21(1), 38-44.

139. RAWLINGS, D. E. AND D. B. JOHNSON *Biomining*. Edtion ed.: Springer, 2007. ISBN 3540349111.

140. ROHWERDER, T., T. GEHRKE, K. KINZLER AND W. SAND Bioleaching review part A: progress in bioleaching: fundamentals and mechanisms of bacterial metal sulfide oxidation. Applied microbiology and biotechnology, 2003, 63(3), 239.

141. ROHWERDER, T. AND W. SAND The sulfane sulfur of persulfides is the actual substrate of the sulfur-oxidizing enzymes from Acidithiobacillus and Acidiphilium spp. Microbiology, Jul 2003, 149(Pt 7), 1699-1710.

142. ROHWERDER, T. AND W. SAND Oxidation of inorganic sulfur compounds in acidophilic prokaryotes. Engineering in Life Sciences, 2007, 7(4), 301-309.

143. SAND, W. AND T. GEHRKE Extracellular polymeric substances mediate bioleaching/biocorrosion via interfacial processes involving iron(III) ions and acidophilic bacteria. Res Microbiol, Jan-Feb 2006, 157(1), 49-56.

144. SAND, W., K. ROHDE, B. SOBOTKE AND C. ZENNECK Evaluation of Leptospirillum ferrooxidans for leaching. Applied and Environmental Microbiology, 1992, 58(1), 85-92.

145. SAUER, M., D. PORRO, D. MATTANOVICH AND P. BRANDUARDI Microbial production of organic acids: expanding the markets. Trends Biotechnol, Feb 2008, 26(2), 100-108.

146. SCHEER, M., A. GROTE, A. CHANG, I. SCHOMBURG, et al. BRENDA, the enzyme information system in 2011. Nucleic Acids Res, Jan 2011, 39(Database issue), D670-676.

147. SCHELLENBERGER, J., J. O. PARK, T. M. CONRAD AND B. O. PALSSON BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics, 2010, 11, 213.

148. SCHELLENBERGER, J., R. QUE, R. M. FLEMING, I. THIELE, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc, Sep 2011, 6(9), 1290-1307.

149. SCHOMBURG, I., A. CHANG AND D. SCHOMBURG BRENDA, enzyme data and metabolic information. Nucleic Acids Res, Jan 1 2002, 30(1), 47-49.

150. SEPÚLVEDA, A. E. M., P. A. M. CORTÉZ, M. A. B. ABARCA, P. A. P. VALDECANTOS, et al. Method to increase the production of extracellular polymeric substances (eps) in a acidithiobacillus ferrooxidans culture by the inhibition of enzymes of tricarboxilic acid cycle. In.: Google Patents, 2011.

151. SHEN, C. R., E. I. LAN, Y. DEKISHIMA, A. BAEZ, et al. Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. Appl Environ Microbiol, May 2011, 77(9), 2905-2915.

152. SHEN, C. R. AND J. C. LIAO Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. Metab Eng, Nov 2008, 10(6), 312-320.

153. SHEN, C. R. AND J. C. LIAO Synergy as design principle for metabolic engineering of 1-propanol production in Escherichia coli. Metab Eng, May 2013, 17, 12-22.

154. SHIVELY, J. M. AND A. A. BENSON Phospholipids of Thiobacillus thiooxidans. J Bacteriol, Nov 1967, 94(5), 1679-1683.

155. SILVERMAN, R. B. *The organic chemistry of enzyme-catalyzed reactions*. Edtion ed.: Academic Pr, 2002. ISBN 0126437319.

156. SOUCAILLE, P., S. I. MEYNIAL, F. VOELKER AND R. FIGGE. Microorganisms and methods for production of 1, 2-propanediol and acetol. In.: U.S. Patent Application 12/532,469, 2008.

157. SUBLETTE, K. L. Production of microbial biomass protein from autotrophic fermentation of hydrogen sulfide. Biotechnology and bioengineering, 1988, 32(3), 408-409.

158. SUTHERLAND, P. AND L. MCALISTER-HENN Isolation and expression of the Escherichia coli gene encoding malate dehydrogenase. J Bacteriol, Sep 1985, 163(3), 1074-1079.

159. SUTHERS, P. F. AND D. C. CAMERON. Production of 3-hydroxypropionic acid in recombinant organisms. In.: US Patent 6,852,517, 2005.

160. TANG, X., Y. TAN, H. ZHU, K. ZHAO, et al. Microbial conversion of glycerol to 1,3-propanediol by an engineered strain of Escherichia coli. Appl Environ Microbiol, Mar 2009, 75(6), 1628-1634.

161. TEPPER, N. AND T. SHLOMI Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics, Feb 15 2010, 26(4), 536-543.

162. THIELE, I. AND B. O. PALSSON A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc, Jan 2010, 5(1), 93-121.

163. THIELE, I., N. SWAINSTON, R. M. FLEMING, A. HOPPE, et al. A community-driven global reconstruction of human metabolism. Nat Biotechnol, May 2013, 31(5), 419-425.

164. TRINH, C. T. Elucidating and reprogramming Escherichia coli metabolisms for obligate anaerobic n-butanol and isobutanol production. Appl Microbiol Biotechnol, Aug 2012, 95(4), 1083-1094.

165. TSENG, H. C., C. H. MARTIN, D. R. NIELSEN AND K. L. PRATHER Metabolic engineering of Escherichia coli for enhanced production of (R)- and (S)-3-hydroxybutyrate. Appl Environ Microbiol, May 2009, 75(10), 3137-3145.

166. VALDÉS, J., J. CÁRDENAS, R. QUATRINI, M. ESPARZA, et al. Comparative genomics begins to unravel the ecophysiology of bioleaching. Hydrometallurgy, 2010, 104(3), 471-476.

167. VALDES, J., I. PEDROSO, R. QUATRINI, R. J. DODSON, et al. Acidithiobacillus ferrooxidans metabolism: from genome sequence to industrial applications. BMC Genomics, 2008, 9, 597.

168. VALDES, J., F. VELOSO, E. JEDLICKI AND D. HOLMES Metabolic reconstruction of sulfur assimilation in the extremophile Acidithiobacillus ferrooxidans based on genome analysis. BMC Genomics, Dec 15 2003, 4(1), 51.

169. VALENTIN, H. E. AND D. DENNIS Production of poly(3-hydroxybutyrate-co-4-hydroxybutyrate) in recombinant Escherichia coli grown on glucose. J Biotechnol, Oct 2 1997, 58(1), 33-38.

170. VARMA, A., B. W. BOESCH AND B. O. PALSSON Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. Appl Environ Microbiol, Aug 1993, 59(8), 2465-2473.

171. VICKERS, C. E., D. KLEIN-MARCUSCHAMER AND J. O. KROMER Examining the feasibility of bulk commodity production in Escherichia coli. Biotechnol Lett, Apr 2012, 34(4), 585-596.

172. WANG, Q., C. LIU, M. XIAN, Y. ZHANG, et al. Biosynthetic pathway for poly(3-hydroxypropionate) in recombinant Escherichia coli. J Microbiol, Aug 2012, 50(4), 693-697.

173. WARBURG, O., K. POSENER AND E. NEGELEIN Ueber den stoffwechsel der tumoren. Biochem Z, 1924, 152, 319-344.

174. WERPY, T., G. PETERSEN, A. ADEN, J. BOZELL, et al. Top Value Added Chemicals From Biomass. Volume 1-Results of Screening for Potential Candidates From Sugars and Synthesis Gas. 2004.

175. WISHART, D. S., C. KNOX, A. C. GUO, R. EISNER, et al. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res, Jan 2009, 37(Database issue), D603-610.

176. WOOD, A. P., J. P. AURIKKO AND D. P. KELLY A challenge for 21st century molecular biology and biochemistry: what are the causes of obligate autotrophy and methanotrophy? FEMS Microbiol Rev, Jun 2004, 28(3), 335-352.

177. WYMAN, C. E., S. R. DECKER, M. E. HIMMEL, J. W. BRADY, et al. Hydrolysis of cellulose and hemicellulose. Polysaccharides: Structural Diversity and Functional Versatility, 2005, 995-1033.

178. XIAO, M., H. YANG, W. XU, S. MA, et al. Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in

mutations of FH and SDH tumor suppressors. Genes Dev, Jun 15 2012, 26(12), 1326-1338.

179. XU, W., H. YANG, Y. LIU, Y. YANG, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell, Jan 18 2011, 19(1), 17-30.

180. YAN, Y., C. C. LEE AND J. C. LIAO Enantioselective synthesis of pure (R,R)-2,3-butanediol in Escherichia coli with stereospecific secondary alcohol dehydrogenases. Org Biomol Chem, Oct 7 2009, 7(19), 3914-3917.

181. YANG, F., M. A. HANNA AND R. SUN Value-added uses for crude glycerol--a byproduct of biodiesel production. Biotechnol Biofuels, 2012a, 5, 13.

182. YANG, M., T. SOGA, P. J. POLLARD AND J. ADAM The emerging role of fumarate as an oncometabolite. Front Oncol, 2012b, 2, 85.

183. YARZABAL, A., C. APPIA-AYME, J. RATOUCHNIAK AND V. BONNEFOY Regulation of the expression of the Acidithiobacillus ferrooxidans rus operon encoding two cytochromes c, a cytochrome oxidase and rusticyanin. Microbiology, Jul 2004, 150(Pt 7), 2113-2123.

184. YIM, H., R. HASELBECK, W. NIU, C. PUJOL-BAXLEY, et al. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nat Chem Biol, Jul 2011, 7(7), 445-452.

185. ZENG, A. P. AND W. SABRA Microbial production of diols as platform chemicals: recent progresses. Curr Opin Biotechnol, Dec 2011, 22(6), 749-757.

186. ZHOU, X. Y., X. X. YUAN, Z. Y. SHI, D. C. MENG, et al. Hyperproduction of poly(4-hydroxybutyrate) from glucose by recombinant Escherichia coli. Microb Cell Fact, 2012, 11, 54.

187. ZHUGE, B., C. ZHANG, H. FANG, J. ZHUGE, et al. Expression of 1,3-propanediol oxidoreductase and its isoenzyme in Klebsiella pneumoniae for bioconversion of glycerol into 1,3-propanediol. Appl Microbiol Biotechnol, Aug 2010, 87(6), 2177-2184.