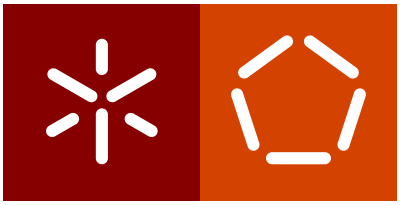
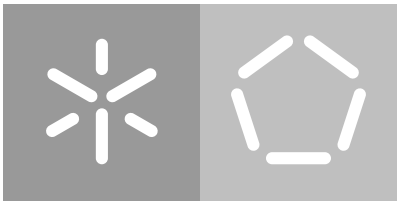


Universidade do Minho
Escola de Engenharia
Departamento de Informática

Rui Ricardo Barros Nunes

Genome-scale metabolic modelling of an extremophile microbial community

February 2022



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Rui Ricardo Barros Nunes

Genome-scale metabolic modelling of an extremophile microbial community

Master dissertation
Master Degree in Bioinformatics

Dissertation supervised by
Oscar Manuel Lima Dias

February 2022

1

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos. Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho

Atribuição
CC BY

https://creativecommons.org/licenses/by/4.0/

ACKNOWLEDGEMENTS

First of all, I would like to say that this work gave me a lot of positive things. I gained a lot of skills in the area as well as it heavily contributed to my personal and professional growth.

I would like to acknowledge professor Oscar Dias, Sophia Santos and Emanuel Cunha for all the help and support. I could not have finished my work without the knowledge they transmitted me.

A special thank you goes to my friends for assisting me on the psychological side. They were a big pillar in maintaining a positive emotional state providing a good environment in which I could do my work.

Lastly, a big thank you to my family who always believed in me and ensured I was alright. I would not have accomplished what I have without their constant support.

2

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

ABSTRACT

Biomining offers an ecological alternative to the standard mining practices by using extremophiles that can endure elevated temperatures and low pH values. Several studies have been performed using *Acidithiobacillus caldus SM-1* and *Acidimicrobium ferrooxidans DSM 10331*, suggesting that these bacteria in a community offer several advantages in bioleaching environments.

Genome-Scale Metabolic (GSM) models simulate the organisms' metabolism through constraint-based approaches. Therefore, the reconstruction of GSM models for *A. caldus* and *A. ferrooxidans* and their integration into a community will offer, besides valuable insights into their metabolism, a unique perspective on the potential interaction mechanisms between the two organisms within the community.

In this work, we developed manually curated GSM models for *A. caldus* with 416 genes, 846 reactions and 646 metabolites, and A. ferrooxidans with 408 genes, 817 reactions and 640 metabolites. Both models were reconstructed using the user-friendly software *merlin*.

We performed the functional annotation of both organisms' genomes to identify their metabolic characteristics, which allowed generating a draft of the metabolic network. Manual curation efforts through literature, genomic information, phylogenetically close organisms and biological databases allowed refining the metabolic network. Furthermore, the models were validated using Cobrapy and Mewpy which allowed analysing flux distribution and interactions in different environmental conditions, and the results were compared with the literature and experimental data. Lastly, the community model was built using the organisms' validated GSM models.

In silico phenotypic simulations of the community model revealed that *A. caldus* exchanged lipid-production related compounds whilst *A. ferrooxidans* donated hydrogen sulfide assisting the former with its more complex sulfur metabolism. Moreover, the results suggest a more significant influence of *A. ferrooxidans* in the community's growth rate whilst *A. caldus* assists *A. ferrooxidans* in biomass production.

These models can serve as a starting point to study and model the community's behaviour in several bioleaching conditions.

Keywords: *Acidithiobacillus caldus SM-1*, *Acidimicrobium ferrooxidans DSM 10331*, Extremophiles, Genome-scale metabolic model, Microbial community, Biomining

RESUMO

A biomineração oferece uma alternativa ecológica às práticas de mineração comuns através do uso de extremófilos que são capazes suportar elevadas temperaturas e baixos valores de pH. Vários estudos foram realizados usando os microrganismos *Acidithiobacillus caldus SM-1* e *Acidimicrobium ferrooxidans DSM 10331* em comunidade, sugerindo várias vantagens em ambientes de biolixiviação.

Os modelos metabólicos à escala genómica permitem a modelação do metabolismo através de abordagens baseadas em restrições. Portanto, a construção de uma comunidade com contendo o modelo da *A. caldus* e outro da *A. ferrooxidans* poderá oferecer novas perspetivas sobre os seus respetivos metabolismos, assim como sobre os mecanismos de interação entre os dois organismos dentro da comunidade.

Neste trabalho, foram reconstruídos dois modelos metabólicos à escala genómica com um elevado nível de curação manual utilizando a ferramenta *merlin*. O modelo da *A. caldus* conta com 416 genes e 846 reações enquanto que o da *A. ferrooxidans* possuí 408 genes e 817 reações.

Os modelos foram funcionalmente anotados a fim de identificar as características metabólicas dos organismos, gerando um esboço da rede metabólica. Esta rede metabólica foi depois curada manualmente a fim de a refinar. Para isto foi usado informação presente na literatura, dados genómicos, organismos filogeneticamente próximos e bases de dados biológicas. Posteriormente, os modelos foram validados através de uma análise da distribuição de fluxo com diferentes condições ambientais e os resultados foram comparados com a literatura e dados experimentais. Por fim, o modelo da comunidade foi construído usando os modelos validados dos dois organismos.

Simulações fenotípicas *in silico* do modelo da comunidade revelaram uma troca de compostos relacionados com produção de lípidios por parte da *A. caldus*, enquanto *A. ferrooxidans* doou sulfato de hidrogénio, auxiliando o primeiro no seu metabolismo de enxofre mais complexo. Por fim, os resultados sugerem uma maior influência de *A. ferrooxidans* na taxa de crescimento da comunidade enquanto que *A. caldus* auxilia o primeiro na produção de biomassa.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

A

ADP Adenosine Pyrophosphate.

ATP Adenosine Triphosphate.

AUREME Automatic Reconstruction of Metabolic Models.

B

BIGG Biochemical, Genetic and Genomic.

BIOISO Biological networks In Silico Optimization.

BLAST Basic Local Alignment Search Tool.

BRENDA BRaunschweig Enzyme Database.

C

CMM Community Metabolic Models.

COBRA COnstraint-Based Reconstruction and Analysis.

D

DNA Deoxyribonucleic acid.

E

EAS Evolutionary Algorithms.

EBI EMBL Nucleotide Database.

EC Enzyme Comission.

EFMS Elementary Flux Modes.

EPS Exopolysaccharides.

F

FBA Flux Balance Analysis.

FRAMED FRAmework for Metabolic Engineering and Design.

FVA Flux Variability Analysis.

G

GOLD Genomes Online Database.

GPR gene-protein reaction.

GSM Genome-Scale Metabolic.

I

ID identifiers.

K

KEGG Kyoto Encyclopedia of Genes and Genomes.

L

LPS lipopolysaccharide.

M

MERLIN MEtabolic model Reconstruction using genome scaLe INformation.

MFA Metabolic Flux Analysis.

MILP mixed-integer linear programming.

MOMA Minimization of Metabolic Adjustment.

N

NCBI National Center for Biotechnology Information.

NGS Next-Generation Sequencing.

P

PFBA parsimonious Flux Balance Analysis.

PGDB Pathway/ Genome Databases.

PMF Proton Motive Force.

PRPP 5-Phospho-alpha-D-ribose 1-diphosphate.

R

RAST Rapid Annotation of microbial genomes using Subsystems Technology.

RAVEN Reconstruction, Analysis and Visualization of Metabolic Networks.

RISCS reduced inorganic sulfur compounds.

RNA Ribonucleic acid.

ROOM Regulatory on/off minimization of metabolic flux changes.

S

SA Simulated Annealing.

SBML Systems Biology Markup Language.

T

TC Transport Classification.

TCDB Transporter Classification Database.

TRANSYT Transport Systems Tracker.

TREMBL Translated EMBL Database.

U

UNIPROT Universal Protein Resource.

3

INTRODUCTION

3.1 CONTEXT AND MOTIVATION

The huge amount of data that *Next-Generation Sequencing (NGS)* technologies can provide, opened a whole new set of opportunities in the fields of systems biology as well as in metabolic engineering. The data retrieved from NGS provides the reconstruction of GSM models [3], delivering new insights into the metabolism of the organism of interest, as well as its functioning. A model can be seen as a guide to the wet lab experiments due to its faster and cheaper phenotype predictions [4].

GSM models are bioinformatic tools frequently used for the identification of potential drug targets [5] as well as over/under production of compounds of interest [6, 7]. The complete procedure for the development of GSM models was firstly described by Thiele and Palsson, in 2010, and involves four major steps: Draft reconstruction, Refinement of reconstruction, Conversion of reconstruction into a computable format and Network evaluation [8]. Although several tools have been developed to automate several steps of the reconstruction, such as *merlin* [3], revision of literature is a constant necessity throughout the reconstruction.

The study of extremophiles, microorganisms that live in extreme environments, can provide several benefits benefits through their extremozymes [9] as well as their bioactive compounds [10]. In 2015, Giddings and Newman described the importance of extremophile's bioactive compounds in the discovery of new potential drugs. Moreover, the extremozymes found in extremophiles, present the same function as their non-extreme counterparts, with the advantage of bestowing the capability to perform their function under extreme environments without denaturing [11]. This provides several benefits for the industrial area since biocatalysts are widely used and most procedure conditions are severe [12]. In 2018, Oyama et al., suggested the use of the microorganisms *Acidithiobacillus caldus SM-1* and *Acidimicrobium ferrooxidans DSM 10331* in order to optimize the process of biomining. Biomining offers an alternative process through the use of extremophiles that are able to oxidize and reduce sulfur-based compounds and/or iron as energy source [13]. The use of these microorganisms instead of the traditional process offers several advantages since it has a lower energy requirement and does not generate harmful gaseous emissions [14].

A. caldus SM-1 is a moderately thermoacidophilic, obligately chemolithotrophic, gram-negative bacteria with an optimal growth temperature around 40-45 degrees Celsius and pH around 2-2.5. It is the most frequently used bacteria in the biomining process due to its ability to oxidize elemental sulfur and reduced inorganic sulfur compounds producing the acidity that is essential for biomining. It is also used for the removal of accumulated sulfur that would otherwise retard the oxidation of ores [15]. *A. ferrooxidans DSM 10331* is a moderately thermophilic, ferrous-iron-oxidizing acidophile gram-positive bacteria with an optimal growth temperature around 48 degrees Celsius. This organism has the ability to oxidize ferrous iron to ferric iron in order to produce ATP [16].

Since that in a natural environment, bacteria are generally found to live in communities, it is imperative to comprehend the interactions between them to understand how a single bacterium works in the presence of others and how to manipulate one inducing changes in the other. Thus, the reconstruction of a GSM model for each organism in the community would provide, as it does to single organisms, precious insights into the mechanism of each one and the interactions between them.

3.2 OBJECTIVES

The main goal of this work is to build GSM models for the bacteria *A. caldus SM-1* and *A. ferrooxidans DSM 10331*. From this work, we expect to add knowledge to extremophiles metabolism and interactions providing a guide for future experiments on these two organisms. In order to achieve this, the following goals need to be fulfilled:

- Obtain a high-quality functional annotation of both organism genomes;
- Generate a draft for both metabolic networks;
- Perform refinement and manual curation of the network using information from literature;
- Convert the metabolic network into a stoichiometric model;
- Validate the reconstructed models using experimental data;
- Analyze both organisms and community metabolic networks to find potential compound or proteins with biological interest;

3.2.1 *Document organization*

Chapter 2, 'State of the art', contains literature information regarding the subject areas addressed in this work. It starts with a systems biology overview followed by a detailed description of a GSM model reconstruction as well as a summary of other studies performed

with reconstructed GSM models and computational tools frequently used in this matter. The chapter ends with an overview of extremophiles as well as of the bacteria present in this study.

Chapter 3, 'Materials and methods', contains a detailed description of the development of this work. This chapter starts by explaining the efforts made in genome functional annotation followed by the generation of the models' draft metabolic network. Next, it is explained the formulation of the organisms' biomass equations as well as the manual curation performed on the models' metabolic network. Lastly, the efforts made into the models' validation is explained as well as the procedure for the community model construction.

Chapter 4, 'Results and discussion', covers the results of this work and their respective discussion. It starts with the results of the genomes' annotation followed by an overview of the models' metabolic network. Moreover, a detailed description of the organisms' biomass equations is given and the results obtained from the models' validation is provided. Lastly, the results obtained from the simulations of the community model are presented.

Chapter 5, 'Conclusions and future work', contains the conclusions taken from this work as well as a few perspectives on possible future work that can be done with these models.

4

STATE OF THE ART

4.1 SYSTEMS BIOLOGY

The emergence of the field of systems biology comes through a necessity of processing the exponential growth of omics data being generated and stored correlated with numerous applications of great benefit. This field is focused on analyzing the different levels of components and their interactions from a single molecule to the whole organism [17, 18] in order to comprehend their organization and to predict its behavior [19, 20] through the construction of mathematical models. This methodology is more easily applied to the prokaryotic organisms due to a simpler compartmentalization and fewer genomic information. Nonetheless, eukaryotic models are of great relevance (e. g., protein post-translation mechanisms) with the increasing number of eukaryotic models being made in recent years [21].

The introduction of whole-genome high-throughput sequencing techniques promoted the completion of several sequencing projects generating large-scale data sets [22] allowing more complex analyses opening a whole new set of opportunities in the fields of systems biology as well as in metabolic engineering.

The data retrieved from NGS enables the reconstruction of GSM models [3], delivering new insights into the metabolism of the organism of interest as well as its mechanisms under several environmental and genetic conditions. Thus, a model can be seen as a guide to wet lab experiments due to its faster and cheaper phenotype predictions [4]. The knowledge retrieved from GSM models can be of great interest for several fields ranging from industrial purposes with the creation of cell factories [23] to medical use through drug targeting studies [24].

4.2 GENOME-SCALE METABOLIC MODELS

Whole-genome sequencing and annotation alongside with stored biochemical data from several biological databases [25, 26] made the reconstruction of GSM models possible. Preferably, these models would integrate different levels of information, ranging from reactions stoichiometry to reactions kinetics and regulatory information. However, there is still a massive gap in knowledge regarding kinetic and regulatory data since this dynamic information only exists

for part of some well-studied organisms [27, 28]. Nonetheless, it is still possible to predict several features of the metabolic systems through steady-state analysis e. g., growth rate under a variety of environmental and genetic conditions as well as prediction of essential genes.

Currently, prokaryotic models are more abundant when compared to eukaryotic models. In fact, eukaryotes have a higher quantity of genomic information leading to a more specialized compartmentalization for each cell as well as a different type of tissues in higher eukaryotes, making these models more demanding [8].

GSM models are valuable bioinformatics tools enabling the *in silico* simulation of the organism or cell phenotype behaviour. For this, it is necessary to retrieve information for each reaction, such as the substrates and products, its stoichiometry, reversibility, and location [29].

The complete procedure for the development of GSM models was firstly described by Thiele and Palsson, in 2010, and involves four major steps: Draft reconstruction, Refinement of reconstruction, Conversion of reconstruction into computable format and Network evaluation [8].

Throughout the reconstruction, a GSM model heavily depends on the information available in several biological databases regarding genome sequences and annotation and/or the functional capabilities of the proteins [30]. This genomic information is used to create a draft metabolic network of the intended organism or cell, which, after several evaluations is converted into a GSM model alongside with the addition of an objective function, energy requirements and several other constraints. One of the most used objective functions is the biomass equation that accounts with all the building blocks needed to construct a single organism or cell. The biomass equation is not generated in the draft metabolic network since it is not generated by the genome but rather by the necessity of each cell.

Finally, the model should be exported to a universally accepted format, such as *Systems Biology Markup Language (SBML)* file in order to be tested through several simulation methods. These simulations are intended to test the predictive behaviour of the model, hence the coherence of the model is checked by comparing the *in silico* results to available experimental data. In this process, several steps of the reconstruction are reviewed in an iterative manner until the GSM model is completed and ready to use for applications such as the identification of potential drug targets [5] as well as over/under production of compounds of interest [6, 7]. Although several tools have been developed to automate several steps of the reconstruction, such as merlin [3], revision of literature is a constant necessity if the goal is a GSM model with good predictive capabilities.

Moreover, biological databases assist the user throughout the reconstruction of the model making them an essential tool in systems biology. Throughout the next chapters it is specified the support of the databases in the several steps of the reconstruction. In table 1 are listed some useful databases for the reconstruction of a GSM model.

Table 1: Relevant databases for the reconstruction of a GSM model

Database	Description	Web address	Reference
EMBL Nucleotide Database (EBI)	Bioinformatics sequence analysis tools as well as full-featured text search engine with cross-referencing and data retrieval capabilities.	https://www.ebi.ac.uk/ena	[31]
National Center for Biotechnology Information (NCBI)	Genomic and biochemical data containing several databases with information regarding whole genome sequences and annotation, protein functions and domains, taxonomy, literature and chemical properties.	https://www.ncbi.nlm.nih.gov/	[32]
BioCyc	Collection of *Pathway/ Genome Databases (PGDB)* as well as software tools to explore the data.	https://biocyc.org/	[33]
BKM-react	Biochemical reaction database containing known enzyme-catalyzed and spontaneous reactions.	http://bkm-react.tu-bs.de/	[34]
BRaunschweig Enzyme Database (BRENDA)	Biochemical and molecular information of ezymes. EC system.	https://www.brenda-enzymes.org/	[35]
ExPASy Bioformatics Resources Portal	Provides access to scientific databases and software tools in proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics among others.	https://www.expasy.org/	[36]
KEGG	Information regarding genes, metabolites, reactions, and pathways being able to build interaction networks.	https://www.genome.jp/kegg/	[37]
MetaCyc	Curated database of experimentally tested metabolic pathways.	https://metacyc.org/	[38]
SABIO-RK	Curated database of biochemical reactions, their kinetic rate equations with parameters and experimental conditions.	http://sabio.villa-bosch.de/	[39]
Transporter Classification Database (TCDB)	Transporter proteins functional information. *Transport Classification (TC)* system.	http://www.tcdb.org/	[40]
Universal Protein Resource (UniProt)	Protein sequence functional information.	https://www.uniprot.org/	[41]
Biochemical, Genetic and Genomic (BiGG) Models	GSM network reconstructions with standardized identifiers.	http://bigg.ucsd.edu/	[42]
BioModels	Curated mathematical models of biological systems.	https://www.ebi.ac.uk/biomodels/	[43]
ModelSEED	Provides access to tools for the reconstruction, exploration, comparison, and analysis of metabolic models as well as to biochemical reactions and genome annotations.	http://modelseed.org/	[44]

4.2.1 *Genome annotation*

The genome annotation is a crucial task since the whole model depends on the correct assignment of gene functional annotation. This step starts with data analysis obtained from the genome annotation which can be accessed in several public genomic databases such as *Genomes Online Database (GOLD)* and NCBI. It is essential to retrieve relevant annotated data such as gene names and functions, unique identifiers, EC [45] and TC [46] numbers which can be retrieved from KEGG as well as BRENDA and TCDB, respectively. It is also important to identify the subunits of the protein complexes due to the possibility of more than one gene being required to encode for one enzyme.

Furthermore, to perform a genome (re)annotation of the intended organism's genome sequence using tools such as *Basic Local Alignment Search Tool (BLAST)* [47] and HMMER [48] is required, due to the importance of high reliability in this step. These tools operate based on sequence similarity retrieving a similarity score which represents a confidence level of a given gene function assignment. The assignment of this gene function in the model will depend on this confidence score. *merlin* uses these tools against the NCBI and EBI databases and compares the results with the uniprot database in order to achieve a curated annotation.

In metabolic models, only the genes encoding enzymes and membrane transport proteins and metabolic genes will be included and corresponding metabolic/transport reactions will be added. Metabolic genes with a low annotation score should be manually reviewed to assign a putative function. However, one should be careful in doing so since the miss-attribution of the function can lead to gaps in the network or even the inclusion of reactions and /or metabolites inexistent in the given organism.

BRENDA, as well as UniProt, are usually used to confirm the genes annotation, reducing several possible errors in the assignment of the EC number. Also, databases such as KEGG, ExPASy, BioCyc, and MetaCyc, (table 1) can be used since they provide valuable information regarding the set of reactions deducted from the organism's genome sequence.

4.2.2 *Metabolic network assembly*

In this step, the draft metabolic network is assembled. It starts by building the reaction set catalyzed by the metabolic genes (genes that encode for enzymes and transport proteins).

4.2.2.1 *Assembly of the reaction set*

To build a draft metabolic network, the assembly of the reaction set is necessary. The metabolic genes need to be identified and the respective metabolic reactions can be connected with the draft reconstruction by associating the EC and TC numbers to the respective substrates and products. This information can be retrieved from several pathway databases such as KEGG,

ExPASy, TCDB and BRENDA, mentioned in table 1, by using the respective identifiers to which each reaction is associated [49]. At the end of this stage, the user should have, the name of the reactions and genes as well as the respective reactants and products. Reaction directionality and stoichiometry as well as spontaneous reactions, GPR associations and compartmentalization will also be added to the reaction set being further discussed in the next few chapters.

Moreover, the EC nomenclature system looks to categorize all enzymes accordingly to their functions [45]. For this, a four number code was created in order to standardize the classification, whereas the first number classifies the enzyme into one of the six major classes (oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases). As the number goes further to the right the classification increases its specificity.

Analogously, the TC has the objective of creating a universal classification system for membrane transport proteins [46] based on functional and phylogenetic information. The standard code is represented by four numbers and one letter (second digit) with the first number representing one of the seven major transport proteins classes (channels and pores, electrochemical potential-driven transporters, primary active transporters, group translocators, transmembrane electron carriers, accessory factors involved in transport and incompletely characterized transport systems). In the same way as the EC classification system, the further to the right the number goes the more specific the classification is.

It is worth noting that many false positives can be present in this first list of reactions set. Usually, proteins related to *Deoxyribonucleic acid (DNA)* methylation or rRNA modification as well as enzymes involved in the acid nucleic metabolism and signal transfer are given an EC number, however, these reactions are not included in the metabolic model [8]. Also, reactions catalyzed by enzymes without EC numbers assigned, such as transport and exchange reactions and missing reactions known to exist in an organism, need to be further added to the reactions set of the metabolic network. In order to add these reactions and to validate the metabolic network, a careful examination of literature is required. Moreover, an enzyme can be related to several reactions being important to include in the model only the reactions associated with the intended organism.

4.2.2.2 *Spontaneous reactions and Reaction stoichiometry*

The next step in the reconstruction is the addition of nonenzymatic and spontaneous reactions to the metabolic network. Notably, only the spontaneous reactions that have at least one metabolite linking them to the rest of the network should be added, in an effort to avoid too many dead-end metabolites [26]. Commonly, associating these spontaneous reactions to an artificial gene and protein facilitates the analysis of reaction and gene essentiality studies. This artificial GPR association also facilitates the differentiation between spontaneous reactions and reactions without a known gene [8]. Moreover, spontaneous reactions can be retrieved

from literature or several pathway databases such as KEGG. Several tools such as *merlin* and MicrobesFlux [50] retrieve these reactions automatically from the database mentioned above.

In table 1 are listed some useful databases for the reconstruction of a GSM model. At the end of this step, every reaction should have an equal charge and number of atoms on both sides of the reaction (balanced equations). The inability to perform this step correctly may lead to the synthesis of protons or energy (*Adenosine Triphosphate (ATP)*) out of nothing [8]. Furthermore, the cofactor utilization in reaction stoichiometry may be related to an important issue. When is not known if certain enzymes NADH or NADPH, or even both, for many metabolic reactions is common to add the same reaction twice for the use of NADH and NADPH. This may raise a problem since the presence of reactions that accept both cofactors lead to an increase in net transhydrogenation reactions which is not very likely to occur under normal conditions. A possible solution for this is to discard some of these reactions from the network or to consider them irreversible [26].

Moreover, it is also important to analyze the reaction directionality, which can be modelled through the reaction lower and upper bounds. In order to accomplish this, several approaches can be taken:

- revision of literature for biochemical data of the intended organism as well as published models stored in databases such as BiGG and ModelSEED (table 1). Also, biochemical databases such as BRENDA and KEGG (table 1) can present this information. However, one should be careful since these databases present the same directionality for a specific reaction regardless of the organism.
- Estimation of standard Gibbs free energy of formation ($\Delta fG'^{o}$) and reaction ($\Delta rG'^{o}$) in a biochemical system [8]. These values can be obtained from KEGG.
- Usually, reactions comprising the transfer of phosphate from ATP to an acceptor are irreversible except for the ATP synthetase.
- Commonly, reactions with quinones are irreversible.

The miss-assigned direction can have a significant impact on the reliability of the model. If either no information is available, and none of the rules applies, one should carefully leave the reaction reversible [8].

4.2.2.3 *Gene Protein Reactions associations*

Special attention to cases where the rule one-gene-to-one-enzyme-to-one-reaction does not apply is needed.

Promiscuous enzymes are cases where one gene encodes one enzyme which can catalyse several reactions, thus associating one gene to several reactions. In these cases, the EC number is associated with one gene and several reactions *identifiers (ID)*.

Isoenzymes are cases where more than one gene encodes for the same enzyme, catalyzing the same reaction or set of reactions. In the metabolic network, in such cases, reactions are associated with multiple genes with the Boolean rule OR meaning that an enzyme can independently catalyse the reaction.

Lastly, in the presence of an enzymatic complex one or more genes encode for one subunit of the complex with these genes being associated with one or more reactions. Usually, all the subunits need to be present in order to catalyze the reaction specifying the Boolean rule AND, however, this case is not always valid. In some cases, is possible that in two isoenzymes one of them is an enzymatic complex. Such cases should be treated as $isoenzyme_1$ OR ($isoenzyme_{2a}$ AND $isoenzyme_{2b}$) [8].

4.2.2.4 *Compartmentation and localization*

In this step, a higher level of complexity is added to the model due to the creation of cell compartments defining the organelle(s) in which an enzyme operates. Additionally, intracellular transport reactions should be carefully added since there is not much genomic information regarding these reactions and too many intracellular transport reactions can originate futile cycles.

In prokaryotes, the compartments available are often cytosol, periplasmic space, and extracellular space. For lower eukaryotes, this division is not so simple, since there is a higher complexity associated with the organism. For these organisms, several cell compartments can be created, namely, Golgi apparatus, lysosome, mitochondrion, endoplasmic reticulum, or glyoxysome whereas for higher eukaryotes, it is even necessary to take into account multi-tissue interaction [51]. Moreover, the exchange reactions (drains) between the extracellular space and the interior space (periplasmic space in gram-negative bacteria, cytosol in all other cases) are considered inputs or outputs of the system [26].

The compartmentalization of the model allows the user to differentiate similar reactions with the same metabolites that occur in distinct compartments. These metabolites are distinguished by the addition of a compartment identifier to its name reflecting the metabolite localization. In this way, it is possible to process one metabolite in different compartments as distinct metabolites are important, especially for the metabolites with no transporters associated with them and in which diffusion does not usually occur.

Information regarding compartmentalization can be found in several online databases such as UniProt as well as in literature data. Moreover, enzyme localization can be predicted through its amino acid sequence and physiologic data of the organism by using software tools such as TargetP [52] or PSort [53]. Commonly, when no relevant information is found for an enzyme localization, the enzyme is assigned to the cytosol.

4.2.2.5 *Manual curation*

After the reconstruction of the draft network, it is normal to find gaps within the metabolic network. These gaps can occur due to a missing reaction in the pathway leading to the accumulation of undesired metabolites. Likewise, the rest of the pathway will not occur due to the missing substrate, which would be produced by the absent enzyme(s). Reactions without any gene associated might reveal these gaps in the model [49]. In this step, an extensive revision of literature of the organism of study is necessary, or, in case of information unavailability, revision of literature of phylogenetic close organisms. Also, databases such as KEGG should be used since it can present manually drawn pathway maps with molecular interaction and reaction networks [49].

In order to standardize this refinement and to increase its efficiency, a pathway-by-pathway analysis should be done. Thus, one should start from the main pathways leading to the secondary pathways and end with reactions with no pathways associated with them, making it easier to identify network gaps [8].

Each entry of the reaction set should be revised and reactions with no literature evidence should be discarded from the model. Moreover, not every reaction associated with an enzyme occurs in the intended organism being necessary to confirm which reactions are known to take place in the organism [26]. Furthermore, since metabolites may be protonated or deprotonated within a cell, one should determine the charged formula for each depending on its location. Also, the assignment of ambiguous identifiers such as incomplete EC numbers as well as generic terms, can miss-assign a reaction or even discard one from the network.

At the end of this stage, one should have a debugged GSM metabolic network ready to be converted into a mathematical model. It is worth noting that since the reconstruction of the GSM model is an iterative process, the metabolic network obtained at this stage may need further modification depending on the results and inconsistencies obtained in the next stages.

4.2.3 *Conversion from metabolic to a mathematical model*

In this stage, the reaction set is converted to a mathematical model through, firstly the addition of objective function (usually biomass), energy requirements and other constraints and later the conversion of the metabolic network to a stoichiometric matrix.

4.2.3.1 *Biomass and Energy requirements*

Before converting the metabolic network to a mathematical model, it is necessary to add the biomass equation to the reaction set. This biomass equation accounts for every biomass component and their relative contribution to the overall biomass composition necessary for cellular growth. This equation can either be constituted of building blocks (e.g., amino acids and nucleotides) or macromolecules with drains of building blocks or macromolecules guiding

the flux to the biomass equation being created. In the case of the biomass equations being constituted of macromolecules, a set of reactions for the assembly of these from the building blocks should be created. For both cases, the fractional amount of each biomass component needs to be specified depending on the intended organism [26]. These coefficients represent an individual impact on the overall cellular biomass composition. Therefore, the formation of the biomass equation can be represented as follows:

$$\sum_{k=1}^{p} C_k.X_k \longrightarrow Biomass \tag{1}$$

Where:

- C_k represents the coefficient of a biomass component (metabolite or macromolecule)
- X_k represents a component of the biomass

The flux through this reaction represents the specific growth of the intended organism being measured in grams of biomass per time unit (h^{-1}) since all biomass precursors are converted to mmol gDW^{-1}. Thus, the biomass equation sums the moles of each component necessary to produce 1 g dry weight of cells. The determination of the coefficients for each biomass component should be experimentally determined or retrieved from the physiologic and biochemical data of the target organism. If such data cannot be found, data from a phylogenetically close organism can be used. However, one should be careful since studies have shown that the slightest change in the biomass composition can induce substantial changes in the model prediction ability [54]. When no experimental data is available the relative contribution of amino acids and nucleic acids can be estimated from the genomic and/or transcriptomic data [54] and an approach to estimate cofactors in prokaryotes has also been proposed [55]. In the case of lipids and carbohydrates, a thorough revision of literature is necessary.

The biomass equation should account for growth-associated energy requirements in the form of molecules of ATP necessary for 1 gram of biomass synthesized. These energy requirements are represented as the hydrolysis of ATP into *Adenosine Pyrophosphate (ADP)* and orthophosphate being related to maintenance of the membrane potential, turnover of macromolecules, and polymerization of amino acids and nucleotides required for cell replication [26]. The values for these requirements are retrieved from chemostat growth experiments. Alternatively, one can either determine these energy requirements by fitting the model simulation results to experimental data on growth yields [26] or by determining the energy required for macromolecular synthesis [8].

Moreover, non-growth associated energy requirements should also be taken into account. These requirements are associated with cell maintenance processes, such as membrane potential and turgor pressure [56]. In order to incorporate this into the model, one should add an irreversible reaction to the reaction set, converting ATP into ADP and inorganic phosphate.

The coefficients for this reaction or flux rate can either be retrieved from the literature or experimentally determined.

Finally, it is necessary to specify the ratio between ATP synthesis and oxygen consumption for the electron transport chain. In detailed models, this ratio can be calculated through the association between electron transport, proton translocations, respiration and protons usage in transport [57]. However, in less detailed models, one should manually define this ratio by fitting the model to overall growth yields [58].

With the biomass equation defined and both the growth-associated and non-growth associated energy requirements accounted for, the reaction set is ready to be represented as a stoichiometric matrix. For this, the principles of chemical engineering are applied to represent the dynamic mass balance of a metabolite through a single equation where the variation of metabolite concentration throughout time is calculated. However, due to the lack of information regarding kinetic data, an approximation to the steady-state is required. A network with M metabolites and N reactions can be represented as:

$$\sum_{j=1}^{N} S_i.v_j = 0, \quad i = 1, ...M \tag{2}$$

Where:

- v_j represents the rate of a reaction j
- S_{ij} represents the stoichiometric coefficient of a metabolite i in a reaction j
- i represents the concentration of a metabolite
- j represents each reaction

Note that, in a steady-state environment, the concentrations of the metabolites do not change throughout time due to the reduction of mass balance of each metabolite to a set of linear equations [26]. Thus, the rates of consumption of metabolites must be equal to the rates of production being represented in a network with M metabolites and N reactions as:

$$S.v = 0 \tag{3}$$

Where:

- v is the flux vector
- S is the stoichiometric matrix (M x N) where each reaction and metabolites are represented in the columns and rows, respectively whereas the stoichiometric coefficients are represented in each cell

Besides the internal fluxes of the cell, v also accounts for the exchanges fluxes thus, representing metabolite transportation through the cell membrane. Commonly, the number of fluxes

is greater than the number of mass balance constraints (undetermined system), resulting in the space of possible solutions that may satisfy the mass balance constraints, the null space of S [26]. However, with the implementation of additional constraints, the null space of S is reduced to a set of feasible solutions, the flux cone of solutions [30].

Constraints added to the model will reduce the feasible flux space of solutions. The reversibility of reactions is represented through these constraints by imposing a flux range allowed for each reaction. Therefore, the implementation of lower and upper bounds as model constraints is necessary being represented as inequalities:

$$\alpha_j \leq v_j \leq \beta_j, \quad j = 1, ...N \tag{4}$$

Where:

- v_j represents the flux vector
- α_j represents the lower bound
- β_j represents the upper bound

In practical terms, reversible reactions can be constrained between -1000 and 1000 representing minus and plus infinity, respectively. Whereas in irreversible reactions, one of the bounds is set to zero depending on the directionality of the reaction. If the value for the maximum and/or minimum flux of a specific reaction is known, one should constrain with the respective value. The exchange fluxes are likewise constrained, depending on the nutrients available in the medium with the lower bound being in accordance with experimental data of maximum specific uptake rate or to data from limiting substrate studies [26].

In growth-limiting substrate conditions, one should limit the uptake rate by assignment of a specific value to the lower bound of the intended exchange reaction. In the eventuality of a non-existence metabolite in the medium, the corresponding transport fluxes should be constrained to zero [8]. For maximum production of a metabolite, constraints for the upper bounds are necessary. Depending on the environmental conditions being tested, other constraints must be added.

Finally, the stoichiometric representation of the model should be saved in the standard format, , for example SBML, in order to be further used in simulations. These simulations can be made in specialized tools for this purpose, such as *COnstraint-Based Reconstruction and Analysis (COBRA)* toolbox [59] or OptFlux [60].

4.2.4 *Metabolic model evaluation and simulation methods*

With the mathematical model constructed and the constraints applied to it, behaviour predictions, for each environmental condition created, are now possible.

Flux Balance Analysis (FBA) is a frequently used approach to evaluate the predictive behaviour of the model, which is able to compute a set of solutions out of the flux cone [61]. This problem of linear programming calculates an optimal flux distribution for a specified objective function (linear combination of fluxes) in a steady-state system. A set of constraints are given defining the feasible flux space, which will be further constrained with environmental specific conditions. The retrieved result is a particular flux distribution from the edge of the feasible flux space. However, it is possible to attain several optimal solutions through different flux distributions for the same objective function. For such cases, one should use the simulation method *parsimonious Flux Balance Analysis (pFBA)* which will resort to the optimal space of solutions computed by the FBA, retrieving the solution that minimizes the sum of all the fluxes. Several tools have been developed to resolve the optimization problem of FBA as described in later sections. Therefore, the linear programming problem in FBA can be represented as:

$$\begin{aligned} &\text{Maximize} && Z \\ &\text{subject to} && S \cdot v = 0 \\ & && \alpha_j \leq \mathrm{v}_j \leq \beta_j, \quad j = 1, ..., N \end{aligned} \tag{5}$$

Where:

- Z is the linear objective function
- v is the flux vector
- S is the stoichiometric matrix
- α_j and β_j is the lower and upper bound, respectively

The linear objective function (Z) can be to maximize or minimize either the production of biomass or any other metabolite. Commonly, the maximization of biomass production as an objective function for different growth conditions is used, being further compared to experimental data of specific growth rate and growth yields studies [62]. Usually, an organism tends to maximize growth and biomass formation when in the presence of a carbon limiting medium [63], hence one should constrain various reactions in order to test several physiological responses of the organism. If necessary, ATP requirements can be changed accordingly to experimental data, in order to reduce possible inconsistencies between *in silico* and *in vivo* results.

Moreover, active/inactive pathways in a variety of environmental conditions (i.e., aerobic/anaerobic growth) should be analyzed and compared to experimental data. In case of any discrepancy, the model will need further curation of the reaction set. Analysis of dead-end pathways should assist the debugging at this stage. Also, the predictive behaviour of the model in a deletion mutant scenario should be in accordance with experimental data. If these do not match, a review of the genome annotation will be necessary [26].

If the predictive behaviour of the model is considered realistic in the evaluation tests other simulation methods can be performed depending on the application of the model to the user. In order to predict exchange fluxes, one can perform *Metabolic Flux Analysis (MFA)* [64] which characterizes by turning an undetermined system into a determined system, taking also into account experimental data of measured fluxes. MFA is able to provide valuable insights into cellular physiology under specific conditions as well as to predict its metabolic capability after genetic or environmental perturbations.

Moreover, a frequently used approach to test the robustness of the metabolic model is the simulation method *Flux Variability Analysis (FVA)*. This approach computes the minimum and maximum allowable flux through each metabolic reaction to achieve optimal and sub-optimal objectives [65] allowing the user to investigate network flexibility and redundancy as well as to study flux distributions under sub-optimal growth.

4.2.5 *Community models*

The increasing generation of systems biology data enabled the construction of accurate metabolic models possible. Consequently, the advances in metabolic modelling resources and technologies led to a focus on the prediction of community behaviour [66]. However, community phenotypic prediction presents greater challenges than single species metabolic models due to interspecies interactions and more complex objective reactions [67]. Microbial communities have a large impact on nature ranging from soil ecology to environmental engineering. Thus, being able to predict their response to several environmental catalysts is of great relevance. Moreover, the comprehension of the community mechanics is essential in order to be able to design and engineer microbial ecosystems for desirable outputs [68, 69].

A common approach used for the reconstruction of these community metabolic models is through the reconstruction of high-quality individual models. This approach is based on the reconstruction of single-species models which will further be treated as compartments of a single cell, hence forming a community with all its interspecies interactions. However, in order for this approach to be accurate, the species in the study should be well characterized [67].

An alternative approach is to consider the community as a single supra-organism comprising either the genome of all the individual species or using a metagenome of the community. The networks of all organisms are combined by ignoring compartments of each individual species, resulting in a single cell that includes the systems of all the members of the community [70, 71].

Lastly, an approach using an automated single-organism pipeline is also commonly employed, resulting in fast-generating high-throughput individual models for each species. Softwares such as ModelSEED [44], COBRA [59], RAVEN [72] are normally used. Nevertheless, these top-down approaches may not be as accurate as the high-quality manual curated bottom-up approaches, since the models produced hold a certain similarity between them.

Furthermore, in the case of the species genome unavailability, steps such as genome structural annotation, gene calling, and genome functional annotation can be used [67]. The use of omics data in such community model's can improve the model's prediction capacity as well as the knowledge of the community interspecies interactions [73].

4.2.6 *Computational Tools*

The manual reconstruction of the process above for the GSM model is a time-consuming procedure [74]. Protocols have been described [8], reviews made [49, 29] and hundreds of GSM models have been manually built. However, the evolution of both technology and knowledge allowed for the development of several automated tools further aiding the reconstruction of GSM models [1]. These tools speed up the process by automating several steps of the reconstruction, such as the construction of the draft network, gap-filling, formation of the biomass equation as well as providing useful information for the manual curation of the model. Feature comparison of the tools discussed in this chapter is presented in table 2.

Each tool has its drawbacks and advantages. MetaDraft and AuReMe are very fast in building accurate GSM models for an organism in which already exists curated GSM models of phylogenetic close organisms. MetaDraft offers a user-friendly interface while AuReMe is command line-based focusing on traceability. For large-scale studies, tools such as ModelSEED and Pathway Tools are more suitable since these are the fastest to build a GSM model. Both ModelSEED and *merlin* allow for a sequence FASTA an input file, however, although *merlin* allows the user to curate and re-annotate the functional annotations of the submitted genome, other tools such as RAVEN and Pathway Tools require annotated genomes as input. Furthermore, while *merlin* has an integrated annotation tool, ModelSEED redirects to *Rapid Annotation of microbial genomes using Subsystems Technology (RAST)* annotation system [78]. Moreover, ModelSEED allows users to perform for FBA simulations however, it provides no chance for manual curation. For this, Pathway Tools and *merlin* are more suitable, guiding the user throughout the reconstruction, thus reducing the time and effort required for manual curation. AuReMe and MetaDraft generate draft networks with higher similarity to curated networks than the other tools with both RAVEN and *merlin* being capable of predicting subcellular localization. For less characterized species, RAVEN is a useful tool since it provides biochemical information from different databases. Besides the commonly used databases for biochemical information (KEGG and MetaCyc), *merlin* also uses the database TCDB for a better annotation of the transport proteins [1].

It is worth mentioning that web interfaces have the advantage of centralizing all software and reference data in a single location allowing for an instant benefit regarding model reconstruction methods as well as model underlying data. Whereas the standalone interface tools offer more

Table 2: Adapted from [1]. Feature comparison between several automated GSM model tools with a classification system where 1 is the worst and 5 the best.

Feature	AuReMe	MetaDraft	*merlin*	ModelSeed	Pathway tools	RAVEN
Up to date reference databases/resources	5	5	5	5	5	5
Software availability (easy access as well as free)	5	5	5	5	3	5
Dependence on free software and databases	5	5	5	5	4	3
Comprehensive documentation	3	4	3	4	5	5
It meets current standards of input and output	4	5	5	2	2	5
User-friendly interface	3	5	5	5	5	3
Open-source code	5	5	1	5	1	5
Completeness of model information fields	3	5	5	3	4	4
Similarity between automated draft reconstruction with manually curated models	5	5	3	3	3	3
Automatization	4	3	2	5	4	3
Manual refinement assistance	3	1	5	1	5	3
Provides identifiers from other databases for metabolites and reactions	3	5	1	5	5	5
Traceability	3	2	3	2	3	2
Automatic integration of experimental data for curation	3	1	1	2	3	1
Flexibility in parameter settings	4	4	4	3	3	4
Gap-filling	Yes	No	No	Yes	Yes	Yes
Simulation ready	Yes	Yes	No	Yes	Yes	Yes
Associated databases	KEGG, BiGG Models, MetaCyc	BIGG Models	KEGG, MetaCyc, UniProtKB, TCDB, NCBI	ModelSEED (In-house reaction database)	PGDB, MetaCyc	KEGG, MetaCyc
Reference	[75]	[76]	[3]	[44]	[77]	[72]

privacy to the researcher regarding their private genome sequences/annotations. The tools operated through MATLAB have the disadvantage of requiring a license.

Moreover, gap-filling is one of the most important steps in manual curation and can be very time-consuming. ModelSEED and Pathway Tools have an automated gap-filling while RAVEN suggests candidate reactions and *merlin* provide tools to find gaps. However, automated gap-filling still requires manual refinement since the focus of these automated algorithms is to restore model connectivity and pathway completeness and not accuracy. Most gap-filling algorithms are variants of the GapFill algorithm [79] which presents *mixed-integer linear programming (MILP)* to determine the minimum set of reactions to be added to the model. A major drawback of MILP is the substantial amount of time required to resolve the problem, particularly without a commercial optimization software. However, efforts are being

made, through linear programming, to reduce the computation time required for gap-filling [80, 81, 82].

Moreover, tools with no simulation support require the use of additional platforms, such as COBRA Toolbox, OptFlux, MEWpy among others. COBRApy allows the user to access the tools of the COBRA toolbox through python. The **COBRA toolbox** is a set of tools designed for the reconstruction and analysis of constraint-based metabolic models. This toolbox presents a variety of tools for many different objectives namely, FBA, visualization, reconstruction, gap-filling, sampling, fluxomics, metabolic engineering, support to novel solvers, novel input/Output, support and test suites. **OptFlux** was designed to meet the arising necessities in the metabolic engineering field. This software presents several functionalities such as phenotype simulation through methods of FBA and MFA, *Minimization of Metabolic Adjustment (MOMA)* [83] and *Regulatory on/off minimization of metabolic flux changes (ROOM)* [84]. Moreover, it is capable of performing a pathway analysis through the calculation of *Elementary Flux Modes (EFMs)* [85]. Also, allows for strain optimization through the *Evolutionary Algorithms (EAs)* [86], *Simulated Annealing (SA)* meta-heuristics [87] and OptKnock tool [88].

Furthermore, MEWpy [89] is a Metabolic Engineering Workbench designed in python for metabolic and regulatory modelling approaches as well as phenotype simulations. It encompasses a range of different phenotype prediction methods and EAs for metabolic engineering optimization being able to support ReFramed and COBRApy libraries. Reframed is a refactored version of the tool *FRAmework for Metabolic Engineering and Design (FRAMED)* which is a python tool for analysis and modelling of metabolic models allowing for several constraint-based simulation methods including community simulation. It offers support for the SteadyCom [90] algorithm which is capable of predicting the composition of a community whilst assuring community stability, given an environmental condition. Moreover, [91] is a python tool capable of predicting microbial community interactions through a global or detailed analysis. The global analysis is capable of predicting the competition for metabolites and how can the species interaction decrease the necessity for outside resources. The detailed analysis provides insights into species dependency, species uptake necessity for survival, the capacity of a species to produce a metabolite and an overall score that takes into account the previous three resulting in a probability of cross-feeding interaction.

The evolution of NGS technologies lead to an increase in the number of complete genomes sequenced. This growth, coupled with recent advances in GSM model reconstruction methodologies, is allowing us to build a model for each sequenced organism. However, there is a need for further improvement of the GSM model reconstruction methods towards shortening the time and efforts required for manual curation with more accurate results. Draft models generated from automated tools still require manual curation before presenting precise predictions [74] being common to find inaccurate reactions, GPR associations as well as incorrectly constrained reactions. This may be caused by incorrect mass and energy balance information stored in the databases [78] as well as an erroneous assignment of genes to reactions.

Moreover, the lack of organism-specific information presents a drawback for the formation of the biomass equation. In efforts to resolve these problems, some solutions are already being established. It is possible to assess the quality of the draft network through the tool *memote* [92]. Moreover, biomass generation equations of biologically related organisms can assist in the manual curation. Furthermore, the automatic extraction of the biomass equation of the intended organism as well as GPR associations from literature through text mining methods will have a great relevance towards the automation of the manual curation process. Notably, although automated methods for the formation of the biomass equation can be helpful, these do not replace the relevance of the biomass equation to be experimentally determined.

4.2.7 *Model applications*

Several applications can be taken from a GSM model. The analysis of GSM models for the identification of drug targets will facilitate the development of more efficient pharmaceutical compounds with lesser side effects. Additionally, predictive behaviour for different conditions and network analysis for product optimization while maintaining biomass levels are of great use for industrial applications. Also, the analysis of a pan-reactome through models of closely related organisms can provide valuable insights in several fields. Finally, modelling of an organism's interaction promotes the knowledge related to the mechanisms of interaction, contributing to several benefits.

GSM models allow the analysis of critical genes in which the absence of any of them will result in a non-viable network. In 2004, Yeh and co-workers provided the reconstruction model for *Plasmodium falciparum* [93], which was later presented with a list of potential drug targets by Fatumo and co-workers, in 2007 [94]. Moreover, Bordbar and co-workers, in 2010, [95] presented an interaction between the alveolar macrophage submodel of the Recon 1 [96] *Homo sapiens* metabolic reconstruction and the *i*NJ661 *Mycobacterium tuberculosis in silico* strain [97], hence building a host-pathogen model. This *in silico* interaction provided a better understanding of the pathogenic mechanisms, thus aiding in the identification of potential new drug targets. Altogether, these studies, coupled with many others, not mentioned here, promoted by the evolution of systems biology, provide precious insights into the identification of drug targets [30].

Moreover, the optimization of GSM models through knockout of genes or up- or down-regulation of gene expression for the over-production of metabolites of interest is rapidly increasing. In 2018, Yang and co-workers, with the objective of over-express D-phenyllactic acid, an aromatic polymer, used the model of the metabolically engineered *Escherichia coli* strain XB201T (*i*JO1366 [98]). It was proven *in silico* that the knockout of two genes (*tyrB*, *aspC*) would enhance the production of the intended compound [99]. In another study, the application of several algorithms was used in order to optimize the production of dodecanedioic acid in the

Yarrowia lipolytica GSM model, *i*YLI647, [100]. Optimizations like these lead to a re-design of an organism's metabolic network, promoting the development of cell factories for industrial and medical interests.

In other cases, the analysis of GSM models can promote the functional identification of unidentified reactions and/or enzymes. Using the *E. coli* model *i*JO1366, genes previously thought to be non-essential were predicted to be essential in a critical gene analysis [101]. This discrepancy was thought to be caused by unidentified reactions affecting the essentiality of a gene after its knockout. In order to test this, several genes were selected for experimental validation due to their sequence similarity. Through knockout of some of the genes the authors concluded that the missing genes can be compensated by other genes in alternative reactions, thus identifying possible isoenzymes [101]. Moreover, an algorithm was developed with the objective of finding promiscuous enzymes in a target organism through the use of its model [102]. In short, the algorithm looks for sequence homology in a gene similarity tree generating matrix of primary and potential promiscuous functions of enzymes encoded by the corresponding genes. A 'replacer' gene with a potential promiscuous function similar to the primary function of another conditionally essential gene is then identified. This algorithm has been used in several studies [44], proving its usefulness. Ultimately, high-quality GSM models facilitate the identification of new enzyme functions and enzyme promiscuity being very beneficial due to the lack of experimental data for many enzyme functions.

Furthermore, the use of GSM models for the analysis of the pan-reactome, the whole set of reactions of an organism, of phylogenetic close organisms, either for different strains of a species [103, 104] or for different species of a genus [105, 106], enhances the acquired knowledge of metabolic traits and lifestyles for these organisms. A study conducted to test the differences in the pan-reactome of 410 *Salmonella* strains using a reconstructed GSM for each strain revealed differences in the accessory reactome, a set of reactions present in only some strains. This study provided new information regarding preferred growth environments of the *Salmonella* as well as information about their evolution [105]. Identically, in 2018, other study analyzed the pan-reactome of 24 different species of *Penicillium* using a reconstructed GSM model for each species [106]. The study showed that despite the primary metabolism being highly conserved, the biosynthetic pathways for secondary metabolites were very different, thus contributing to the genomic differences amongst the 24 species of the genus. As shown above, analysis of the pan-reactome of close phylogenetic organisms coupled with automatic GSM reconstruction tools can present high biological importance.

Additionally, modeling of metabolic interactions between different cells or organisms has significantly increased its importance [76, 107, 108]. In 2019, a study conducted using GSM models to assess the impact of costless metabolites, defined in the study as metabolites that do not negatively affect the fitness cost of the organism when excreted, on the growth of different microorganisms under several environmental conditions and mediums. The authors concluded that the secretion of costless metabolites has a positive impact on the growth of

other microorganisms and, consequently in the taxonomic diversity observed in nature [108]. In another study, the interaction between four different gut bacteria using their GSM model provided precious insights into the contribution/competition on the production of short-chain fatty acids, amino acids, and gases which are known compounds to have essential roles in both intermicrobial metabolic interactions and the regulation of human metabolism [109]. The use of more microorganisms and exchanged metabolites will further increase the relevance of the data retrieved. Similarly, in 2018, Kumar and co-workers, reconstructed *Community Metabolic Models (CMM)* through the incorporation of GSM models of each species of the gut microbiota [110]. These CMM allowed for the identification of a lower rate of production of essential amino-acids in malnourished children. Moreover, the interaction between a host and its pathogen can also be studied through the use of GSM models [111]. Also in 2018, another study examined the impact of a pathogen in the host's plant photosynthetic pathways through the use of its GSM models [112]. Through this modulation, the authors concluded a negative impact of the pathogen in the Calvin cycle fluxes and further carbon fixation providing additional insights into the defence and attack mechanism of the plant and pathogen, respectively.

Although having described many applications of the GSM model, metabolic engineering has yet to achieve its full potential. The rapid increase in GSM model applications is correlated with the increased amount of biological data and information as well as automatic GSM model reconstruction tools availability. GSM models achieved the integration of data such as protein allocation [113, 114], cellular macromolecular composition [115, 116], and protein structural information [117, 104] with enzyme–substrate interactions, structure of protein–protein complexes, and post-translational modification still having to be further attained [74]. For this, higher coverage of GPR associations, reconciling model inconsistencies, developing novel mathematical modelling techniques for high-throughput data and the incorporation of cellular processes beyond metabolism [118] have yet to be achieved.

4.3 EXTREMOPHILES

Extremophiles are defined as organisms capable of thriving in environments with significant chemical and physical barriers for life support as described in table 3. These extremophiles belong to Archaea, Prokarya, and Eukarya domains [10] have developed strategies and mechanisms to survive under these extreme environments [9]. Studying the mechanisms of these extremophiles provides fundamental insights having great interest in biotechnological e. g., extremozymes and commercial purposes e. g., production of biofuel [119].

Many acidophiles are found in acid mine drainage water or mine spoils with enough oxidation of elemental sulfur and sulfide minerals to produce sulfuric acid resulting in a higher pH value of the environment. Acidophiles thrive in high pH environments by excreting acid out of the cell in order to maintain a pH gradient through the plasma membrane. In this way,

Table 3: Adapted from [2]. Extremophile classification and characteristics.

Environment parameter	Class	Defining growth condition	Environment/Source	Example organism
High temperature	Thermophile	60-80 °C	Hot-springs	*Synechococcus lividus, Sulfolobus sp.*
High temperature	Hyperthermophile	>80 °C	Submarine hydrothermal vents	*Pyrolobus fumarii, strain 121*
Low temperature	Psychrophiles	<15 °C	Ice, snow	*Psychrobacter, Methanogenium spp.*
Low pH	Acidophiles	pH<5	Acid mine drainage, volcanic springs	*Picrophilus oshimae/torridus, Stygiolobus azoricus*
High pH	Alkaliphiles	pH>9	Soda lakes	*Bacillus firmus OF4, Haloanaerobium alcaliphilum*
High pressure	Barophiles	High pressure	Deep ocean	*Pyrococcus sp., Colwellia sp*
High salinity	Halophiles	2 - 5 *M* NaCl	Salt lakes, last mines	*Halobacteriaceae, Dunaliella salina, Halanaerobacter sp.*
Radiation	Radiophile	High levels of radiation	Radioactive waste from mining	*Deinococcus radiodurans, Thermococcus gammatolerans*
Desiccation	Xerophile	Anhydrobiotic	Desert, rock surfaces	*Artemia salina, Deinococcus sp., lichens, Methanosarcina barkeri*
Rock-dwelling	Endolith	Resident in rock	Upper subsurfaces to deep subterranean	Lichens, cyanobacteria, *Desulfovibrio cavernae*

the organism ensures an internal pH range of 5,0–7,5 allowable for biological reactions to occur. Commonly these organisms have reverse membrane potentials, highly impermeable to H_3O^+, with an elevated concentration of secondary transporters and extracellular enzymes that can operate at a high pH value. Acidophiles not only are resistant to metal-rich environments but can also generate energy or ATP from metals being a precious source of acid-stable enzymes [10].

Thermophiles prosper in environments with elevated temperatures, ranging from 61–79 °C being usually found in a variety of environments, namely, in hot springs, bioreactors, deep oil wells, geothermal plants, coal piles, compost heaps, deep-sea hydrothermal vents. These organisms adapted to the higher temperatures by modifying the lipid compositions of the membranes such that they are in a liquid crystalline state, conferring a lower rate of proton permeation to the cell. The enzymes from these organisms present the same function as their non-extreme counterparts, with the advantage of bestowing the capability to perform their function under high temperatures, extreme pH values, high substrate concentrations, and high pressure without denaturing. This provides several benefits for the industrial area since biocatalysts are widely used and most procedure conditions are severe [9]. Thus, the study of thermophiles presents several benefits in an industrial paradigm, having contributed already

in processes such as anaerobic fermentative processes for water treatment, fuel production, sulfur removal from crude oil, etc [10].

The study of extremophiles generated great benefits through diverse fields. The discovery of the DNA polymerase from thermophilic bacteria [120] boosted the advances in molecular biology and life sciences in the last decades. On the opposite side, the cold-active enzymes from psychrophiles are highly used in food processing to improve milk fermentation, to store frozen yogurt, and to improve ice-cream production [121]. Moreover, the use of thermophiles for the production of hydrogen through hydrogenases and anaerobic fermentation is widespread [122]. Another thermophile example is in the production of biofuel which showed to exceed the mesophilic microorganisms based on traditional methodologies since the process involves high temperature and low ph [123]. Moreover, the nucleoside phosphorylase from hyperthermophiles was used for the synthesis of nucleoside analogues for antiviral treatment [124] since it was discovered that the use of these enzymes lowers the risk of microbial contamination and viscosity of higher solubility of substrates [125]. Therefore, the use of extremophiles is proving to be of great benefit in biotechnological and commercial applications showing an elevated relevance in the fields of medicine (e.g., antibiotics and antitumors), food technology (e.g., phytases and phosphatases), biofuel production (e.g., proteases and lipases), cosmetic industry (carotenoids), biomining and contaminated soil remediation (xenobiotic-degrading enzymes), agriculture (plant growth inducers), and organic residue cycling (cellulose and lignocellulose) [126].

Notably, these extremophiles are commonly found in nature living in a highly social multi-species microbial community, which through collective response to their environment, leads to a response at a larger scale [127]. The resulting interactions can either be for space competition or cooperation with beneficial metabolic exchanges boosting their chances of survival [91]. These mutualistic exchanges can lead to interspecies dependencies further increasing the importance of the study of extremophiles communities [128].

4.3.1 *Acidithiobacillus caldus SM-1*

Acidithiobacillus caldus SM-1 is a moderately thermophilic and obligately acidophilic proteobacterium with its optimal growth temperature ranging between 40 °C and 45 °C and pH ranging between values of 2 and 2,5 [15]. This rod-shaped (0.4 x 2.0 μm) gram-negative bacterium is motile with one or more flagella using CO_2 as the primary carbon source for autotrophic growth [129]. Although *A. caldus* is not able to grow heterotrophically, it has been shown its capacity to grow mixotrophically with glucose or yeast extract as carbon sources and tetrathionate as energy source [130].

This obligately aerobic bacterium [131] is widely used in biomining [132, 133] as well as in acid mine drainage [15] due to its sulfur-oxidizing capability. In biomining, it is used for

oxidization of elemental sulfur and *reduced inorganic sulfur compounds (RISCs)*, generating the required acidity for the process and later removal of the accumulated elemental sulfur, which would retard the oxidation of ores [134]. The RISCs oxidation by *A. caldus* has shown many advantages being already proved that mechanisms behind this oxidization are related to ATP generation via electron transport phosphorylation [15].

4.3.2 *Acidimicrobium ferrooxidans DSM 10331*

Acidimicrobium ferrooxidans DSM 10331 is a moderately thermophilic, iron-oxidizing acidophile with an optimal growth temperature of between 45-50 ºC and pH value of 2. It is able to grow under heterotrophic conditions with yeast extract as carbon source [16]. Under anaerobic conditions uses hydrogen as an electron donor and ferric iron as an electron acceptor [135].

This gram-positive small (0.4 μm x 1-1.5 μm) rod-shaped bacterium was firstly isolated in 1996 from a hot spring in Iceland by Clark and Norris, having remained extremely phylogenetically isolated as the single type strain in the actinobacterial subclass *Acidimicrobidae* for several years [16]. Under autotrophic growth, *A. ferrooxidans DSM 10331* present a low requirement for CO_2 forming small colonies when grown on ferrous iron. Nowadays, this microorganism has been used for its ability to oxidize ferrous iron to ferric iron in order to produce ATP being extremely useful in the biomining field [136].

4.3.3 *A. caldus SM-1 and A. ferrooxidans DSM 10331 community*

Several studies have been using *A. caldus SM-1* and *A. ferrooxidans DSM 10331* for metal sulfide oxidation, suggesting a preferable interaction by these bacteria [137, 138, 139, 140, 141]. However, little is known regarding their interaction. Hence, studying these organisms in a community can provide valuable insights into their interaction mechanisms.

5

SCHEDULE

- Weeks 1-12: Studying the state-of-the-art; using *merlin* and implemented methods to perform *A. caldus SM-1* and *A. ferrooxidans DSM 10331* genomes' functional annotation; writing pre-thesis.
- Weeks 13-19: Developing a draft of the *A. caldus SM-1* and *A. ferrooxidans DSM 10331* metabolic networks.
- Weeks 20-25: Applying concepts of constraint-based modelling and flux balance analysis.
- Weeks 26-34: Using scientific literature information and experimental data for both metabolic model validation.
- Weeks 35-40: Analyze both organism metabolic networks to find potential compounds or proteins of interest.
- Weeks 13-40: Writing the thesis.

6

MATERIALS AND METHODS

In this section, a description for each step of the metabolic model reconstruction using the semi-automated reconstruction tool *merlin* is provided.

6.1 GENOME FILES

Genome files regarding the specific organism were automatically retrieved from the NCBI assembly directory using the taxonomy ID 990288 and 525909 for *A. caldus* and *A. aferrooxidans*, respectively.

6.2 GENOME ANNOTATION

A functional annotation against each genome was performed with an e-value threshold of 1.0e-30 and a maximum of 100 homolog sequences for each gene. For time-consuming purposes and to ensure priority of curated data, a first BLASTp against the UniProt-SwissProt database was performed. Genes with no hits were then submitted to a second BLASTp against the UniProt-*Translated EMBL Database (TrEMBL)* database. For each homolog sequence of each gene *merlin* assigns a score being represented as:

$$score = \alpha \times Score_f + (1 - \alpha) \times Score_t \tag{6}$$

Where:

- α is a weighting parameter
- $Score_f$ is the frequency of each EC number for every gene homology result
- $Score_t$ is the taxonomy score for the organisms associated with every homology result

merlin automatically assigns the homology result with the best score to every gene, automatizing otherwise a time consuming process.

6.2.1 *Automatic workflow*

In order to add robustness to the annotation results, an additional feature was performed. In this step, a list of phylogenetic close organisms was provided to *merlin* as represented in tables 4 and 5. For each gene, the software searches the blast hits for the organism/genus provided in the automatic workflow list and assigns a confidence label. This is done to each database generated by the two blasts. In this way, *merlin* prioritizes hits from related organisms reducing the chance of false positives. A lower e-value threshold is set for the more distant organisms in order to reduce the false-positive hits by demanding a higher level of similarity between the query and homologue sequences.

Table 4: Automatic workflow used for *A. ferrooxidans*

Taxa type	Taxon	E-value
species	*Acidimicrobium ferrooxidans* (strain DSM 10331 / JCM 15462 / NBRC 103882 / ICP)	1E-20
genus	*Acidimicrobium*	1E-20
genus	*Actinobacteria*	1E-20
genus	*Rubrobacter*	1E-30
genus	*Bifidobacterium*	1E-30
genus	*Frankia*	1E-30
genus	*Nocardioides*	1E-30
genus	*Propionibacterium*	1E-40
genus	*Bacillus*	1E-40

Table 5: Automatic workflow used for *A. caldus*

Taxa type	Taxon	E-value
species	*Acidithiobacillus caldus (strain SM-1)* (strain DSM 10331 / JCM 15462 / NBRC 103882 / ICP)	1E-20
genus	*Acidithiobacillus*	1E-20
genus	*Thiobacillus*	1E-20
genus	*Thioalkalivibrio*	1E-30
genus	*Aquifex*	1E-30
genus	*Picrophilus*	1E-30
genus	*Sulfolobus*	1E-30
genus	*Thermoplasma*	1E-40
genus	*Escherichia*	1E-40

6.3 ASSEMBLY OF THE DRAFT METABOLIC NETWORK

merlin resorts to KEGG to import the metabolic data. All spontaneous reactions, enzymatic reactions with an EC number encoded by the organism's genome as well as the metabolites present in those reactions are integrated into the metabolic model, generating a draft metabolic network.

6.3.1 *Compartmentalization*

For the model compartmentalization a web-based tool, Psort 3.0 was used. This tool predicts the subcellular localization of proteins given a protein FASTA file of the organism. *merlin* allows for the integration of the results of this tool in the long format.

6.3.2 *Transporter proteins and exchange reactions*

The integration of the tool *Transport Systems Tracker (TranSyT)* in *merlin* allows for the annotation of the transporter proteins through sequence homology against the TCDB by associating genes with the TC number. Then, TranSyT integrates the transport reactions between the compartments available into the model. For every reaction between the extracellular and intracellular, space this tool creates an exchange reaction for the respective metabolite. These reactions allow dictating which metabolites are able to be imported or exported of the metabolic model. In this way, it is possible to simulate growth mediums, environmental conditions as well as intra-community metabolic interactions.

6.3.3 *Gene-Protein-Reaction associations*

Genes are connected to reactions through the GPR associations. This is important to correctly predict mutant phenotypes through gene knockout. *merlin* does these associations by resorting to the information present in KEGG BRITE, retrieving the respective GPR rules for each gene through homology of orthologs genes.

6.4 BIOMASS FORMULATION AND ENERGY REQUIREMENTS

The biomass reaction represents the molecular composition of the cell to be modelled. It should account for all the metabolites required for the cell to replicate. For this, experimental data of the organism was used, when available, since an erroneous biomass equation can have an impact on the model's performance. Whenever this information could not be found, data from phylogenetically related organisms or other models was used. In this work, an approach

using abstract macromolecules entities in the biomass was preferred for both organisms being its relative content stoichiometrically represented by grams of macromolecule per gram of biomass.

Both biomass equations are composed of seven macromolecules: DNA, *Ribonucleic acid (RNA)*, Protein, Cofactors, Carbohydrates, Cell wall and Lipids. A total of seven reactions were created to form these macromolecules through the respective building blocks. The relative macromolecule contents of the biomass were based on the stoichiometric model of *Acidithiobacillus Ferrooxidans* [142] and on the *Bacillus subtilis* model [143] for the *A. caldus* and *A. ferrooxidans* models, respectively.

Moreover, *merlin*'s tool *"e-Biomass"* infers, through genomic data, the stoichiometric coefficient of the precursors of DNA, RNA and protein. For the Cofactors equation, a set of universal cofactors were used as described in [55]. Changes in some precursors were made whenever information of specific cofactors of the organism was available. A reaction to produce a generic fatty acid was added to both models. Information regarding fatty acid composition and the respective molar fraction was based on the literature [144, 145]. This generic fatty acid will then substitute the R groups from the lipids in the biomass equation of the respective models.

Lastly, it is required to create an ATP hydrolysis reaction to account for the growth- and non-growth associated maintenance where the first is represented by the stoichiometry and the last through the reaction flux boundaries. Growth associated maintenance was calculated as described in [8] and the values of the non-growth associated maintenance were based on the *Acidithiobacillus Ferrooxidans* model [142] for the *A. caldus* model and on the *Mycobacterium tuberculosis* model [146] for the *A. ferrooxidans*.

6.5 Manual curation of the draft network

After the implementation of the biomass reaction into the model, it is necessary to add a minimum defined medium in order for the model to become feasible.

6.5.1 *Reaction reversibility and directionality*

Since all reactions imported from KEGG are reversible it is necessary to correct their reversibility and directionality. *merlin*'s tool *"Correct Reversibility"* allows to automatically correct this by resorting to the Zeng database [147] and choosing the organism type. However, multiple reactions still need to be manually curated. For this, databases such as ModelSEED, BiGG and MetaCyc were used.

6.5.2 *Gap-Filling*

In order to ensure the model complies with steady-state assumptions, an analysis of the model's network connectivity was necessary. An unconnected network would result in the accumulation of dead-end metabolites which would fail the steady-state assumption in which the variation of the metabolite's concentration must be zero. *merlin*'s tool *"Blocked reactions"* highlights dead-end reactions as well as reactions that lead to a blocked reaction. This allows to easily identify the blocked metabolites in the metabolic network. Moreover, the tool *"Draw in browser"* colours enzymes and reactions of a selected KEGG pathway depending on the current state of the model, as shown in figure 1.

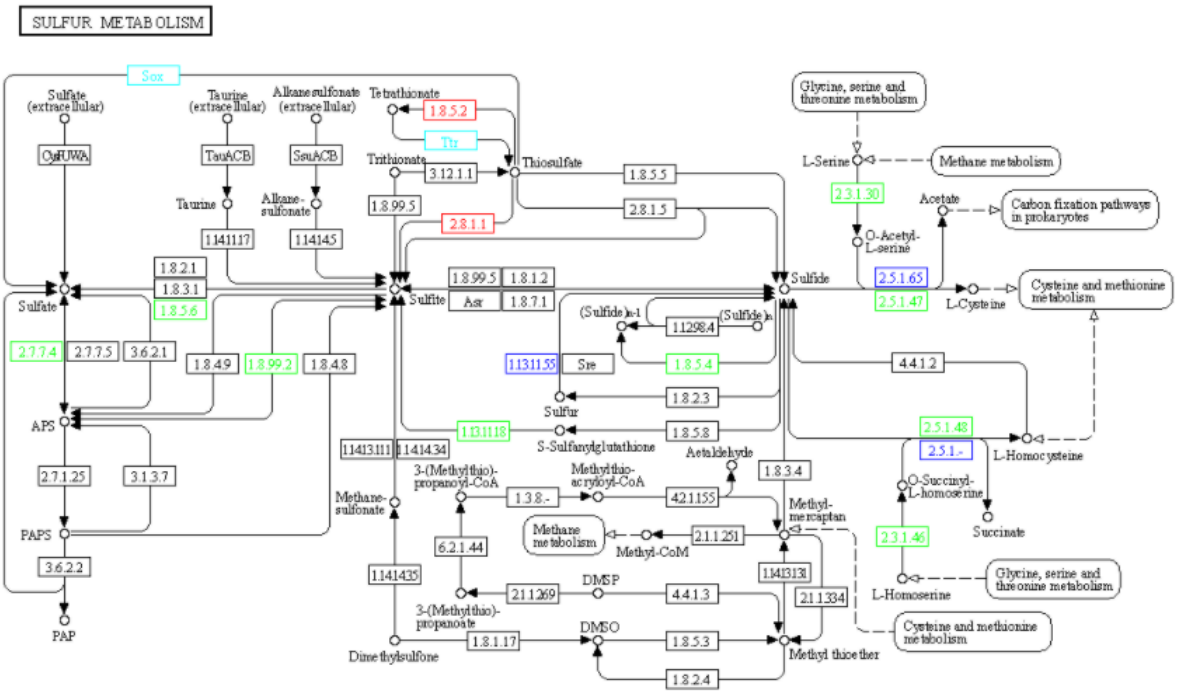


Figure 1: KEGG metabolic pathway with the *merlin* feature *Draw in browser*

Where:

- Green - Enzyme in model
- Dark blue - Reaction in the model, however enzyme missing from the model
- Light blue - Reaction that leads to a blocked reaction
- Red - Blocked reaction

Moreover, the tool *Biological networks In Silico Optimization (BioISO)* identifies where in the metabolic network the flux stops, given an objective function. This tool can be of great use

i. e. to ensure the flux reaches the biomass equation. A pathway-by-pathway analysis was performed and model feasibility was ensured. Commonly, several miss-annotations were detected and corrected using genome information and KEGG's organism reference pathways. However, the presence of some reactions was only supported by literature and in such cases, the reactions were added to the model but the enzyme was not associated with any gene.

6.5.3 *Mass balancing*

The balance of several reactions was corrected since an incorrect balance can lead to misleading flux distributions and model predictions. For this, *merlin* has an automated tool that highlights unbalanced reactions as well as provides the molecules and respective quantity on both sides of the reaction. Databases such as BiGG, ModelSEED and MetaCyc were used as well as genome-scale models of phylogenetic close organisms.

6.6 MODEL VALIDATION

In order to validate both models, several strategies were employed. Different environmental conditions were tested. For each condition, an analysis of the flux distribution and exchanged metabolites was made. These simulations were performed using cobra toolbox and Mewpy [89]. pFBA simulations were made and the results were compared to experimental data, whenever available.

6.6.1 *Environment conditions*

A set of environmental conditions were tested depending on the characteristics of the organisms being evaluated. Moreover, a condition with no open drains was tested to access model redundancy. For all environmental conditions, the simulation method pFba was used with the objective function being maximizing the biomass production.

6.6.1.1 *A. caldus*

A. caldus is an obligatory aerobic organism, thus it is only able to grow in the presence of oxygen. Moreover, *A. caldus* is able to grow autotrophically with sulfur or tetrathionate as energy sources, as replicated in tables 15 and 16, respectively. Values for CO2 uptake were based on literature [148, 149].

A. caldus is also able to grow mixotrophically with glucose as carbon source and tetrathionate as energy source as scrutinized in table 19, respectively. The value of glucose uptake was based on literature [150].

6.6.1.2 *A. ferrooxidans*

A. ferrooxidans is a facultative aerobe, thus aerobic and anaerobic environmental conditions were tested. A second biomass equation was created for anaerobic conditions and cell growth was ensured for both conditions. Moreover, *A. ferrooxidans* is a facultative autotrophic organism, meaning that is able to fix carbon through CO_2, whenever present. Although it is also able to grow heterotrophically on yeast extract, it is not possible to replicate such environmental condition since a medium with yeast extract is not defined and hence it cannot be simulated.

6.6.2 *Flux distribution*

An analysis of flux distribution was made for both models. Carbon and energy metabolism were scrutinized. Crucial sectors of the metabolism such as the Calvin cycle, glycolysis/gluconeogenesis and pentose phosphate were thoroughly analysed.

Regarding the energy metabolism, a pathway 'Oxidative phosphorylation' was created since this pathway was missing. Moreover, *A. caldus* generates ATP through oxidation and reduction of sulfur, thus sulfur metabolism was analyzed. *A. ferrooxidans* generates H+ gradient through oxidative reductions of iron which was also analysed.

6.7 MICROBIAL COMMUNITY

The community was built and simulated using the tool ReFramed [151]. Each model was treated as a compartment and an additional extracellular space common to both models was added. A community biomass reaction was added taking into account the biomass of both organisms. A community minimum medium comprised of each organism's minimum medium was used and an aerobic autotroph condition was tested.

Moreover, SteadyCom [90] was used to infer the weight of each organism to the community. Lastly, SMETANA [91] was used to analyse the probability of interactions between the members of the community and the results were compared to the ReFramed simulation results.

7

RESULTS AND DISCUSSION

7.1 GENOME ANNOTATION

A GSM model reconstruction starts with functional annotation of the genome if the structural annotation has already been performed. In addition, only the annotation of metabolic genes (genes encoding for an enzyme or transport protein) will be taken into account. As represented in table 6, these genes correspond to around 30% and 40% of the entire genome for the *A. caldus* and *A. ferrooxidans* models, respectively. Out of these annotated genes, 176 and 122 were associated with partial EC numbers for each of the respective organisms.

Table 6: Annotated genes of each organism.

	A. caldus	***A. ferrooxidans***
Genes	3186	1964
Annotated genes	940	785
% of annotated genes	29.50	39.97

7.1.1 *A. caldus*

A total of 940 genes were annotated after the *'Automatic workflow'* corresponding to 29.5% of the entire genome. There were no confidence label 'A' matches in the automatic workflow for the UniProt-SwissProt database due to the absence of available entries for *A. caldus*, as depicted in figure 2. There were 148, 110, 23 and 95 genes assigned to the confidence label 'B', 'C', 'D' and 'E', respectively. The further it goes taxonomically from *A. caldus* the lesser the probability of a homolog sequence, hence the frequency of the labels should decrease as the software runs the list provided to the *'Automatic Workflow'*. *Escherichia* was attributed to the last confidence label since is the model organism for the gram-negative bacteria, thus having an abundance of curated information, resulting in 262 genes assigned to the label 'I'. A total of 213 genes were left as 'Default' since none of the confidence labels could be applied. The genes that could not be annotated through the UniProt-SwissProt database had at least one homologue sequence

from *A. caldus* in the annotation through the UniProt-TrEMBL database. Consequently, all genes were assigned with the label 'A'.

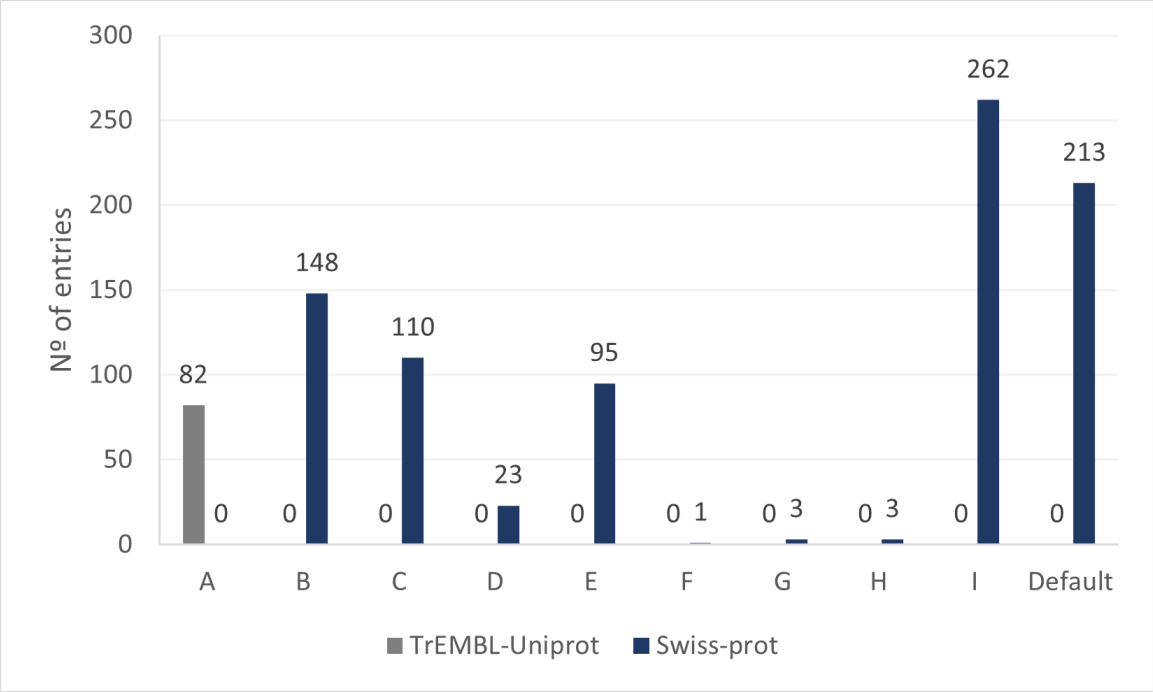


Figure 2: Automatic workflow results for the *A. caldus* model representing the number of genes assigned to each confidence label for both databases.

A total of 840 genes were left with the same EC number for both databases, as shown in figure 3. The EC number of 100 genes was changed after the *'Automatic Workflow'* for the UniProt-SwissProt database and no changes were made for the genes annotated through the UniProt-TrEMBL database.

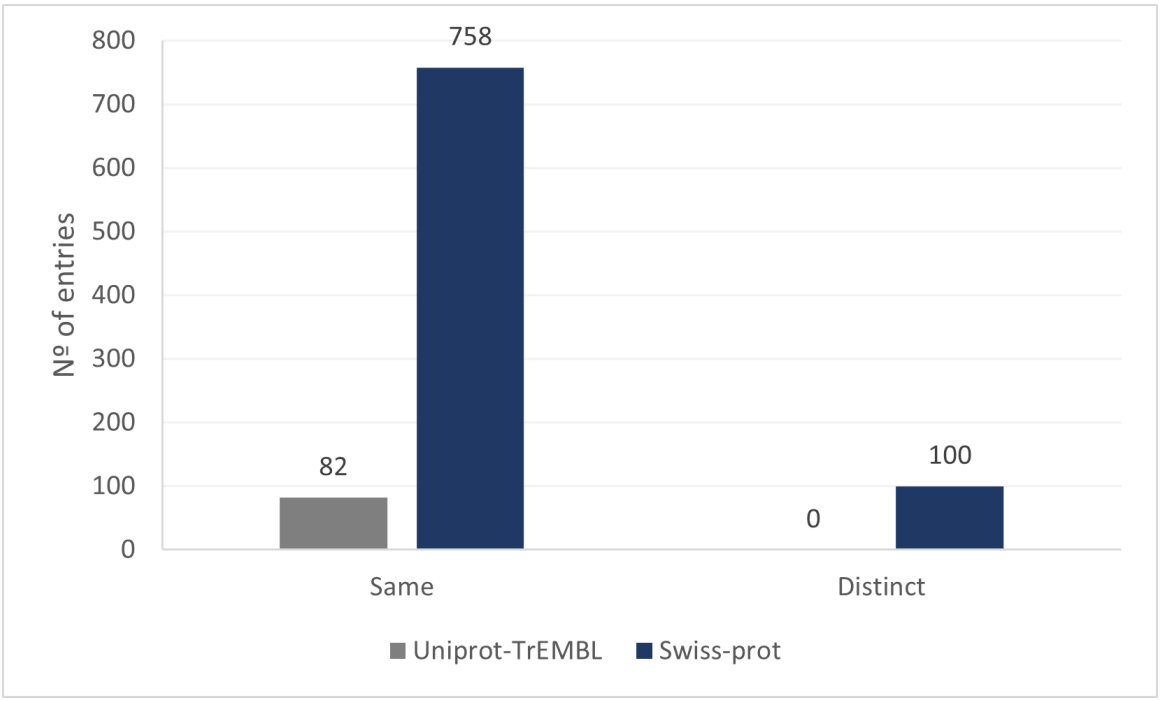


Figure 3: EC number changed with automatic workflow for the *A. caldus* model.

7.1.2 *A. ferrooxidans*

A total of 8, 0, 0 genes were left with the confidence label 'A', 'B', 'C', respectively, as represented in figure 4. The lack of labels 'B' and 'C' may be explained by the few respective genera's curated entries which, in this case, did not correspond to any homolog sequence. Moreover, in the same way as for *A. caldus*, the gram-positive model genus *Bacillus* was selected for the last confidence label, resulting in the annotation of 234 genes. A total of 305 were left as 'Default'. Similar to *A. caldus*, the only genes annotated in the UniProt-TrEMBL database were through entries of the own organism.

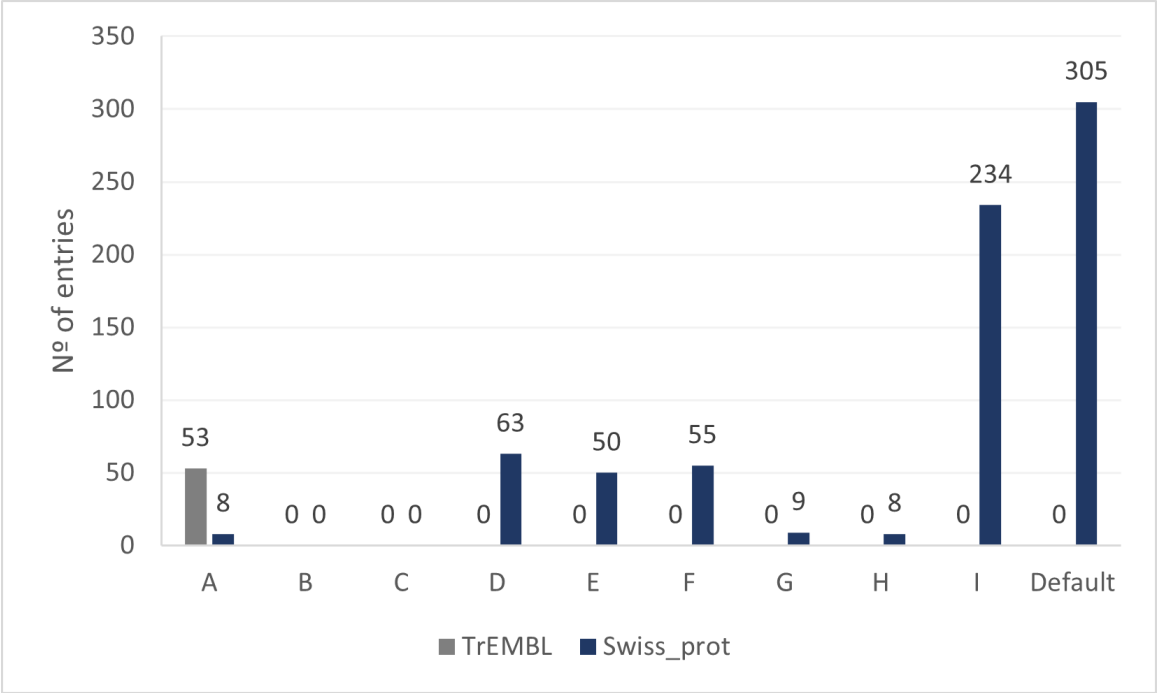


Figure 4: Automatic workflow results for the *A. ferrooxidans* model representing the number of genes assigned to each confidence label for both databases.

As shown in figure 5 a total of 632 and 7 genes EC numbers were left unchanged for the UniProt-SwissProt and UniProt-TrEMBL databases, respectively. A total of 100 and 46 gene's EC numbers were changed for the Swiss-Prot and TrEMBL databases, respectively, due to a given priority of the 'Automatic Workflow' selected genera which shifted the functional annotation from the best possible result to a more accurate one since in the construction of these models is preferable quality over quantity of information.

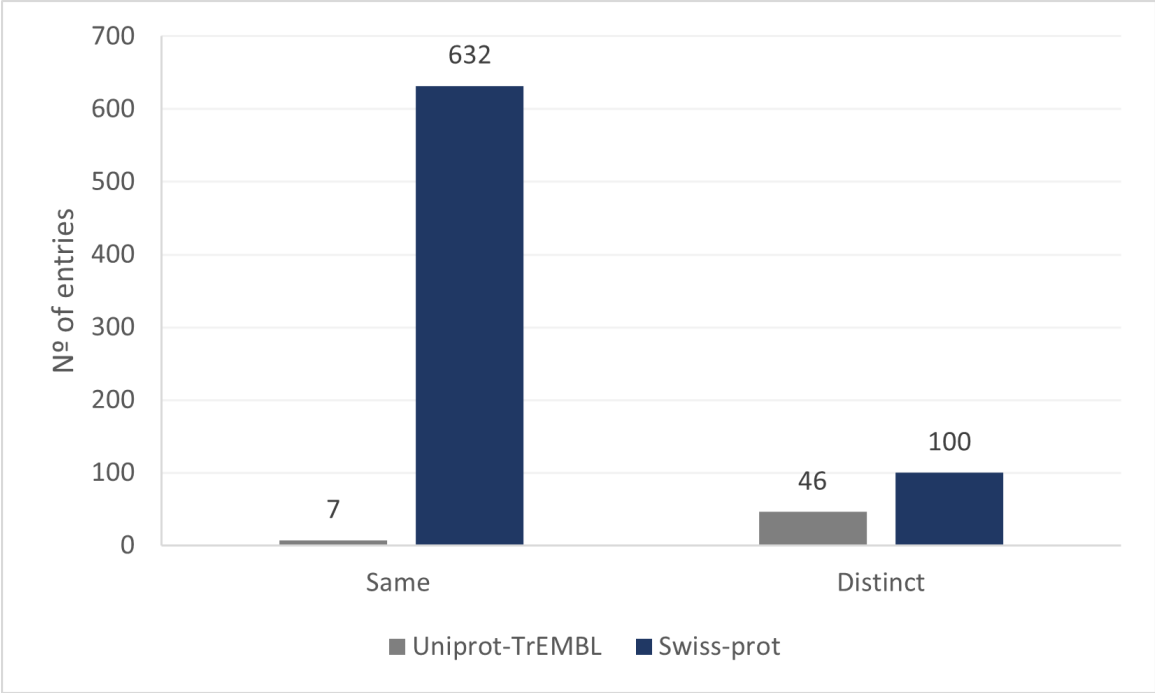
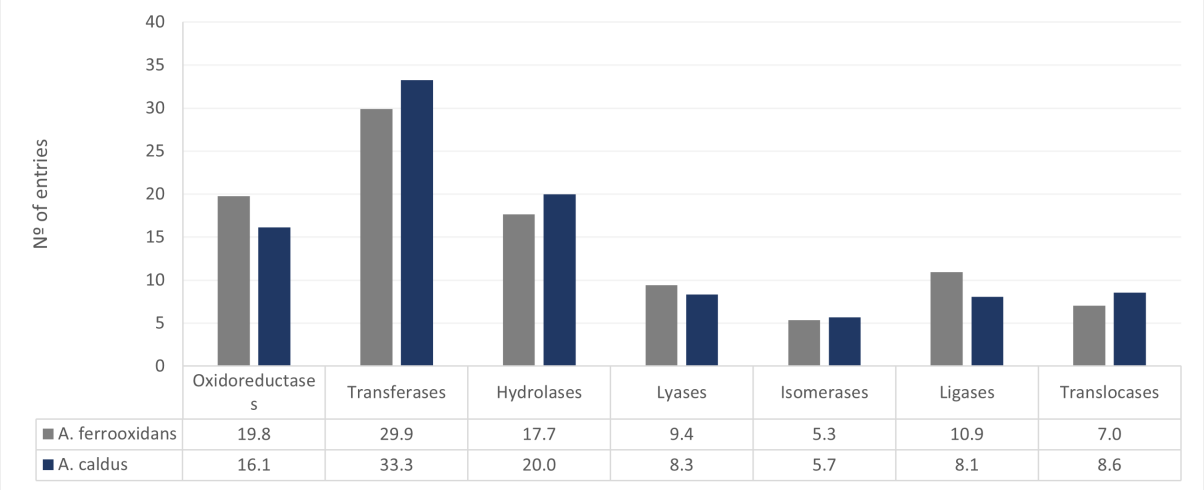


Figure 5: EC number changed with automatic workflow for the *A. ferrooxidans* model.

Enzyme classification system divides enzymes into 7 families: EC 1. Oxidoreductases, EC 2. Transferases, EC 3. Hydrolases, EC 4. Lyases, EC 5. Isomerases, EC 6. Ligases and EC 7. Translocases. As seen in figure 6, the first three classes hold the majority of enzymes in both models whilst the EC 5. isomerases represent the less frequent class in both models. Around 17% of the total EC numbers for both models represent an incomplete EC number. These were not included in the model unless a reaction could be catalysed by an incomplete EC number.



	Oxidoreductases	Transferases	Hydrolases	Lyases	Isomerases	Ligases	Translocases
A. ferrooxidans	19.8	29.9	17.7	9.4	5.3	10.9	7.0
A. caldus	16.1	33.3	20.0	8.3	5.7	8.1	8.6

Figure 6: Overview of enzymes class present in each model.

7.2 METABOLIC NETWORK

After genome annotation, the KEGG reactions associated with the resulting EC numbers were imported into the model and a draft metabolic network was generated.

An overall comparison between the draft networks and the curated ones was made, being represented in table 7. The number of reactions, metabolites and pathways in the models decreased from the draft to the curated models due to the refinement of the network during the manual curation. Furthermore, the curated models contain around 20% of the total reactions added manually. This may be due to the lack of functional information on the organisms in this study as well as on phylogenetically close organisms, which consequently would result in several missing reactions.

Moreover, unique reactions are reactions that are present in one model and absent in the other by comparing between the draft models and the curated ones. Therefore, the *A. caldus* draft model contains 319 reactions that are not present in the *A ferrooxidans* draft model with the latter possessing 404 unique reactions. Whilst in the curated models, the value of unique reactions slightly increased for the *A. caldus* model and decreased for the *A. ferrooxidans* model. The same comparison can be made for the metabolites, however, in this case, the number of unique metabolites decreased in both models from the draft to the curated ones.

Table 7: Overview of the metabolic network of draft and curated models.

	Draft		**Curated**	
	A. caldus	***A. ferrooxidans***	***A. caldus***	***A. ferrooxidans***
Reactions in model	945	1030	846	817
Reactions inserted by homology (%)	660 (69.84)	729 (70.78)	457 (54.02)	470 (57.53)
Reactions inserted manually (%)	-	-	184 (21.75)	161 (19.71)
Unique reactions (%)	319 (33.76)	404 (39.22)	339 (40.07)	305 (37.33)
Spontaneous reactions (%)	117 (12.38)	123 (11.94)	11 (1.30)	12 (1.47)
Total metabolites	1066	1202	646	640
Unique metabolites (%)	246 (23.08)	382 (31.78)	184 (28.48)	174 (27.19)
Pathways in model	115	126	86	95

7.2.1 *Compartmentalization*

7.2.1.1 *A. caldus*

A total of 5 compartments were predicated for the *A. caldus* model, as represented in table 8. However, the model only contains 3 compartments: cytoplasm, periplasm and extracellular. 'Cytoplasmatic membrane' represent transport reactions from the cytoplasm to the periplasm whilst 'outer membrane' represent the exchange reactions from the periplasm to the extracellular space. Genes with unknown localization were assigned to the cytoplasm.

Table 8: Compartments predicted by PSORTb 3.0 for the *A. caldus* model.

Compartment	**Gene count**
Cytoplasmic	1417
Cytoplasmic_membrane	699
Extracellular	31
Outer_membrane	48
Periplasmic	66
Unknown	925

7.2.1.2 *A. ferrooxidans*

A total of 4 compartments were predicated for the *A. ferrooxidans* model, as shown in table 9. As *A. ferrooxidans* is a gram-positive bacteria it lacks the periplasm compartment. Therefore an abstract compartment, *Proton Motive Force (PMF)*, was created in order to be able to simulate the H^+ gradient for the generation of ATP. Additionally, the model contains two additional compartments: cytoplasm and extracellular space. Similar to *A. caldus* the predicted compartment 'Cytoplasmic_membrane' represent the transport reactions.

Table 9: Compartments predicted by PSORTb 3.0 for the *A. ferrooxidans* model.

Compartment	**Gene count**
Cellwall	5
Cytoplasmic	1147
Cytoplasmic_membrane	610
Extracellular	28
Unknown	174

7.2.2 *Transport Reactions and drains*

A comparison between the draft and curated models was performed. Although the number of drains is similar, the curated models contain more transport reactions, as represented in table 10. This is explained by the higher number of compartments in the curated models. The draft models only contain two compartments (inside and outside) therefore the transport reactions existing in these models are only between these 2 compartments.

Table 10: Comparison of the number of transport reactions and drains between the draft and curated models.

	Draft		**Curated**	
	A. caldus	***A. ferrooxidans***	***A. caldus***	***A. ferrooxidans***
Drains	72	74	65	72
Transporters	91	99	156	112

7.2.2.1 *A. caldus*

A total of 156 transport proteins were annotated in the *A. caldus* model. The most common direction of the transporters was from the outside to inside the model totalling around 35% of the transporters, as represented in figure 7.

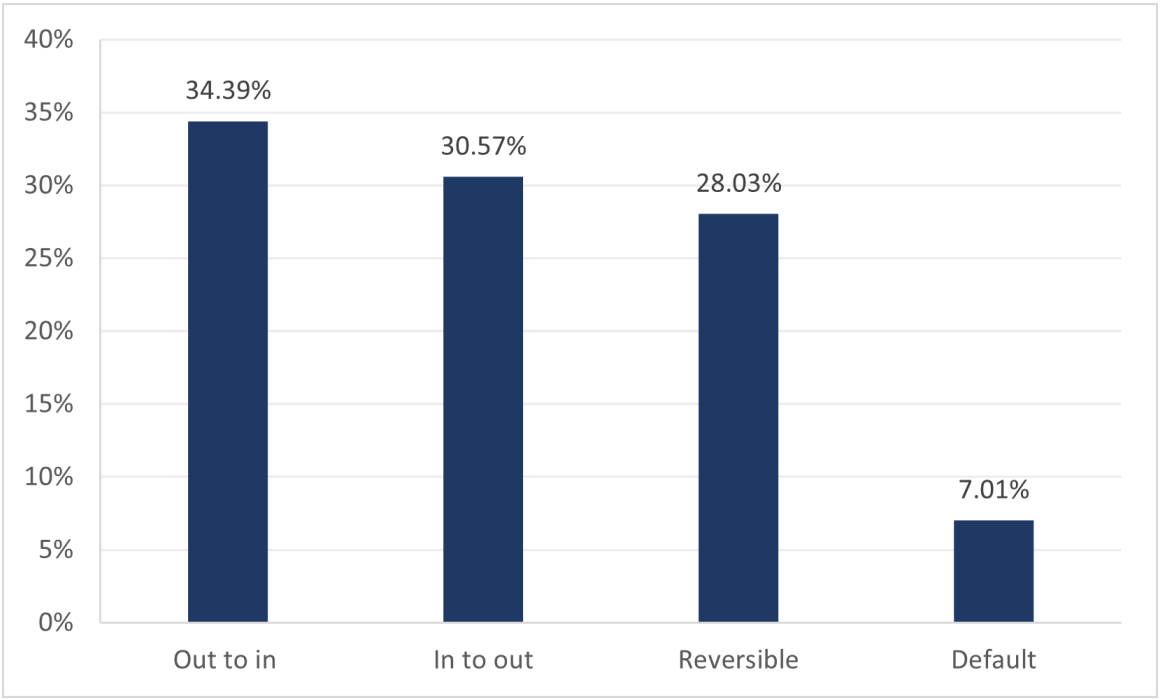


Figure 7: Direction of the transport proteins in the *A. caldus* model.

As represented in figure 8, uniport transport proteins are the most common type of transporters present in the model, totalling around 38% of the model's transporters. This type of transport proteins allows for the transport of only one metabolite at the same time. Symport proteins allow for the transport of more than one metabolite at the same time in the same direction, which contrasts with antiport proteins that allow for the transport of more than one metabolite in opposite directions. These transport proteins total around 27% and 14%, respectively, of the transporters in the model. Lastly, ABC transporters are transport proteins associated with energy cost, in other words, they require the consumption of ATP to be activated, totalling around 21% of the transporters in the model.

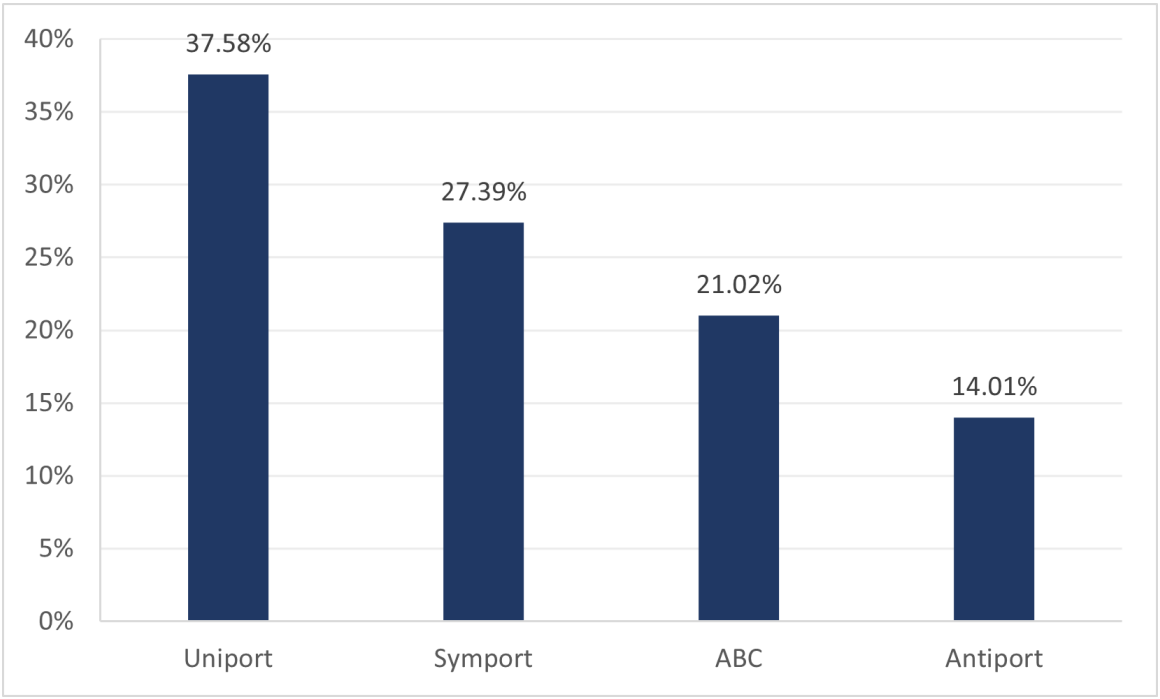


Figure 8: Type of the transport proteins in the *A. caldus* model.

7.2.2.2 *A. ferrooxidans*

Similar to the *A. caldus* model, the most common direction of the transport proteins in the *A. ferrooxidans* model is from the outside to inside totalling around 48%, as represented in figure 9. However, there is a difference of around 30% between this direction and the reversible transporters, whilst in the *A. caldus* model it is only of 6%.

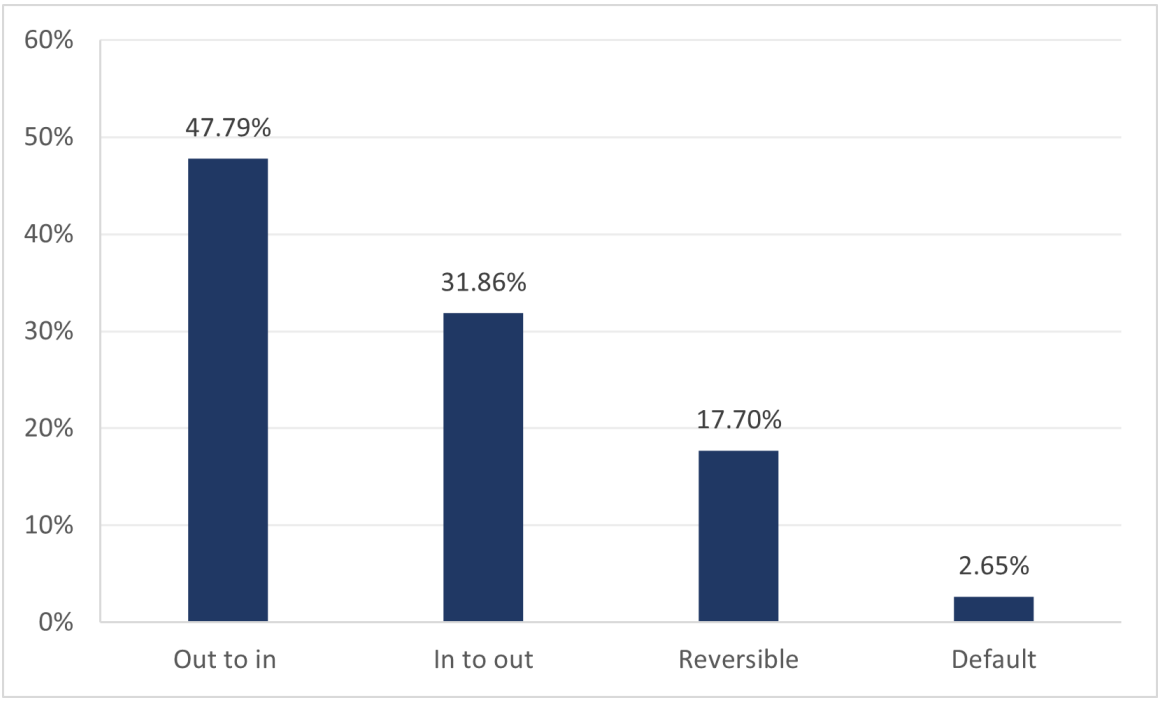


Figure 9: Direction of the transport proteins in the *A. ferrooxidans* model.

The most common type of transport proteins of the *A. ferrooxidans* model is the ABC transporters with 58% of the model's total transporters, as can be seen in figure 10. Uniport and symport proteins are also present in this model with a relative frequency of 22% and 16%, respectively. The least common type of transporters is the cofactor and redox transport proteins. As the name suggests cofactor transporters allow the movement of a cofactor whilst redox transporters facilitate the movement of metabolites against the gradient without the use of ATP.

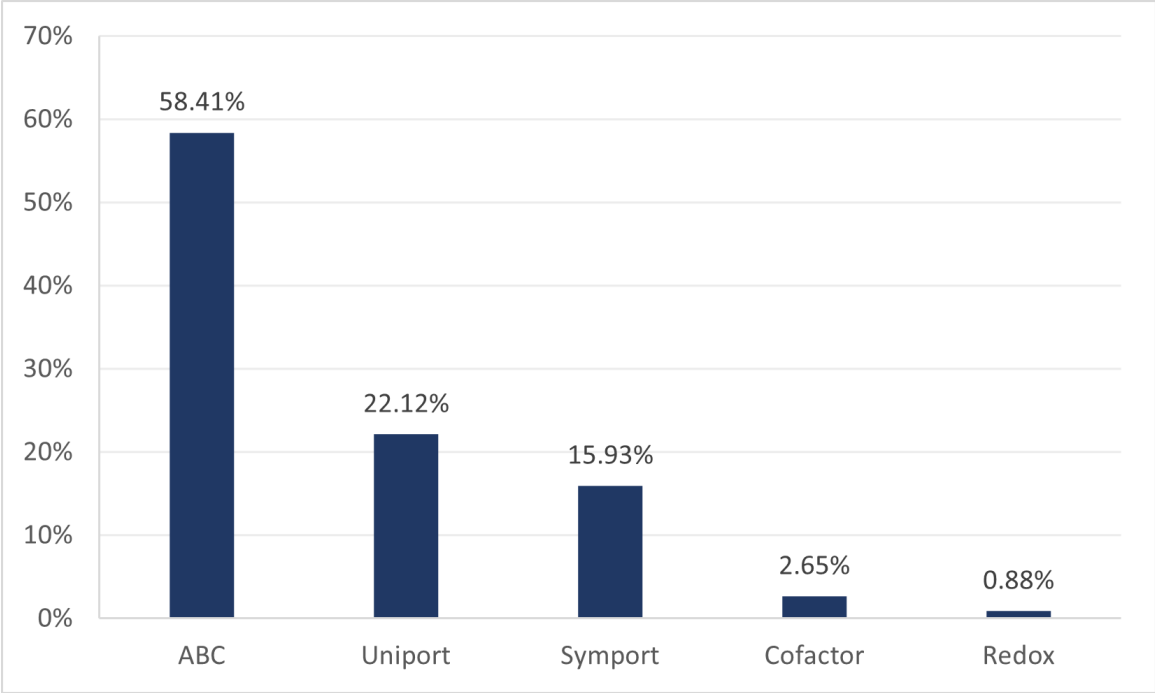


Figure 10: Type of the transport proteins in the *A. ferrooxidans* model.

7.2.3 *Gap-filling*

Network connectivity is essential for the model's performance. Dead-end metabolites lead to blocked reactions, therefore it is necessary to ensure that all metabolites in the model have at least one reaction that can consume them and another that can produce them, otherwise the model would fail the steady-state assumption. Moreover, incorrect reaction's directionality can lead to gaps in the network which will further lead to dead-end metabolites and blocked reactions [8].

As represented in table 11, *A. caldus* draft model contains 589 blocked reactions which represent around 62% of the total reactions in the network. Moreover, the draft network has 512 dead-end metabolites which represent around 48% of the total metabolites present in the network. Similarly, *A. ferrooxidans* draft model comprises 649 and 601 blocked reactions and dead-end metabolites, respectively. These numbers correspond to 63% and 50% of the total reactions and metabolites, respectively, present in the draft network. After manual curation, no gaps were present in the models' metabolic network.

Table 11: Number of blocked reactions and dead-end metabolites of the draft models.

	A. caldus	***A. ferrooxidans***
Blocked reactions	589	649
Dead-end metabolites	512	601

7.2.4 *Unbalanced reactions*

Whenever present in the model, unbalanced reactions can lead to erroneous model predictions since the principles behind mass conservation could not be applied and the model would fail the steady-state assumption. An analysis of these reactions was performed and the unbalanced reactions were corrected. As represented in figure 11, the draft *A. caldus* and *A.ferroxidans* models contained, respectively, 149 and 157 unbalanced reactions which represent around 15% of each metabolic network. After manual curation, these reactions were all corrected with the exception of drains as well as the reactions for macromolecule formation which will be naturally unbalanced due to their chemical formula being represented as 'R'.

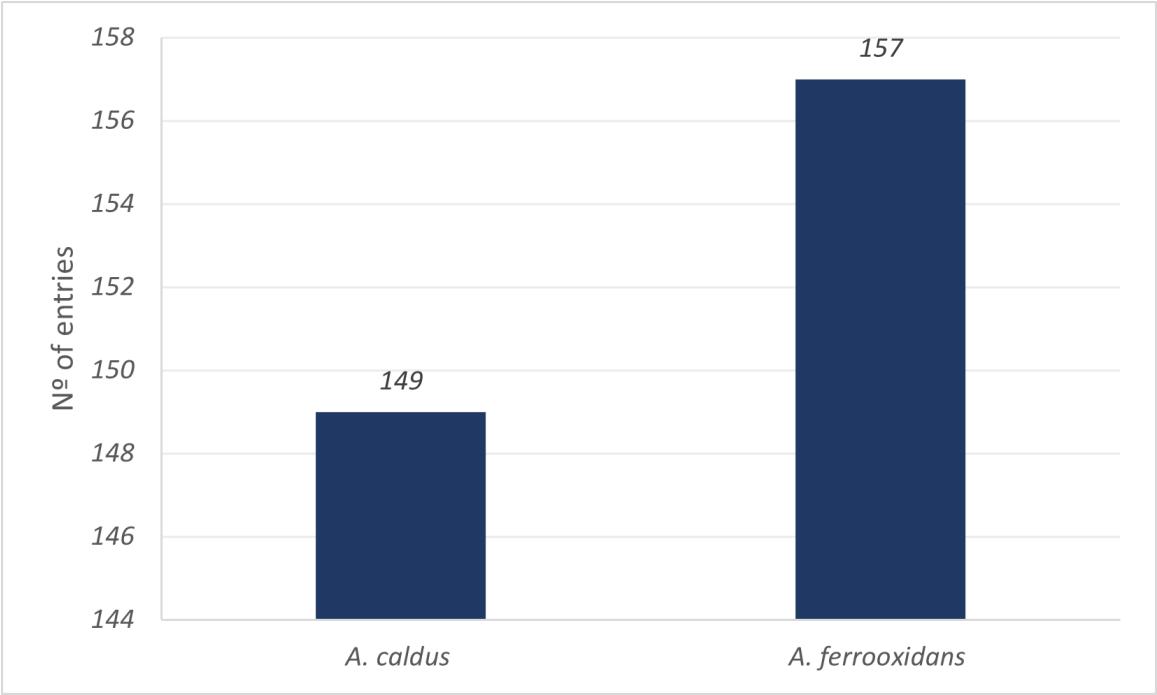


Figure 11: Comparison of the percentage of unbalanced reactions between the draft and curated models.

7.2.5 *GPR*

7.2.5.1 *A. caldus*

An analysis of the GPR associations present in the model was performed. There are 4 different situations in which the GPR rule can be made. 'Default' associations is when one gene is associated with only one reaction whilst 'Promiscuous enzymes' are situations where the same gene is associated to multiple reactions. Moreover, 'Enzymatic complex' is when the Boolean rule 'AND' is present in the GPR rule and 'Isoenzymes' are cases when the Boolean rule 'OR' is present in the GPR rule [8].

The model contains a total of 697 GPR associations. Out of this total, the model presents around 27%, 24%, 23% and 1% for 'Default' associations, 'Isoenzymes', 'Promiscuous enzymes' and 'Enzymatic complex', respectively. Lastly, around 25% of the model's reactions do not contain a GPR association. These are reactions present in the model that are not associated with a gene. These reactions are exchange reactions and macromolecule formation reactions which are naturally not associated with a gene, as well as manually inserted reactions that are known to be present in the organism, however, they could not be associated with a gene.

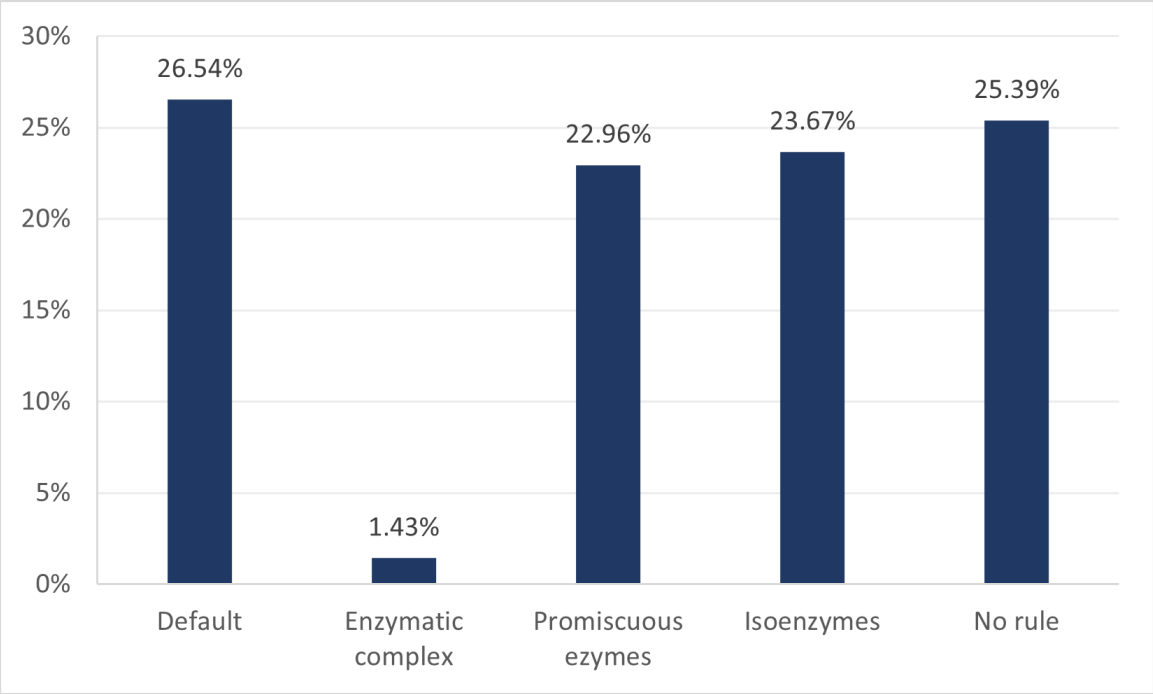


Figure 12: GPR rules in the *A. caldus* model.

7.2.5.2 *A. ferrooxidans*

Similar to the *A. caldus* model, GPR rules were associated with the respective reactions for the *A. ferrooxidans* model. A total of 644 GPR associations were made with around 28% being 'Default' associations and 22% having no GPR rule. The results for this model are similar to the results of the *A. caldus* model.

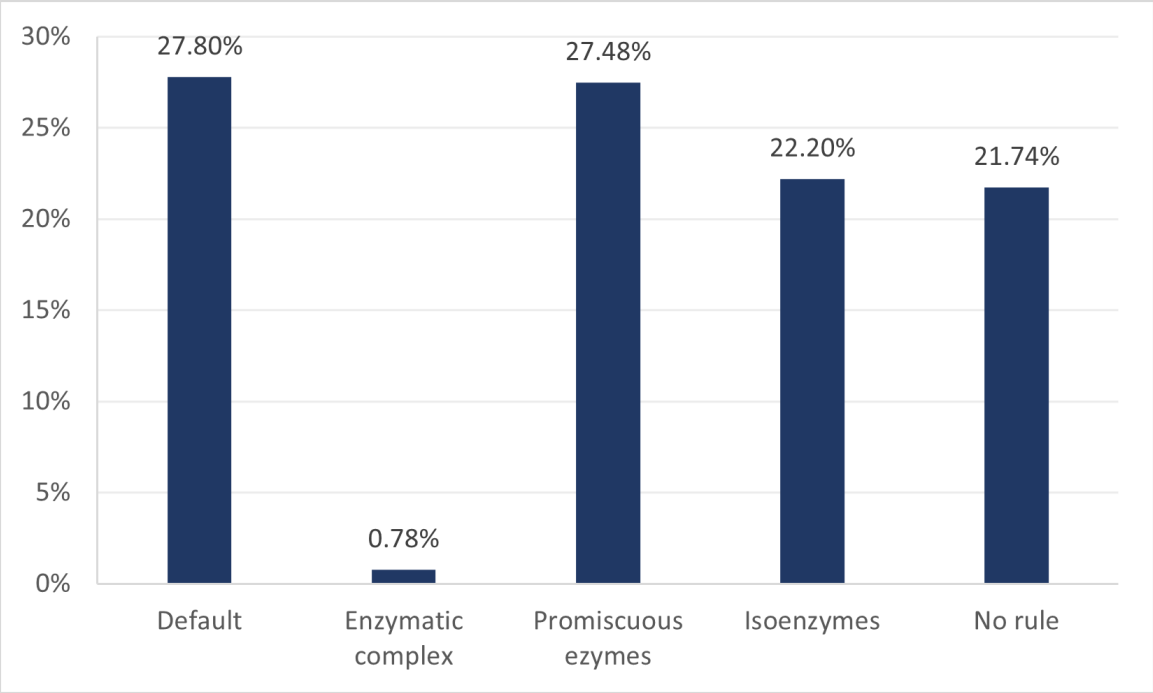


Figure 13: GPR rules in the *A. ferrooxidans* model.

7.2.6 *Model Comparison*

Several basic model characteristics were compared between the models being reconstructed and other curated models present in the BiGG database. For this, it was chosen the model of the reference organisms for the gram- positive and negative bacteria, *Bacillus subtilis* and *Escherichia coli K-12 MG1655*, respectively. In addition, it was also chosen the model of an organism that shares the same phylum of the bacteria in the study, being represented in table 12. Although the models shared a similar percentage of genes present in the model, the reconstructed models contain a slightly greater number of reactions per gene. This may be due to a lack of information on the organisms in the study as well as of the phylogenetically close organisms. A few of the reactions added to the models were unable to be associated with a gene due to a lack of information.

Table 12: Model comparison

Organism	*Bacillus subtilis*	*Mycobacterium tuberculosis*	*A. ferrooxidans*	*Escherichia coli K-12 MG1655*	*Pseudomonas putida*	*A. caldus*
Model	iYO844	iEK1008	-	iJO1366	iJN746	-
Organism Genes	4114	4402	1964	4405	5350	3186
Model Genes	844	1008	408	1367	746	416
% of genes in model	20.52%	22.90%	20.77%	31.03%	13.94%	13.06%
Reactions	1250	1226	817	2583	950	846
Nº of reactions/Gene	1.48	1.22	2.00	1.89	1.27	2.03
Metabolites	990	998	640	1805	911	646

7.3 BIOMASS EQUATION

7.3.1 *A. caldus*

As represented in table 13, the biomass composition of the *A. caldus* model is composed of 7 macromolecule entities. The respective fraction (represented in grams of molecule per gram of dry weight) of each entity was retrieved by *merlin* for DNA, RNA and protein. For the remaining macromolecules, the values were based from another model [142] and then normalised in order for the sum of each molar fraction to totalize 1. An *Exopolysaccharides (EPS)* reaction was considered, however it was discarded since the organism does not need the EPS formation in order to produce biomass.

Table 13: Biomass composition of *A. caldus*.

Macromolecule	gmolecule/gDW	Source
DNA	0,03	*merlin*
RNA	0,15	*merlin*
Protein	0,59	*merlin*
Cofactors	0,04	[142]
Carbohydrates	0,03	[142]
Cell Wall	0,07	[142]
Lipids	0,10	[142]

As represented in supplementary tables 32 and 33, the DNA and RNA equations produce diphosphate alongside the macromolecule DNA and RNA, respectively. The stoichiometry (millimole of the precursor per gram of the macromolecule) is inferred by the *merlin* e-biomass tool through genomic information.

Similar to the DNA and RNA equations, the protein reaction, as represented in supplementary table 34, takes into account the amino acids and their respective stoichiometry inferred by the tool *'e-Biomass'* through the organism proteomics data. The equation produces the macromolecule 'Protein' and H_2O as well as the tRNA of each amino acid.

The cofactor composition, as represented in supplementary table 35, was based on a set of universal cofactors for bacteria as described in [55]. The quinone was changed to Ubiquinone-8 as it is described in the literature as the main *Acidithiobacillus* quinone [152]. Moreover, the cofactor pyridoxal phosphate was removed as there was no genomic or literature evidence of the presence of this compound in the organism.

Bacteria often produce a reserve hydrocarbonate in order to be able to store carbon whenever there is more than what is required. As shown in supplementary table 36, starch was taken into account for the formation of the biomass being based in genomic evidence [15].

The macromolecule 'Cell Wall' accounts for the *lipopolysaccharide (LPS)* and peptidoglycan composition, as shown in supplementary table 37. For this KEGG's complex compounds were used in order for the flux to go through the respective pathways. Otherwise the model would produce the simple compounds individually without taking into account the whole metabolite. Moreover, the peptidoglycan composition was based on literature [153] whilst the LPS composition as well as the respective molar faction was based on the *Acidithiobacillus ferrooxidans* (*At. ferrooxidans*) model.

The 'Lipid' macromolecule contains a set of several phospholipids present in the membrane of the bacteria. As no information was found in the literature relative to the lipid composition of *A. caldus*, this composition was based on the *At. ferrooxidans* model as well as on genomic information, being represented in supplementary table 38.

Although the fatty acid composition is not directly involved in the biomass equation, fatty acids are relevant for the generation of the 'Lipid' macromolecule. Each phospholipid contains a variable number of fatty acid chains, depending on the 'R' groups of its composition. For this, a generic fatty acid was formulated as represented in supplementary table 39 with its composition being based on literature [144].

7.3.2 *A. ferrooxidans*

Similar to the *A. caldus* model, the biomass composition of *A. ferrooxidans* contains 7 macromolecule entities, as represented in table 14. The relative quantities of each macromolecule were based on the *merlin* 'e-Biomass' tool for the DNA, RNA, and Protein macromolecules, while for the remaining macromolecules was based on literature [143, 154]. Since *A. ferrooxidans* is able to grow in anaerobically, a biomass equation was formulated for this condition. The main difference between the two biomass reactions is the Cofactor macromolecule as the anaerobic biomass reaction takes into account an anaerobic Cofactor reaction.

Table 14: Biomass composition of *A. ferrooxidans*.

Macromolecule	gmolecule/gDW	Source
DNA	0,03	*merlin*
RNA	0,07	*merlin*
Protein	0,53	*merlin*
Cofactors	0,05	[143]
Carbohydrates	0,03	[154]
Cell Wall	0,22	[143]
Lipids	0,08	[143]

Both DNA and RNA equations, as represented in supplementary tables 40 and 41, are inferred through *merlin*'s 'e-biomass' tool producing diphosphate as well as DNA and RNA, respectively. The stoichiometry coefficients are based on genomic information from the assembly retrieved files.

Similar to DNA and RNA, *merlin*'s 'e-biomass tool also predicts amino acid composition through the organism's proteome. The Protein reaction is represented in supplementary table 42.

Moreover, the cofactor composition was based on a set of universal cofactors for bacteria, as described in the literature [55], which are represented in supplementary table 43. Menaquinone-9 was added to the composition as it was described in the literature of a phylogenetically close organism as the major quinone [145]. Moreover, a cofactor composition for anaerobic conditions without the heme compound was also added to the model since this compound requires oxygen to be produced.

Since no information was found in the literature for carbohydrates used by the organism, genomic information was used and the composition is represented in supplementary table 44.

Furthermore, the 'Cell Wall' macromolecule is composed by the complex compounds, as represented in supplementary table 45. Although gram-positive bacteria do not contain LPS in the cell wall, they contain teichoic and/or lipoteichoic acids depending if these are attached to the cell wall or to the membrane, respectively. Since no information was found in the literature regarding these carbohydrates chains, genomic information was used. For the peptidoglycan composition information of a related organism was used [145].

Analogously to the *A. caldus* model, the fatty acid composition is not directly involved in the biomass. However, this composition is relevant for the formation of the 'Lipid' macromolecule. The fatty acid composition, represented in supplementary table 46, was based on literature [145].

Lastly, the phospholipid composition was based on genomic information as well as on literature [155, 156, 157]. As represented in supplementary table 47, the *A. ferrooxidans* major phospholipid is phosphatidylglycerol.

7.4 MODEL VALIDATION

7.4.1 *Environmental conditions*

7.4.1.1 *Medium absence*

A first environmental condition was tested for both organisms which consisted in closing all drains, therefore there would be no uptake of any compound by the models. This test resulted in no biomass production for both organisms, as expected. This means that no ATP and carbon were being generated inside the models.

7.4.1.2 *A. caldus*

For the validation of the *A. caldus* model three environmental conditions were tested: two autotrophic conditions with different energy sources and one mixotrophic condition. The growth medium used for both aerobic conditions is represented in tables 15 and 16.

Table 15: Aerobic condition with sulfur as energy source

Reaction ID	Metabolite Name	Metabolite Formula	Lower Bound
R_Drain_H2O__extr	H2O	H_2O	-1000.0
R_EX_C00011__extr	CO2	CO_2	-0.7106
R_EX_C00007__extr	Oxygen	O_2	-1000.0
R_EX_C00087__extr	Sulfur	S	-1000.0
R_EX_C14818__extr	Fe2+	Fe	-1000.0
R_EX_C00009__extr	Orthophosphate	H_3PO_4	-1000.0
R_EX_C00014__extr	Ammonia	NH_3	-1000.0

Table 16: Aerobic condition with tetrathionate as energy source

Reaction ID	Metabolite Name	Metabolite Formula	Lower Bound
R_EX_C02084__extr	Tetrathionate	$H_2S_4O_6$	-1000.0
R_Drain_H2O__extr	H2O	H_2O	-1000.0
R_EX_C00011__extr	CO2	CO_2	-0.7106
R_EX_C00059__extr	Sulfate	H_2SO_4	-1000.0
R_EX_C00007__extr	Oxygen	O_2	-1000.0
R_EX_C14818__extr	Fe2+	Fe	-1000.0
R_EX_C00009__extr	Orthophosphate	H_3PO_4	-1000.0
R_EX_C00014__extr	Ammonia	NH_3	-1000.0

Small differences were found between the two autotrophic conditions. Regarding the requirement for the energy source, the model predicts a lesser consumption rate of tetrathionate while optimizing biomass production rate, as can be seen in tables 17 and 18. Moreover, the model predicted a production of sulfate which results from the oxidation of sulfur-based compounds which goes in agreement with experimental data [158], as represented in tables 17 and 18. Furthermore, the predicted growth rate was 0.01852 h^{-1} which goes in accordance with the literature [159, 160, 161, 162].

Table 17: Consumption and production rate of the *A. caldus* model for the autotrophic condition with sulfur as the energy source.

	Reaction ID	Metabolite name	Metabolite formula	Consumption rate (mmol/gDW/h)
Consumption	R_EX_C00007__extr	Oxygen	O2	-28.89487
	R_Drain_H2O__extr	H2O	H_2O	-25.98921
	R_EX_C00087__extr	Sulfur	S	-21.27575
	R_EX_C00011__extr	CO2	CO_2	-0.7106
	R_EX_C00014__extr	Ammonia	NH_3	-0.1913
	R_EX_C00009__extr	Orthophosphate	H_3PO_4	-0.01663
	R_EX_C14818__extr	Fe2+	Fe	-0.0001
	Reaction ID	**Metabolite name**	**Metabolite formula**	**Production rate (mmol/gDW/h)**
Production	R_EX_C00059__extr	Sulfate	H2SO4	21.27229
	R_EX_C00080__extr	H+	H	9.42567
	R_Drain_biomass__extr	Biomass	-	0.01852
	R_Drain_5_Deoxyadenosine_C05198__extr	5'-Deoxyadenosine	$C_{10}H_{13}N_5O_3$	0.0001
	R_EX_C00237__extr	CO	CO	0.0001
	R_drain_Glycolate__extr	Glycolate	$C_2H_4O_3$	0.0001

Table 18: Consumption and production rate of the *A. caldus* model for the autotrophic condition with tetrathionate as the energy source.

	Reaction ID	Metabolite name	Metabolite formula	Consumption rate (mmol/gDW/h)
Consumption	R_EX_Co0007__extr	Oxygen	O_2	-31.69491
	R_Drain_H2o__extr	H2O	H_2O	-30.26152
	R_EX_Co2084__extr	Tetrathionate	$H_2S_4O_6$	-9.49754
	R_EX_Co0011__extr	CO2	CO_2	-0.7106
	R_EX_Co0014__extr	Ammonia	NH_3	-0.1913
	R_EX_Co0009__extr	Orthophosphate	H_3PO_4	-0.01663
	R_EX_C14818__extr	Fe2+	Fe	-0.0001
	Reaction ID	**Metabolite name**	**Metabolite formula**	**Production rate (mmol/gDW/h)**
Production	R_EX_Co0059__extr	Sulfate	H_2SO_4	37.98669
	R_EX_Co0080__extr	H+	H	3.53656
	R_Drain_biomass__extr	Biomass	-	0.01852
	R_Drain_5_Deoxyadenosine_C05198__extr	5'-Deoxyadenosine	$C_{10}H_{13}N_5O_3$	0.0001
	R_EX_Co0237__extr	CO	CO	0.0001
	R_drain_Glycolate__extr	Glycolate	$C_2H_4O_3$	0.0001

A. caldus is able to grow under mixotrophic conditions using glucose as the carbon source and tetrathionate as the energy source. This was used to simulate the mixotrophic conditions with the growth medium being represented in table 19.

Table 19: Mixotrophic condition

Reaction ID	Metabolite Name	Metabolite Formula	Lower Bound
R_Drain_H2o__extr	H2O	H_2O	-1000.0
R_EX_Co0011__extr	CO2	CO_2	-0.7106
R_EX_Co0059__extr	Sulfate	H_2SO_4	-1000.0
R_EX_Co0007__extr	Oxygen	O_2	-1000.0
R_EX_Co2084__extr	Tetrathionate	$H_2S_4O_6$	-1000.0
R_EX_Co0267__extr	alpha-D-Glucose	$C_6H_{12}O_6$	-0.4438
R_EX_C14818__extr	Fe2+	Fe	-1000.0
R_EX_Co0009__extr	Orthophosphate	H_3PO_4	-1000.0
R_EX_Co0014__extr	Ammonia	NH_3	-1000.0

Small differences were found in the consumption rate of tetrathionate and oxygen between the mixotrophic and autotrophic conditions, as can be seen in tables 20 and 18. Moreover, the model predicted a growth rate of 0.08791 h^{-1} having a greater growth rate than in autotrophic

conditions which is corroborated by literature [163, 164]. This occurs since in this condition there are two carbon sources available to the model.

Table 20: Consumption and production rate of the *A. caldus* model for the mixotrophic condition.

	Reaction ID	Metabolite name	Metabolite formula	Consumption rate (mmol/gDW/h)
Consumption	R_EX_C00007__extr	Oxygen	O2	-33.17236
	R_Drain_H2O__extr	H2O	H_2O	-29.42972
	R_EX_C02084__extr	Tetrathionate	$H_2S_4O_6$	-9.93972
	R_EX_C00014__extr	Ammonia	NH_3	-0.90814
	R_EX_C00011__extr	CO2	CO_2	-0.7106
	R_EX_C00267__extr	alpha-D-Glucose	$C_6H_{12}O_6$	-0.4438
	R_EX_C00009__extr	Orthophosphate	H_3PO_4	-0.07893
	R_EX_C14818__extr	Fe2+	Fe	-0.00047
	Reaction ID	**Metabolite name**	**Metabolite formula**	**Production rate (mmol/gDW/h)**
Production	R_EX_C00059__extr	Sulfate	H2SO4	39.74245
	R_EX_C00080__extr	H+	H	4.54081
	R_Drain_biomass__extr	Biomass	-	0.08791
	R_Drain_5_Deoxyadenosine_C05198__extr	5'-Deoxyadenosine	$C_{10}H_{13}N_5O_3$	0.00047
	R_EX_C00237__extr	CO	CO	0.00047
	R_drain_Glycolate__extr	Glycolate	$C_2H_4O_3$	0.00047

7.4.1.3 *A. ferrooxidans*

Regarding the *A. ferrooxidans* model an aerobic and anaerobic condition was tested with the respective growth medium being represented in tables 21 and 23. For the aerobic condition, the model predicts a very low consumption rate of oxygen, when compared to the *A. caldus* model, as can be seen in table 22. Moreover, as represented in tables 22 and 24, the model predicts a similar growth rate between the two environmental conditions as observed in other studies [165, 166].

Table 21: Aerobic condition

Reaction ID	Metabolite Name	Metabolite Formula	Lower Bound
R_EX_C00011__extr	CO2	CO_2	-0.7106
R_EX_C00059__extr	Sulfate	H_2SO_4	-1000.0
R_Drain_Oxigen_C00007__extr	Oxygen	O_2	-1000.0
R_EX_C14818__extr	Fe2+	Fe	-1000.0
R_EX_C00009__extr	Orthophosphate	H_3PO_4	-1000.0
R_EX_C00014__extr	Ammonia	NH_3	-1000.0
R_Drain_C00001__extr	H2O	H_2O	-1000.0

Table 22: Consumption and production rate of the *A. ferrooxidans* model for the aerobic condition.

	Reaction ID	Metabolite name	Metabolite formula	Consumption rate (mmol/gDW/h)
Consumption	R_EX_C00011__extr	CO2	CO_2	-0.7106
	R_Drain_Oxigen_C00007__extr	Oxygen	O_2	-0.42744
	R_EX_C00014__extr	Ammonia	NH_3	-0.16307
	R_EX_C00009__extr	Orthophosphate	H_3PO_4	-0.01082
	R_EX_C00059__extr	Sulfate	H_2SO_4	-0.00255
	R_EX_C14818__extr	Fe2+	Fe	-0.00013
	Reaction ID	**Metabolite name**	**Metabolite formula**	**Production rate (mmol/gDW/h)**
Production	R_Drain_C00001__extr	H2O	H_2O	2.17576
	R_EX_C00080__extr	H+	H	1.28666
	R_Drain_biomass__extr	Biomass	-	0.01871
	R_Drain_C00237__extr	CO	CO	0.00013
	R_Drain_C05198__extr	5′-Deoxyadenosine	$C_{10}H_{13}N_5O_3$	0.00013

For the anaerobic condition, ferric iron was used since no information was found on the oxidation of ferrous iron without the use of oxygen. The model also predicted a slightly greater consumption rate of iron as represented in table 24.

Table 23: Anaerobic condition

Reaction ID	Metabolite Name	Metabolite Formula	Lower Bound
R_EX_C14819__extr	Fe3+	Fe	-1000.0
R_EX_C00011__extr	CO2	CO_2	-0.7106
R_EX_C00059__extr	Sulfate	H_2SO_4	-1000.0
R_EX_C00009__extr	Orthophosphate	H_3PO_4	-1000.0
R_EX_C00014__extr	Ammonia	NH_3	-1000.0
R_Drain_C00001__extr	H2O	H_2O	-1000.0

Table 24: Consumption and production rate of the *A. ferrooxidans* model for the anaerobic condition.

	Reaction ID	Metabolite name	Metabolite formula	Consumption rate (mmol/gDW/h)
Consumption	R_EX_C14819__extr	Fe3+	Fe	-1.54052
	R_EX_C00011__extr	CO2	CO_2	-0.7106
	R_EX_C00014__extr	Ammonia	NH_3	-0.16365
	R_EX_C00009__extr	Orthophosphate	H_3PO_4	-0.01101
	R_EX_C00059__extr	Sulfate	H_2SO_4	-0.0026
	Reaction ID	**Metabolite name**	**Metabolite formula**	**Production rate (mmol/gDW/h)**
Production	R_EX_C00080__extr	H+	H	2.99231
	R_EX_C14818__extr	Fe2+	Fe	1.54052
	R_Drain_C00001__extr	H2O	H_2O	1.32069
	R_Drain_biomass_anaerobic__extr	Biomass_anaerobic	-	0.01874
	R_Drain_C00237__extr	CO	CO	0.00015
	R_Drain_C05198__extr	5′-Deoxyadenosine	$C_{10}H_{13}N_5O_3$	0.00015

7.4.2 *Flux distribution*

7.4.2.1 *Carbon distribution*

The organisms shared a similar carbon metabolism which is represented in figure 14. The models fixate carbon through the Calvin-Benson cycle. The carbon will then go to glycolysis through the production of glycerate 3-phosphate generating pyruvate and small amounts of ATP.

The carbon will then continue the Calvin cycle which overlaps with the gluconeogenesis to form D-glucose 1-phosphate. The metabolite will be a precursor of UDP-glucose and ADP-glucose which will form trehalose in the *A. ferrooxidans* model and starch in the *A. caldus* model.

Moreover, D-Glyceraldehyde 3-phosphate will enter the pentose phosphate pathway to generate *5-Phospho-alpha-D-ribose 1-diphosphate (PRPP)* which is an important compound for the production of amino acids as well as nucleic acids.

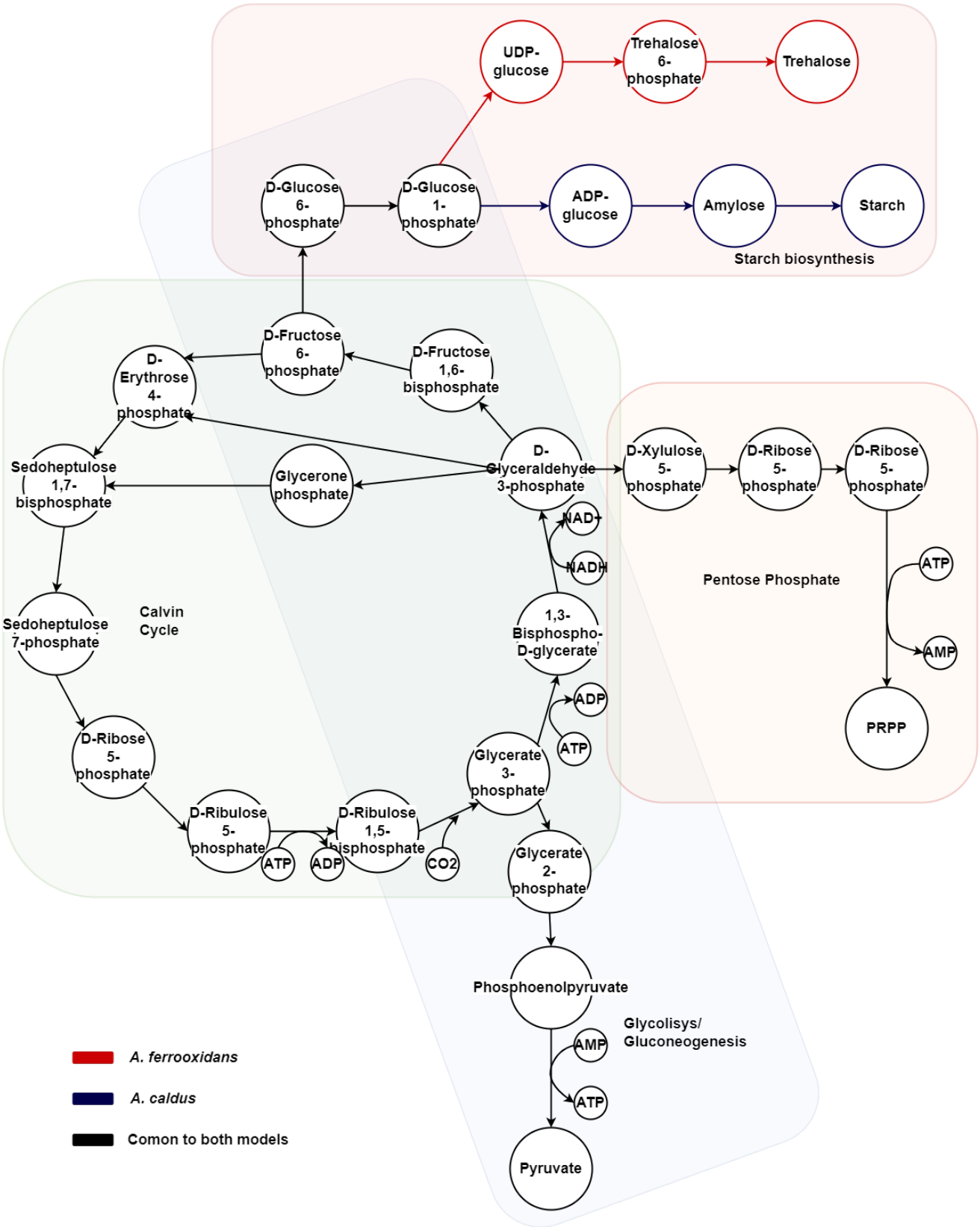


Figure 14: Carbon metabolism of the autotrophic metabolism of both models.

Moreover, the models present a complete Calvin cycle as well as the glycolysis pathway. Regarding the citrate cycle, *A. caldus* contains an incomplete cycle, as described in the literature [15], with the missing reactions being: R02164, R02570, R03316 and R00621.

A. ferrooxidans showed a complete citrate cycle, however, it is only partially used in the simulations performed. The model uses this pathway to produce succinate which is used in the oxidative phosphorilation.

7.4.2.2 *Energy Metabolism*

A. caldus is able to obtain energy through oxidation and reduction of sulfur-based compounds, as represented in figure 15. Whenever sulfur is given in the medium, the compound will enter into the periplasm and then into the cytoplasm through the respective transporters. In the cytoplasm, sulfur will produce sulfite and hydrogen sulfide. The latter will be used for the production of amino acids and will also cycle to sulfur in order to generate H^+ into the periplasm. These compounds will produce thiosulfate which will further generate sulfate realising H^+ to the periplasm. Sulfite will also produce Adenosine 5'-phosphosulfate (APS) which will further generate sulfate and small amounts of ATP going in agreement with the literature [167].

Moreover, whenever tetrathionate is given to the model instead of sulfur the metabolism is similar. After the tetrathionate reaches the cytoplasm a first reaction converting the metabolite into sulfur and thiosulfate occurs. Once tetrathionate is converted to thiosulfate and sulfur the flux distribution of the sulfur metabolism will then proceed accordingly with figure 15.

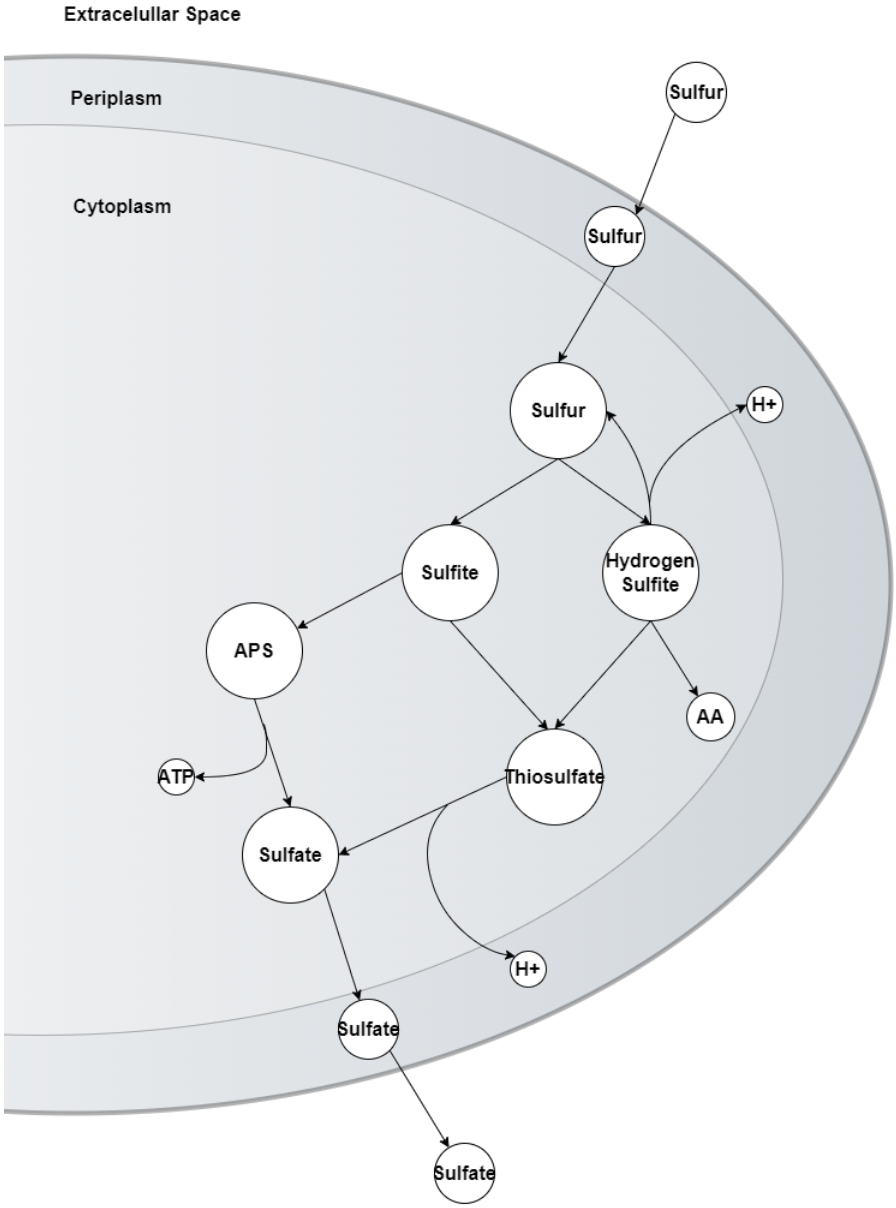


Figure 15: Sulfur metabolism from the *A. caldus* model.

7.5 COMMUNITY

The *A. caldus* and *A. ferrooxidans* community was built using reFramed. The metabolites' and reactions' names were changed depending on which organisms they are located in. Therefore, if the organism is *A. caldus* a suffix '_acaldus' is added to the name. Extracellular metabolites

have a suffix '_extr'. Furthermore, the drains of the community were created using the drains of both models.

The community's biomass is defined as follows:

$$Macromolecule_A + Macromolecule_B + ... -> Biomass_{Acaldus} + Biomass_{Community} \quad (7)$$

$$Macromolecule_A + Macromolecule_B + ... -> Biomass_{Aferrooxidans} + Biomass_{Community} \quad (8)$$

$$Community\_growth : Biomass_{Community} -> \quad (9)$$

Where:

- Equation 7 represents the biomass production of the *A. caldus* model;
- Equation 8 represents the biomass production of the *A. ferrooxidans* model;
- Equation 9 represents the community's drain for the community's biomass.

Moreover, the community model was built by treating the models as compartments within a community extracellular space, common to both models. Therefore, if a metabolite within the *A. caldus* model is exported by the community it takes the path: cytoplasm -> periplasm -> extracellular space -> outside.

7.5.1 *Environmental condition*

A community medium using the required metabolites for aerobic autotrophic growth of both organisms was used, as represented in table 25, and a pFBA maximizing the community's biomass production was performed. The community had a major consumption rate of sulfur for energy production as well as oxygen. Moreover, the community model predicted a sulfate production rate of 21.1923 mmol.gDW.h^{-1} and a biomass production rate of 0.0206 h^{-1}, as shown in table 26.

Table 25: Environmental condition used for community simulation

Reaction ID	Metabolite Name	Metabolite Formula	Lower Bound
R_EX_C00001__extr	H2O	H2O	-1000.0
R_EX_C00011__extr	CO2	CO2	-0.7106
R_EX_C00059__extr	Sulfate	H2SO4	-1000.0
R_EX_C00007__extr	Oxygen	O2	-1000.0
R_EX_C00087__extr	Sulfur	S	-1000.0
R_EX_C14818__extr	Fe2+	Fe	-1000.0
R_EX_C00009__extr	Orthophosphate	H3PO4	-1000.0
R_EX_C00014__extr	Ammonia	NH3	-1000.0

Table 26: Consumption and production rate of the community model.

	Reaction ID	Metabolite name	Metabolite formula	Consumption rate (mmol/gDW/h)
Consumption	R_EX_C00007__extr	Oxygen	O_2	-28.7637
	R_EX_C00001__extr	H2O	H_2O	-25.9426
	R_EX_C00087__extr	Sulfur	S	-21.1951
	R_EX_C00011__extr	CO2	CO_2	-0.7106
	R_EX_C00014__extr	Ammonia	NH_3	-0.1798
	R_EX_C00009__extr	Orthophosphate	H_3PO_4	-0.0119
	R_EX_C14818__extr	Fe2+	Fe	-0.0001
	Reaction ID	**Metabolite name**	**Metabolite formula**	**Production rate (mmol/gDW/h)**
Production	R_EX_C00059__extr	Sulfate	H_2SO_4	21.1923
	R_EX_C00080__extr	H+	H	9.0577
	R_EX_e_Biomass__extr	Biomass	-	0.0206
	R_EX_C05198__extr	5'-Deoxyadenosine	$C_{10}H_{13}N_5O_3$	0.0001
	R_EX_C00237__extr	CO	CO	0.0001

7.5.2 *SteadyCom*

The SteadyCom tool predicted a microbial abundance of 94.94% of the *A. ferrooxidans* and 5.06% of the *A. caldus*, as represented in table 27. The results of this tool suggest that *A. ferrooxidans* provides a major role in the community's maintenance. Although these results are in agreement with the literature [140], the dynamics of the community heavily depends on the leaching experiment being performed since the minerals present can have a big impact on the growth of community's constituents [138, 168]. Although sulfur-oxidizers do not have a primary role in

the oxidation process, some researchers have found that these organisms can outnumber the iron-oxidizers [14].

Table 27: SteadyCom output for abundance of species within the community.

Community species'	Community species' abundance (%)	Community's growth rate (h^{-1})
A. caldus	5.06	0.020508
A. ferrooxidans	94.94	

7.5.3 *SMETANA*

SMETANA outputs a probability of cross-feeding for exchanged metabolites. In table 28 are represented the predicted exchanged metabolites with over a 30% chance of cross-feeding.

Table 28: Predicted interactions between *A. ferrooxidans* and *A. caldus* by SMETANA

Donor	Bigg ID	Kegg ID	Metabolite Name	SMETANA
A. ferrooxidans	cgly	C01419	Cys-Gly	0.93
	pi	C00009	Orthophosphate	0.54
	o2	C00007	Oxygen	0.52
	fe2	C14818	Fe2+	0.5
	ppi	C00013	Diphosphate	0.46
A. caldus	o2	C00007	Oxygen	0.6
	so4	C00059	Sulfate	0.5
	tsul	C00320	Thiosulfate	0.43
	fe2	C14818	Fe2+	0.36

Notably, the tool predicts an exchange of fe2+, thiosulfate and sulfate. These metabolites are associated with the energy metabolism which takes a key role in bioleaching experiments [169].

7.5.4 *Organism interaction*

An analysis of the interactions between the two organisms within the community was performed. As represented in supplementary tables 48 and 49, *A. caldus* consumed the majority of the oxygen and CO_2, predicting a bigger energy/carbon necessities than *A. ferrooxidans*.

Moreover, the organisms exchanged small amounts of hydrogen sulfide and bicarbonate from *A. ferrooxidans* to *A. caldus*, as shown in table 29. Furthermore, from *A. caldus* to *A. ferrooxidans* the organisms exchanged phosphatidylethanolamine, hexadecanoic acid, D-Alanyl-D-alanine, phosphatidylserine, gamma-L-Glutamyl-L-cysteine, phosphatidylglycerol and phosphatidate. Whilst *A. ferrooxidans* required less oxygen and carbon than *A. caldus*, this prediction suggests

that *A. caldus* assists *A. ferrooxidans* in lipid production. In exchange, *A. ferrooxidans* aids *A. caldus* with the more complex sulfur metabolism.

Table 29: Metabolites exchanged within the community.

Donor	Metabolite ID	Metabolite Name	Production rate (|mmol/gDW/h|)
A. caldus	M_Co0669__extr	gamma-L-Glutamyl-L-cysteine	0.07383
	M_Co0249__extr	Hexadecanoic acid	0.00808
	M_Co0344__extr	Phosphatidylglycerol	0.00200
	M_Co0993__extr	D-Alanyl-D-alanine	0.00077
	M_Co0350__extr	Phosphatidylethanolamine	0.00051
	M_Co2737__extr	Phosphatidylserine	0.00005
	M_Co0416__extr	Phosphatidate	0.00001
A. ferrooxidans	M_Co0288__extr	Bicarbonate	0.20287
	M_Co0283__extr	Hydrogen sulfide	0.07102
	M_Co0013__extr	Diphosphate	0.00025

It is worth noting that while maximizing the community's biomass production rate the model predicted a zero growth rate for the *A. caldus* model. These results suggest that the main role of this organism within the community is to assist *A. ferrooxidans* whilst the latter contribute to the growth of the community.

Furthermore, a reaction consuming the biomass of both organisms was added to the model and a pFBA maximizing this reaction was performed. The stoichiometry of both metabolites was set to 1 with the reaction being represented in equation 10.

$$Biomass_{Aferrooxidans} + Biomass_{Acaldus} - > \tag{10}$$

Small differences were found, however, it was noted that by maximizing the growth rate of both organisms the community presented a lower growth rate, as represented in table 30. As can be seen in supplementary tables 50 and 51, *A. caldus* requires more carbon to grow leaving less available to *A. ferrooxidans* which consequently will have a lower growth rate affecting the community. Moreover, as can be seen in table 31, *A. caldus* contributed less to the *A. ferrooxidans* lipid production whilst the latter donated a reduced amount of hydrogen sulfide to *A. caldus* due to the greater focus on their own respective biomass production rate.

Table 30: Consumption and production rate of the community model while maximizing both organisms' growth rate.

	Reaction ID	Metabolite name	Consumption rate (|mmol/gDW/h|)	Difference
Consumption	R_EX_C00007__extr	Oxygen	28.83442	0.07072
	R_EX_C00087__extr	Sulfur	21.23885	0.04375
	R_EX_C00001__extr	H2O	25.96892	0.02632
	R_EX_C00014__extr	Ammonia	0.18585	0.00605
	R_EX_C00009__extr	Orthophosphate	0.01435	0.00245
	R_EX_C14818__extr	Fe2+	0.00012	0.00002
	R_EX_C00011__extr	CO2	0.71060	0.00000
	Reaction ID	**Metabolite name**	**Production rate (|mmol/gDW/h|)**	**Difference**
Production	R_EX_C00080__extr	H+	9.25258	0.19488
	R_EX_C00059__extr	Sulfate	21.23569	0.04339
	R_EX_C00160__extr	Glycolate	0.00005	0.00005
	R_EX_C05198__extr	5'-Deoxyadenosine	0.00012	0.00002
	R_EX_C00237__extr	CO	0.00012	0.00002
	R_Community_growth_alt	Biomass	0.00976	-0.01084

Table 31: Metabolites exchanged within the community while maximizing both organisms' growth rate.

Donor	Metabolite ID	Metabolite Name	Production rate (|mmol/gDW/h|)	Difference
A. caldus	M_C00669__extr	gamma-L-Glutamyl-L-cysteine	0.03553	-0.03830
	M_C00249__extr	Hexadecanoic acid	0.00364	-0.00444
	M_C00344__extr	Phosphatidylglycerol	0.00095	-0.00105
	M_C00993__extr	D-Alanyl-D-alanine	0.00037	-0.00041
	M_C00350__extr	Phosphatidylethanolamine	0.00024	-0.00027
	M_C02737__extr	Phosphatidylserine	0.00002	-0.00003
	M_C00416__extr	Phosphatidate	0.00001	-0.00001
A. ferrooxidans	M_C00288__extr	Bicarbonate	0.09472	-0.10815
	M_C00283__extr	Hydrogen sulfide	0.03420	-0.03682
	M_C00013__extr	Diphosphate	0.00000	-0.00025

In bioleaching experiments, *A. caldus* has an indirect contribution by reducing the pH of the medium which prevents the precipitation of iron as secondary minerals [138]. Moreover, during mineral oxidation, it is commonly released ferrous iron and thiosulfate which later decomposes into sulfur. Iron oxidizers take a key role in oxidizing the ferrous iron into ferric

iron whilst Sulfur oxidizers oxidize the elemental sulfur [168]. Whilst the oxidation of ferrous iron to ferric iron allows for a greater mineral dissolution [170], the removal of elemental sulfur keeps the minerals' catalytic surface clean. Otherwise, this accumulated sulfur would compromise the dissolution of ferric iron through the formation of unwanted passivation layers [171, 172].

Moreover, these autotrophic sulfur-oxidizers release organic carbon which is consumed by the mixotrophic/heterotrophic iron oxidizers [141]. In return, the latter generates ferric iron which stimulates mineral dissolution producing sulfur and CO_2. This CO_2 can then be used as an energy source by the autotrophic sulfur-oxidizers [138].

Although it was not possible to simulate a mineral oxidation environment the community showed a major consumption rate of sulfur exhibiting the greater sulfur oxidation capacities of *A. caldus*. Moreover, Clark and Norris demonstrated that *A. ferrooxidans* is able to efficiently utilize CO_2 whenever in CO_2 limiting conditions [136]. This was accurately predicted by the community model through the low consumption rate of CO_2 when compared to *A. caldus* which required more CO_2 and did not produce biomass. Furthermore, it was shown that *A. caldus* improves the oxidation of ferrous iron by the *A. ferrooxidans* in bioleaching conditions [138]. Although *A. ferrooxidans* exhibited a greater consumption rate of ferrous iron in the community when compared to the isolated model, the difference between the two consumption rates is minimal.

8

CONCLUSIONS AND FUTURE WORK

In this work, a highly curated GSM model of the organism *A. caldus SM-1* and *A. ferrooxidans* were reconstructed. Additionally, a community model with these GSM models was built. The manual curation efforts of the reconstruction were based on genomic information, literature, phylogenetically related organisms and biological databases for the models to be as accurate as possible. The *A. caldus* model can grow in autotrophic and mixotrophic conditions while in the presence of oxygen. Moreover, the *A. ferrooxidans* model can grow in aerobic and anaerobic autotrophic conditions. Furthermore, the community model can simulate the phenotypic behaviour of the organisms and their dynamics under aerobic autotrophic conditions, which can serve as a guide for microbial interactions and in bioleaching experiments.

The carbon metabolism of both models was thoroughly analysed. The models fixate carbon through a complete Calvin cycle which aids the glycolysis/gluconeogenesis pathway, which is complete in both models. Special attention was given to the energy metabolism of *A. caldus* as is well described in the literature. Accordingly, the model allows for the uptake of sulfur and tetrathionate which results in the production of ATP and sulfate. Furthermore, the *A. caldus* model presents an incomplete citrate cycle, whilst the *A. ferrooxidans* model predicts the partial use of this cycle to generate succinate, which will be further used in the oxidative phosphorylation for the production of ATP.

The main objectives of this study were to reconstruct the models and analyse the interactions and general behaviour of *A. caldus* and *A. ferrooxidans* within the community. The community model predicted a significant consumption rate of sulfur, highlighting the greater sulfur-oxidizing capacities of *A. caldus*, with SteadyCom predicting a greater microbial abundance of *A. ferrooxidans*. Moreover, while maximizing the community's biomass production rate, the model predicted an *A. ferrooxidans*' growth rate of 0.0206 h^{-1} while *A. caldus* did not grow. Yet, it was observed that *A. ferrooxidans* presented a greater growth rate in the community when compared to the single model simulations. Additionally, the organism interactions' analysis showed that *A. caldus* exchanged lipid-production related compounds whilst *A. ferrooxidans* donated hydrogen sulfide. Moreover, while maximizing the growth rate of both organisms within the community, a lower growth rate of the community was observed. *A. caldus* showed a greater carbon consumption rate, which consequently leaves less available to *A. ferrooxidans*

resulting in a lower growth rate of the latter. It seems that the community's growth rate is affected due to a greater focus on their respective individual growth rate. These results suggest that *A. caldus* has a role in the community in which it assists *A. ferrooxidans* growth while the latter provides a vital role in the community growth rate. The models and respective genome files, *merlin* workspaces and the in-house built simulation scripts are available at `https://github.com/ruirnunes/MicrobialCommunity`.

Although this study provides several insights into the interaction mechanisms of these two bacteria, other approaches can be assessed. Both models can be further curated as the biomass equation of the organisms was based on phylogenetically related organisms. Whilst this does not affect the models' performance, an accurate biomass reaction would improve the robustness of the models. Moreover, the lack of literature regarding these organisms imposed a problem throughout both models' reconstruction. As new insights into the organisms' metabolism are discovered, they can be incorporated into the models improving their accuracy. Moreover, several bioleaching critical conditions cannot be simulated in these models. However, mineral oxidation reactions and their kinetics considered and used with these models. It would be interesting to analyse, for example, the community model's behaviour while optimizing mineral oxidation reactions with energy-limiting sources.

BIBLIOGRAPHY

[1] Sebastián N. Mendoza, Brett G. Olivier, Douwe Molenaar, and Bas Teusink. A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biology*, 2019.

[2] Siddharth Hegde and Lisa Kaltenegger. Colors of extreme exo-earth environments. *Astrobiology*, 2013.

[3] Oscar Dias, Miguel Rocha, Eugénio C. Ferreira, and Isabel Rocha. Reconstructing genome-scale metabolic models with merlin. *Nucleic Acids Research*, 2015.

[4] Hendrik Rohn, Anja Hartmann, Astrid Junker, Björn H. Junker, and Falk Schreiber. FluxMap: A VANTED add-on for the visual exploration of flux distributions in biological networks. *BMC Systems Biology*, 2012.

[5] Luana Presta, Emanuele Bosi, Leila Mansouri, Lenie Dijkshoorn, Renato Fani, and Marco Fondi. Constraint-based modeling identifies new putative targets to fight colistin-resistant A. baumannii infections. *Scientific Reports*, 2017.

[6] Yong Mei Zhang, Stephen W. White, and Charles O. Rock. Inhibiting bacterial fatty acid synthesis, 2006.

[7] Deok Sun Lee, Henry Burd, Jiangxia Liu, Eivind Almaas, Olaf Wiest, Albert László Barabási, Zoltán N. Oltvai, and Vinayak Kapatral. Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple Staphylococcus aureus genomes identify novel antimicrobial drug targets. *Journal of Bacteriology*, 2009.

[8] Ines Thiele and Bernhard Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 2010.

[9] Felipe Sarmiento, Rocío Peralta, and Jenny M. Blamey. Cold and hot extremozymes: Industrial relevance and current trends, 2015.

[10] Lesley-Ann Giddings and David J. Newman. Bioactive Compounds from Terrestrial Extremophiles. 2015.

[11] Ram Karan, Melinda D. Capes, and Shiladitya DasSarma. Function and biotechnology of extremophilic enzymes in low water activity, 2012.

[12] Joseph Gomes and Walter Steiner. The biocatalytic potential of extremophiles and extremozymes, 2004.

[13] Lissette Valenzuela, An Chi, Simon Beard, Alvaro Orell, Nicolas Guiliani, Jeff Shabanowitz, Donald F. Hunt, and Carlos A. Jerez. Genomics, metagenomics and proteomics in biomining microorganisms, 2006.

[14] Douglas E. Rawlings. Heavy metal mining using microbes, 2002.

[15] Xiao Yan You, Xu Guo, Hua Jun Zheng, Ming Jiang Zhang, Li Jun Liu, Yong Qiang Zhu, Baoli Zhu, Sheng Yue Wang, Guo Ping Zhao, Ansgar Poetsch, Cheng Ying Jiang, and Shuang Jiang Liu. Unraveling the Acidithiobacillus caldus complete genome and its central metabolisms for carbon assimilation. *Journal of Genetics and Genomics*, 2011.

[16] Alicia Clum, Matt Nolan, Elke Lang, Tijana Glavina del Rio, Hope Tice, Alex Copeland, Cheng Jan-Fang, Susan Lucas, Feng Chen, David Bruce, Lynne Goodwin, Sam Pitluck, Natalia Ivanova, Konstantinos Mavromatis, Natalia Mikhailova, Amrita Pati, Amy Chen, Krishna Palaniappan, Markus Göker, Stefan Spring, Miriam Land, Loren Hauser, Chang Yun-Juan, Cynthia C. Jeffries, Patrick Chain, Jim Bristow, Jonathan A. Eisen, Victor Markowitz, Philip Hugenholtz, Nikos C. Kyrpides, Klenk Hans-Peter, and Alla Lapidus. Complete genome sequence of Acidimicrobium ferrooxidans type strain (ICP T). *Standards in Genomic Sciences*, 2009.

[17] Lisa Rizzetto and Duccio Cavalieri. Friend or foe: Using systems biology to elucidate interactions between fungi and their hosts, 2011.

[18] B. Palsson. The challenges of in silico biology, 2000.

[19] Hiroaki Kitano. Systems biology: A brief overview, 2002.

[20] Roberta Mustacchi, Stefan Hohmann, and Jens Nielsen. Yeast systems biology to unravel the network of life, 2006.

[21] Cristiana Gomes de Oliveira Dal'Molin and Lars Keld Nielsen. Plant genome-scale reconstruction: from single cell to multi-tissue modelling and omics analyses, 2018.

[22] Dina Petranovic and Goutham N. Vemuri. Impact of yeast systems biology on industrial biotechnology. *Journal of Biotechnology*, 2009.

[23] Kate Campbell, Jianye Xia, and Jens Nielsen. The Impact of Systems Biology on Bioprocessing, 2017.

[24] Beste Turanli, Cheng Zhang, Woonghee Kim, Rui Benfeitas, Mathias Uhlen, Kazim Yalcin Arga, and Adil Mardinoglu. Discovery of therapeutic agents for prostate cancer using genome-scale metabolic modeling and drug repositioning. *EBioMedicine*, 2019.

[25] Bernhard Palsson. Metabolic systems biology, 2009.

[26] Isabel Rocha, Jochen Förster, and Jens Nielsen. Design and application of genome-scale reconstructed metabolic models. *Methods in Molecular Biology*, 2007.

[27] Markus W. Covert, Eric M. Knight, Jennifer L. Reed, Markus J. Herrgard, and Bernhard O. Palsson. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 2004.

[28] Masaru Tomita. Whole-cell simulation: A grand challenge of the 21st century, 2001.

[29] Adam M. Feist, Markus J. Herrgård, Ines Thiele, Jennie L. Reed, and Bernhard Palsson. Reconstruction of biochemical networks in microorganisms, 2009.

[30] Oscar Dias and Isabel Rocha. Systems Biology in Fungi. In *Molecular Biology of Food and Water Borne Mycotoxigenic and Mycotic Fungi*. 2015.

[31] Guenter Stoesser, Mary Ann Tuli, Rodrigo Lopez, and Peter Sterk. The EMBL Nucleotide Sequence Database, 1999.

[32] Richa Agarwala, Tanya Barrett, Jeff Beck, Dennis A. Benson, Colleen Bollin, Evan Bolton, Devon Bourexis, J. Rodney Brister, Stephen H. Bryant, Kathi Canese, Mark Cavanaugh, Chad Charowhas, Karen Clark, Ilya Dondoshansky, Michael Feolo, Lawrence Fitzpatrick, Kathryn Funk, Lewis Y. Geer, Viatcheslav Gorelenkov, Alan Graeff, Wratko Hlavina, Brad Holmes, Mark Johnson, Brandi Kattman, Viatcheslav Khotomlianski, Avi Kimchi, Michael Kimelman, Masato Kimura, Paul Kitts, William Klimke, Alex Kotliarov, Sergey Krasnov, Anatoliy Kuznetsov, Melissa J. Landrum, David Landsman, Stacy Lathrop, Jennifer M. Lee, Carl Leubsdorf, Zhiyong Lu, Thomas L. Madden, Aron Marchler-Bauer, Adriana Malheiro, Peter Meric, Ilene Karsch-Mizrachi, Anatoly Mnev, Terence Murphy, Rebecca Orris, James Ostell, Christopher O'Sullivan, Vasuki Palanigobu, Anna R. Panchenko, Lon Phan, Borys Pierov, Kim D. Pruitt, Kurt Rodarmer, Eric W. Sayers, Valerie Schneider, Conrad L. Schoch, Gregory D. Schuler, Stephen T. Sherry, Karanjit Siyan, Alexandra Soboleva, Vladimir Soussov, Grigory Starchenko, Tatiana A. Tatusova, Francoise Thibaud-Nissen, Kamen Todorov, Bart W. Trawick, Denis Vakatov, Minghong Ward, Eugene Yaschenko, Aleksandr Zasypkin, and Kerry Zbicz. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 2018.

[33] Peter D. Karp, Christos A. Ouzounis, Caroline Moore-Kochlacs, Leon Goldovsky, Pallavi Kaipa, Dag Ahrén, Sophia Tsoka, Nikos Darzentas, Victor Kunin, and Núria López-Bigas. Expansion of the Biocyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research*, 2005.

[34] Maren Lang, Michael Stelzer, and Dietmar Schomburg. BKM-react, an integrated biochemical reaction database. *BMC Biochemistry*, 2011.

[35] I. Schomburg. BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, 2002.

[36] Panu Artimo, Manohar Jonnalagedda, Konstantin Arnold, Delphine Baratin, Gabor Csardi, Edouard De Castro, Séverine Duvaud, Volker Flegel, Arnaud Fortier, Elisabeth Gasteiger, Aurélien Grosdidier, Céline Hernandez, Vassilios Ioannidis, Dmitry Kuznetsov, Robin Liechti, Sébastien Moretti, Khaled Mostaguir, Nicole Redaschi, Grégoire Rossier, Ioannis Xenarios, and Heinz Stockinger. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, 2012.

[37] Hiroyuki Ogata, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, and Minoru Kanehisa. KEGG: Kyoto encyclopedia of genes and genomes, 1999.

[38] P.D. Karp. The MetaCyc metabolic pathway database. *Metabolic Engineering in the Post Genomic Era*, 2002.

[39] Ulrike Wittig, Martin Golebiewski, Renate Kania, Olga Krebs, Saqib Mir, Andreas Weidemann, Stefanie Anstein, Jasmin Saric, and Isabel Rojas. SABIO-RK: Integration and curation of reaction kinetics data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006.

[40] Milton H. Saier, Vamsee S. Reddy, Dorjee G. Tamang, and Åke Västermark. The transporter classification database. *Nucleic Acids Research*, 2014.

[41] Amos Bairoch, Brigitte Boeckmann, Serenella Ferro, Elisabeth Gasteiger, Nicole Redaschi, Cathy H. Wu, Hongzhan Huang, Darren A. Natale, Winona C. Barker, Lai Su L. Yeh, Rolf Apweiler, Rodrigo Lopez, Michelle Magrane, Maria J. Martin, and Claire O'Donovan. UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 2004.

[42] Zachary A. King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A. Lerman, Ali Ebrahim, Bernhard O. Palsson, and Nathan E. Lewis. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 2016.

[43] Chen Li, Marco Donizelli, Nicolas Rodriguez, Harish Dharuri, Lukas Endler, Vijayalakshmi Chelliah, Lu Li, Enuo He, Arnaud Henry, Melanie I. Stefan, Jacky L. Snoep, Michael Hucka, Nicolas Le Novère, and Camille Laibe. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 2010.

[44] Christopher S. Henry, Matthew Dejongh, Aaron A. Best, Paul M. Frybarger, Ben Linsay, and Rick L. Stevens. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, 2010.

[45] Alan J Barrett. Enzyme Nomenclature. Recommendations 1992. *European Journal of Biochemistry*, 1995.

[46] M. H. Saier. A Functional-Phylogenetic Classification System for Transmembrane Solute Transporters. *Microbiology and Molecular Biology Reviews*, 2000.

[47] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 1990.

[48] S. R. Eddy. Multiple alignment using hidden Markov models. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 1995.

[49] Christof Francke, Roland J. Siezen, and Bas Teusink. Reconstructing the metabolic network of a bacterium from its genome, 2005.

[50] Xueyang Feng, You Xu, Yixin Chen, and Yinjie J. Tang. MicrobesFlux: a web platform for drafting metabolic models from the KEGG database. *BMC Systems Biology*, 2012.

[51] Cristiana Gomes de Oliveira Dal'Molin, Lake Ee Quek, Pedro A. Saa, and Lars K. Nielsen. A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. *Frontiers in Plant Science*, 2015.

[52] Olof Emanuelsson, Henrik Nielsen, Søren Brunak, and Gunnar Von Heijne. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, 2000.

[53] Paul Horton, Keun Joon Park, Takeshi Obayashi, Naoya Fujita, Hajime Harada, C. J. Adams-Collier, and Kenta Nakai. WoLF PSORT: Protein localization predictor. *Nucleic Acids Research*, 2007.

[54] Sophia Santos and Isabel Rocha. Estimation of biomass composition from genomic and transcriptomic information. *Journal of integrative bioinformatics*, 2016.

[55] Joana C. Xavier, Kiran Raosaheb Patil, and Isabel Rocha. Integration of Biomass Formulations of Genome-Scale Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in Prokaryotes. *Metabolic Engineering*, 2017.

[56] Adam M. Feist, Christopher S. Henry, Jennifer L. Reed, Markus Krummenacker, Andrew R. Joyce, Peter D. Karp, Linda J. Broadbelt, Vassily Hatzimanikatis, and Bernhard Palsson. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 2007.

[57] Iman Famili, Jochen Förster, Jens Nielsen, and Bernhard O. Palsson. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences of the United States of America*, 2003.

[58] Helga David, Mats Åkesson, and Jens Nielsen. Reconstruction of the central carbon metabolism of Aspergillus niger. *European Journal of Biochemistry*, 2003.

[59] Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastián N. Mendoza, Anne Richelle, Almut Heinken, Hulda S. Haraldsdóttir, Jacek Wachowiak, Sarah M. Keating, Vanja Vlasov, Stefania Magnusdóttir, Chiam Yu Ng, German Preciat, Alise Žagare, Siu H.J. Chan, Maike K. Aurich, Catherine M. Clancy, Jennifer Modamio, John T. Sauls, Alberto Noronha, Aarash Bordbar, Benjamin Cousins, Diana C. El Assal, Luis V. Valcarcel, Iñigo Apaolaza, Susan Ghaderi, Masoud Ahookhosh, Marouen Ben Guebila, Andrejs Kostromins, Nicolas Sompairac, Hoai M. Le, Ding Ma, Yuekai Sun, Lin Wang, James T. Yurkovich, Miguel A.P. Oliveira, Phan T. Vuong, Lemmer P. El Assal, Inna Kuperstein, Andrei Zinovyev, H. Scott Hinton, William A. Bryant, Francisco J. Aragón Artacho, Francisco J. Planes, Egils Stalidzans, Alejandro Maass, Santosh Vempala, Michael Hucka, Michael A. Saunders, Costas D. Maranas, Nathan E. Lewis, Thomas Sauter, Bernhard Palsson, Ines Thiele, and Ronan M.T. Fleming. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, 2019.

[60] Isabel Rocha, Paulo Maia, Pedro Evangelista, Paulo Vilaça, Simão Soares, José P. Pinto, Jens Nielsen, Kiran R. Patil, Eugénio C. Ferreira, and Miguel Rocha. OptFlux: An open-source software platform for in silico metabolic engineering. *BMC Systems Biology*, 2010.

[61] Jeffrey D Orth, Ines Thiele, and Bernhard O Palsson. What is flux balance analysis?, 2010.

[62] Jeremy S. Edwards, Rafael U. Ibarra, and Bernhard O. Palsson. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, 2001.

[63] Claire R. Shen and James C. Liao. Synergy as design principle for metabolic engineering of 1-propanol production in Escherichia coli. *Metabolic Engineering*, 2013.

[64] Gregory N Stephanopoulos, Aristos a Aristidou, and Jens Nielsen. *Metabolic Engineering: Principles and Methodologies*. 1998.

[65] Namrata Tomar and Rajat K. De. Comparing methods for metabolic network analysis and an application to metabolic engineering, 2013.

[66] Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. Current state and applications of microbial genome-scale metabolic models, 2017.

[67] Christopher S. Henry, Hans C. Bernstein, Pamela Weisenhorn, Ronald C. Taylor, Joon Yong Lee, Jeremy Zucker, and Hyun Seob Song. Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction. *Journal of Cellular Physiology*, 2016.

[68] Stephen R. Lindemann, Hans C. Bernstein, Hyun Seob Song, Jim K. Fredrickson, Matthew W. Fields, Wenying Shou, David R. Johnson, and Alexander S. Beliaev. Engineering microbial consortia for controllable outputs, 2016.

[69] Hans C. Bernstein and Ross P. Carlson. Microbial consortia engineering for cellular factories: In vitro to in silico systems, 2012.

[70] Sahar Abubucker, Nicola Segata, Johannes Goll, Alyxandria M. Schubert, Jacques Izard, Brandi L. Cantarel, Beltran Rodriguez-Mueller, Jeremy Zucker, Mathangi Thiagarajan, Bernard Henrissat, Owen White, Scott T. Kelley, Barbara Methé, Patrick D. Schloss, Dirk Gevers, Makedonka Mitreva, and Curtis Huttenhower. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Computational Biology*, 2012.

[71] José P. Faria, Tahmineh Khazaei, Janaka N. Edirisinghe, Pamela Weisenhorn, Samuel M. D. Seaver, Neal Conrad, Nomi Harris, Matthew DeJongh, and Christopher S. Henry. Constructing and Analyzing Metabolic Flux Models of Microbial Communities. 2016.

[72] Hao Wang, Simonas Marcišauskas, Benjamín J. Sánchez, Iván Domenzain, Daniel Hermansson, Rasmus Agren, Jens Nielsen, and Eduard J. Kerkhoven. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. *PLoS Computational Biology*, 2018.

[73] Octavio Perez-Garcia, Gavin Lear, and Naresh Singhal. Metabolic network modeling of microbial interactions in natural and engineered environmental systems, 2016.

[74] Changdai Gu, Gi Bae Kim, Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. Current status and applications of genome-scale metabolic models, 2019.

[75] Méziane Aite, Marie Chevallier, Clémence Frioux, Camille Trottier, Jeanne Got, María Paz Cortés, Sebastián N. Mendoza, Grégory Carrier, Olivier Dameron, Nicolas Guillaudeux, Mauricio Latorre, Nicolás Loira, Gabriel V. Markov, Alejandro Maass, and Anne Siegel. Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models. *PLoS Computational Biology*, 2018.

[76] Mark Hanemaaijer, Brett G. Olivier, Wilfred F.M. Röling, Frank J. Bruggeman, and Bas Teusink. Model-based quantification of metabolic interactions from dynamic microbial-community data. *PLoS ONE*, 2017.

[77] Peter D. Karp, Mario Latendresse, Suzanne M. Paley, Markus Krummenacker, Quang D. Ong, Richard Billington, Anamika Kothari, Daniel Weaver, Thomas Lee, Pallavi Subhraveti, Aaron Spaulding, Carol Fulcher, Ingrid M. Keseler, and Ron Caspi. Pathway tools version 19.0 update: Software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, 2016.

[78] José P. Faria, Miguel Rocha, Isabel Rocha, and Christopher S. Henry. Methods for automated genome-scale metabolic model reconstruction, 2018.

[79] Vinay Satish Kumar, Madhukar S. Dasika, and Costas D. Maranas. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*, 2007.

[80] Ines Thiele, Nikos Vlassis, and Ronan M.T. Fleming. FASTGAPFILL: Efficient gap filling in metabolic networks. *Bioinformatics*, 2014.

[81] Nikos Vlassis, Maria Pires Pacheco, and Thomas Sauter. Fast Reconstruction of Compact Context-Specific Metabolic Network Models. *PLoS Computational Biology*, 2014.

[82] Mario Latendresse. Efficiently gap-filling reaction networks. *BMC Bioinformatics*, 2014.

[83] Daniel Segrè, Dennis Vitkup, and George M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2002.

[84] Tomer Shlomi, Orner Berkman, and Eytan Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, 2005.

[85] Stefan Schuster, David A. Fell, and Thomas Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 2000.

[86] Kiran Raosaheb Patil, Isabel Rocha, Jochen Förster, and Jens Nielsen. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, 2005.

[87] Miguel Rocha, Paulo Maia, Rui Mendes, José P. Pinto, Eugénio C. Ferreira, Jens Nielsen, Kiran Raosaheb Patil, and Isabel Rocha. Natural computation meta-heuristics for the in silico optimization of microbial strains. *BMC Bioinformatics*, 2008.

[88] Anthony P. Burgard, Priti Pharkya, and Costas D. Maranas. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnology and Bioengineering*, 2003.

[89] Vítor Pereira, Fernando Cruz, and Miguel Rocha. MEWpy: a computational strain optimization workbench in Python. *Bioinformatics*, 2021.

[90] Siu Hung Joshua Chan, Margaret N. Simons, and Costas D. Maranas. SteadyCom: Predicting microbial abundances while ensuring community stability. *PLoS Computational Biology*, 2017.

[91] Aleksej Zelezniak, Sergej Andrejev, Olga Ponomarova, Daniel R. Mende, Peer Bork, and Kiran Raosaheb Patil. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 2015.

[92] Christian Lieven, Moritz Emanuel Beber, Brett G Olivier, Frank T Bergmann, Meric Ataman, Parizad Babaei, Jennifer A Bartell, Lars M Blank, Siddharth Chauhan, Kevin Correia, et al. Memote: A community-driven effort towards a standardized genome-scale metabolic model test suite. *BioRxiv*, page 350991, 2018.

[93] Iwei Yeh, Theodor Hanekamp, Sophia Tsoka, Peter D. Karp, and Russ B. Altman. Computational analysis of Plasmodium falciparum metabolism: Organizing genomic information to facilitate drug discovery. *Genome Research*, 2004.

[94] Segun Fatumo, Kitiporn Plaimas, Jan Philipp Mallm, Gunnar Schramm, Ezekiel Adebiyi, Marcus Oswald, Roland Eils, and Rainer König. Estimating novel potential drug targets of Plasmodium falciparum by analysing the metabolic network of knock-out strains in silico. *Infection, Genetics and Evolution*, 2009.

[95] Aarash Bordbar, Nathan E. Lewis, Jan Schellenberger, Bernhard Palsson, and Neema Jamshidi. Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Molecular Systems Biology*, 2010.

[96] Natalie C. Duarte, Scott A. Becker, Neema Jamshidi, Ines Thiele, Monica L. Mo, Thuy D. Vo, Rohith Srivas, and Bernhard Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*, 2007.

[97] Neema Jamshidi and Bernhard Palsson. Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Systems Biology*, 2007.

[98] Jeffrey D. Orth, Tom M. Conrad, Jessica Na, Joshua A. Lerman, Hojung Nam, Adam M. Feist, and Bernhard Palsson. A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011. *Molecular Systems Biology*, 2011.

[99] Jung Eun Yang, Si Jae Park, Won Jun Kim, Hyeong Jun Kim, Bumjoon J. Kim, Hyuk Lee, Jihoon Shin, and Sang Yup Lee. One-step fermentative production of aromatic polyesters from glucose by metabolically engineered Escherichia coli strains. *Nature Communications*, 2018.

[100] Pranjul Mishra, Na Rae Lee, Meiyappan Lakshmanan, Minsuk Kim, Byung Gee Kim, and Dong Yup Lee. Genome-scale model-driven strain design for dicarboxylic acid production in Yarrowia lipolytica. *BMC Systems Biology*, 2018.

[101] Gabriela I. Guzmán, José Utrilla, Sergey Nurk, Elizabeth Brunk, Jonathan M. Monk, Ali Ebrahim, Bernhard O. Palsson, and Adam M. Feist. Model-driven discovery of underground metabolic functions in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 2015.

[102] Matthew A. Oberhardt, Raphy Zarecki, Leah Reshef, Fangfang Xia, Miquel Duran-Frigola, Rachel Schreiber, Christopher S. Henry, Nir Ben-Tal, Daniel J. Dwyer, Uri Gophna, and Eytan Ruppin. Systems-Wide Prediction of Enzyme Promiscuity Reveals a New Underground Alternative Route for Pyridoxal 5'-Phosphate Production in E. coli. *PLoS Computational Biology*, 2016.

[103] Emanuele Bosi, Jonathan M. Monk, Ramy K. Aziz, Marco Fondi, Victor Nizet, and Bernhard Palsson. Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, 2016.

[104] Jonathan M. Monk, Pep Charusanti, Ramy K. Aziz, Joshua A. Lerman, Ned Premyodhin, Jeffrey D. Orth, Adam M. Feist, and Bernhard Palsson. Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. *Proceedings of the National Academy of Sciences of the United States of America*, 2013.

[105] Yara Seif, Erol Kavvas, Jean Christophe Lachance, James T. Yurkovich, Sean Paul Nuccio, Xin Fang, Edward Catoiu, Manuela Raffatellu, Bernhard O. Palsson, and Jonathan M. Monk. Genome-scale metabolic reconstructions of multiple Salmonella strains reveal serovar-specific metabolic traits. *Nature Communications*, 2018.

[106] Sylvain Prigent, Jens C. Nielsen, Jens C. Frisvad, and Jens Nielsen. Reconstruction of 24 Penicillium genome-scale metabolic models shows diversity based on their secondary metabolism. *Biotechnology and Bioengineering*, 2018.

[107] Sergey Stolyar, Steve Van Dien, Kristina Linnea Hillesland, Nicolas Pinel, Thomas J. Lie, John A. Leigh, and David A. Stahl. Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology*, 2007.

[108] Alan R. Pacheco, Mauricio Moel, and Daniel Segrè. Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nature Communications*, 2019.

[109] Dorines Rosario, Rui Benfeitas, Gholamreza Bidkhori, Cheng Zhang, Mathias Uhlen, Saeed Shoaie, and Adil Mardinoglu. Understanding the representative gut microbiota

dysbiosis in metformin-treated Type 2 diabetes patients using genome-scale metabolic modeling. *Frontiers in Physiology*, 2018.

[110] Manish Kumar, Boyang Ji, Parizad Babaei, Promi Das, Dimitra Lappa, Girija Ramakrishnan, Todd E. Fox, Rashidul Haque, William A. Petri, Fredrik Bäckhed, and Jens Nielsen. Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: Lessons from genome-scale metabolic modeling. *Metabolic Engineering*, 2018.

[111] Douglas McCloskey, Bernhard Palsson, and Adam M. Feist. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli, 2013.

[112] Kelly Botero, Silvia Restrepo, and Andres Pinzón. A genome-scale metabolic model of potato late blight suggests a photosynthesis suppression mechanism. *BMC genomics*, 2018.

[113] Benjamín J Sánchez, Cheng Zhang, Avlant Nilsson, Petri-Jaan Lahtvee, Eduard J Kerkhoven, and Jens Nielsen. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, 2017.

[114] Matteo Mori, Terence Hwa, Olivier C. Martin, Andrea De Martino, and Enzo Marinari. Constrained Allocation Flux Balance Analysis. *PLoS Computational Biology*, 2016.

[115] Edward J. O'Brien, Joshua A. Lerman, Roger L. Chang, Daniel R. Hyduke, and Bernhard Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 2013.

[116] Joshua A. Lerman, Daniel R. Hyduke, Haythem Latif, Vasiliy A. Portnoy, Nathan E. Lewis, Jeffrey D. Orth, Alexandra C. Schrimpe-Rutledge, Richard D. Smith, Joshua N. Adkins, Karsten Zengler, and Bernhard O. Palsson. In silico method for modelling metabolism and gene product expression at genome scale. *Nature Communications*, 2012.

[117] Yara Seif, Jonathan M. Monk, Nathan Mih, Hannah Tsunemoto, Saugat Poudel, Cristal Zuniga, Jared Broddrick, Karsten Zengler, and Bernhard O. Palsson. A computational knowledge-base elucidates the response of Staphylococcus aureus to different media types. *PLoS Computational Biology*, 2019.

[118] Zachary A. King, Colton J. Lloyd, Adam M. Feist, and Bernhard O. Palsson. Next-generation genome-scale models for metabolic engineering, 2015.

[119] Prachi Singh, Kunal Jain, Chirayu Desai, Onkar Tiwari, and Datta Madamwar. Microbial community dynamics of extremophiles/extreme environment. In *Microbial Diversity in the Genomic Era*, pages 323–332. Elsevier, 2019.

[120] A. Chien, D. B. Edgar, and J. M. Trela. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *Journal of Bacteriology*, 1976.

[121] Fernanda P. Cid, Joaquín I. Rilling, Steffen P. Graether, Leon A. Bravo, María De La Luz Mora, and Milko A. Jorquera. Properties and biotechnological applications of ice-binding proteins in bacteria, 2016.

[122] T. De Vrije, G. De Haas, G. B. Tan, E. R.P. Keijsers, and P. A.M. Claassen. Pretreatment of Miscanthus for hydrogen production by Thermotoga elfii. In *International Journal of Hydrogen Energy*, 2002.

[123] James A. Coker. Extremophiles and biotechnology: Current uses and prospects, 2016.

[124] Shaozhou Zhu, Dawei Song, Cuiyu Gong, Pingwah Tang, Xingzhou Li, Jianjun Wang, and Guojun Zheng. Biosynthesis of nucleoside analogues via thermostable nucleoside phosphorylase. *Applied Microbiology and Biotechnology*, 2013.

[125] Noura Raddadi, Ameur Cherif, Daniele Daffonchio, Mohamed Neifar, and Fabio Fava. Biotechnological applications of extremophiles, extremozymes and extremolytes, 2015.

[126] Milko Alberto Jorquera, Steffen P Graether, and Fumito Maruyama. Bioprospecting and biotechnology of extremophiles. *Frontiers in bioengineering and biotechnology*, 7:204, 2019.

[127] Reed M Stubbendieck, Carol Vargas-Bautista, and Paul D Straight. Bacterial communities: interactions to scale. *Frontiers in microbiology*, 7:1234, 2016.

[128] Nathan I. Johns, Tomasz Blazejewski, Antonio L.C. Gomes, and Harris H. Wang. Principles for designing synthetic microbial communities, 2016.

[129] Yuqiao Dong, Baota Jiang, Dake Xu, Chengying Jiang, Qi Li, and Tingyue Gu. Severe microbiologically influenced corrosion of S32654 super austenitic stainless steel by acid producing bacterium Acidithiobacillus caldus SM-1. *Bioelectrochemistry*, 2018.

[130] John E. Aston, William A. Apel, Brady D. Lee, and Brent M. Peyton. Toxicity of select organic acids to the slightly thermophilic Acidophile acidithiobacillus caldus. *Environmental Toxicology and Chemistry*, 2009.

[131] Yuqiao Dong, Jiaqi Li, Dake Xu, Guangling Song, Dan Liu, Haipeng Wang, M. Saleem Khan, Ke Yang, and Fuhui Wang. Investigation of microbial corrosion inhibition of Cu-bearing 316L stainless steel in the presence of acid producing bacterium Acidithiobacillus caldus SM-1. *Journal of Materials Science and Technology*, 2021.

[132] Hong Bo Zhou, Wei Min Zeng, Zhi Feng Yang, Ying Jian Xie, and Guan Zhou Qiu. Bioleaching of chalcopyrite concentrate by a moderately thermophilic culture in a stirred tank reactor, 2009.

[133] Weimin Zeng, Guanzhou Qiu, Hongbo Zhou, Juanhua Peng, Miao Chen, Su Nee Tan, Weiliang Chao, Xueduan Liu, and Yansheng Zhang. Community structure and dynamics of the free and attached microorganisms during moderately thermophilic bioleaching of chalcopyrite concentrate. *Bioresource Technology*, 2010.

[134] H. R. Watling. The bioleaching of sulphide minerals with emphasis on copper sulphides - A review. *Hydrometallurgy*, 2006.

[135] Sabrina Hedrich and D. Barrie Johnson. Aerobic and anaerobic oxidation of hydrogen by acidophilic bacteria, 2013.

[136] Darren A. Clark and Paul R. Norris. Acidimicrobium ferrooxidans gen. nov., sp. nov.: Mixed-culture ferrous iron oxidation with Sulfobacillus species. *Microbiology*, 1996.

[137] E. L J Watkin, S. E. Keeling, F. A. Perrot, D. W. Shiers, M. L. Palmer, and H. R. Watling. Metals tolerance in moderately thermophilic isolates from a spent copper sulfide heap, closely related to Acidithiobacillus caldus, Acidimicrobium ferrooxidans and Sulfobacillus thermosulfidooxidans. *Journal of Industrial Microbiology and Biotechnology*, 2009.

[138] Masahito Tanaka, Yuta Yamaji, Yuken Fukano, Kazuhiko Shimada, Jun Ichiro Ishibashi, Tsuyoshi Hirajima, Keiko Sasaki, Mitsuru Sawada, and Naoko Okibe. Biooxidation of Gold-, Silver, and Antimony-Bearing Highly Refractory Polymetallic Sulfide Concentrates, and its Comparison with Abiotic Pretreatment Techniques. *Geomicrobiology Journal*, 2015.

[139] Keishi Oyama, Tsuyoshi Hirajima, Keiko Sasaki, Hajime Miki, and Naoko Okibe. Mechanism of silver-catalyzed bioleaching of enargite concentrate. In *Solid State Phenomena*, 2017.

[140] Keishi Oyama, Kazuhiko Shimada, Jun ichiro Ishibashi, Hajime Miki, and Naoko Okibe. Silver-catalyzed bioleaching of enargite concentrate using moderately thermophilic microorganisms. *Hydrometallurgy*, 2018.

[141] Yusei Masaki, Tsuyoshi Hirajima, Keiko Sasaki, Hajime Miki, and Naoko Okibe. Microbiological Redox Potential Control to Improve the Efficiency of Chalcopyrite Bioleaching. *Geomicrobiology Journal*, 2018.

[142] Miguel A. Campodonico, Daniela Vaisman, Jean F. Castro, Valeria Razmilic, Francesca Mercado, Barbara A. Andrews, Adam M. Feist, and Juan A. Asenjo. Acidithiobacillus ferrooxidans's comprehensive model driven analysis of the electron transfer metabolism and synthetic strain design for biomining applications. *Metabolic Engineering Communications*, 2016.

[143] You Kwan Oh, Bernhard O. Palsson, Sung M. Park, Christophe H. Schilling, and Radhakrishnan Mahadevan. Genome-scale reconstruction of metabolic network in Bacillus

subtilis based on high-throughput phenotyping and gene essentiality data. *Journal of Biological Chemistry*, 2007.

[144] Claudia Knief, Karlheinz Altendorf, and André Lipski. Linking autotrophic activity in environmental samples with specific bacterial taxa by detection of 13C-labelled fatty acids. *Environmental Microbiology*, 2003.

[145] Midori Kurahashi, Yukiyo Fukunaga, Yayoi Sakiyama, Shigeaki Harayama, and Akira Yokota. Iamia majanohamensis gen. nov., sp. nov., an actinobacterium isolated from sea cucumber Holothuria edulis, and proposal of lamiaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, 2009.

[146] Erol S. Kavvas, Yara Seif, James T. Yurkovich, Charles Norsigian, Saugat Poudel, William W. Greenwald, Sankha Ghatak, Bernhard O. Palsson, and Jonathan M. Monk. Updated and standardized genome-scale reconstruction of Mycobacterium tuberculosis H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC Systems Biology*, 2018.

[147] Hongwu Ma and An Ping Zeng. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, 2003.

[148] CG Bryan, CS Davis-Belmar, N Van Wyk, MK Fraser, D Dew, GF Rautenbach, and STL Harrison. The effect of co2 availability on the growth, iron oxidation and co2-fixation rates of pure cultures of leptospirillum ferriphilum and acidithiobacillus ferrooxidans. *Biotechnology and bioengineering*, 109(7):1693–1703, 2012.

[149] Roberto A. Bobadilla Fazzini, Maria Paz Cortés, Leandro Padilla, Daniel Maturana, Marko Budinich, Alejandro Maass, and Pilar Parada. Stoichiometric modeling of oxidation of reduced inorganic sulfur compounds (Riscs) in Acidithiobacillus thiooxidans. *Biotechnology and Bioengineering*, 2013.

[150] Ann P. Wood and Don P. Kelly. Mechanisms of sugar transport by Thiobacillus A2. *Archives of Microbiology*, 1982.

[151] European Molecular Biology Laboratory Daniel Machado. ReFramed documentation — ReFramed 1.1.0 documentation `https://reframed.readthedocs.io/en/latest/`, 2019. (Accessed on 09/12/2020).

[152] Donovan P. Kelly and Ann P. Wood. Reclassification of some species of Thiobacillus to the newly designated genera Acidithiobacillus gen. nov., Halothiobacillus gen. nov. and Thermithiobacillus gen. nov. *International Journal of Systematic and Evolutionary Microbiology*, 2000.

[153] J. Baj and Z. Markiewicz. Characterization of the cell wall murein of Thiobacillus versutus. *Acta microbiologica Polonica*, 1988.

[154] Ana Paula Oliveira, Jens Nielsen, and Jochen Förster. Modeling Lactococcus lactis using a genome-scale flux model. *BMC Microbiology*, 2005.

[155] Takashi Itoh, Kaoru Yamanoi, Takuji Kudo, Moriya Ohkuma, and Tomonori Takashina. Aciditerrimonas ferrireducens gen. nov., sp. nov., an iron-reducing thermoacidophilic actinobacterium isolated from a solfataric field. *International Journal of Systematic and Evolutionary Microbiology*, 2011.

[156] Long Jin, Hangsak Huy, Kwang Kyu Kim, Hyung Gwan Lee, Hee Sik Kim, Chi Yong Ahn, and Hee Mock Oh. Aquihabitans daechungensis gen. nov., sp. nov., an actinobacterium isolated from reservoir water. *International Journal of Systematic and Evolutionary Microbiology*, 2013.

[157] Iftikhar Ahmed, Takuji Kudo, Saira Abbas, Muhammad Ehsan, Takao Iino, Toru Fujiwara, and Moriya Ohkuma. Cellulomonas pakistanensis sp. nov., a moderately halotolerant Actinobacteria. *International Journal of Systematic and Evolutionary Microbiology*, 2014.

[158] H. W. Ryu, H. S. Moon, E. Y. Lee, K. S. Cho, and H. Choi. Leaching Characteristics of Heavy Metals from Sewage Sludge by Acidithiobacillus thiooxidans MET . *Journal of Environmental Quality*, 2003.

[159] S. M. Mousavi, S. Yaghmaei, F. Salimi, and A. Jafari. Influence of process variables on biooxidation of ferrous sulfate by an indigenous Acidithiobacillus ferrooxidans. Part I: Flask experiments. *Fuel*, 2006.

[160] Takami Kai, Tatsuki Nagano, Tomonori Fukumoto, Masaya Nakajima, and Takeshige Takahashi. Autotrophic growth of Acidithiobacillus ferrooxidans by oxidation of molecular hydrogen using a gas-liquid contactor. *Bioresource Technology*, 2007.

[161] G. Cabrera, J. M. Gómez, and D. Cantero. Kinetic study of ferrous sulphate oxidation of Acidithiobacillus ferrooxidans in the presence of heavy metal ions. *Enzyme and Microbial Technology*, 2005.

[162] A. Rashidi, J. Safdari, R. Roosta-Azad, and S. Zokaei-Kadijani. Modeling of uranium bioleaching by Acidithiobacillus ferrooxidans. *Annals of Nuclear Energy*, 2012.

[163] Julie Mason and Don P. Kelly. Mixotrophic and autotrophic growth of Thiobacillus acidophilus on tetrathionate. *Archives of Microbiology*, 1988.

[164] Amiya Kumar Patel. Isolation and Characterization of Thiobacillus ferrooxidans from Coal Acid Mine Drainage. Technical report, 2010.

[165] D. Barrie Johnson, Paula Bacelar-Nicolau, Naoko Okibe, Angharad Thomas, and Kevin B. Hallberg. Ferrimicrobium acidiphilum gen. nov., sp. nov. and Ferrithrix thermotolerans gen. nov., sp. nov.: Heterotrophic, iron-oxidizing, extremely acidophilic actinobacteria. *International Journal of Systematic and Evolutionary Microbiology*, 2009.

[166] Adibah Yahya, Kevin B. Hallberg, and D. Barrie Johnson. Iron and carbon metabolism by a mineral-oxidizing Alicyclobacillus-like bacterium. *Archives of Microbiology*, 2008.

[167] Rui Wang, Jian Qiang Lin, Xiang Mei Liu, Xin Pang, Cheng Jia Zhang, Chun Long Yang, Xue Yan Gao, Chun Mao Lin, Ya Qing Li, Yang Li, Jian Qun Lin, and Lin Xu Chen. Sulfur oxidation in the acidophilic autotrophic Acidithiobacillus spp., 2019.

[168] Naoko Okibe and D. Barrie Johnson. Biooxidation of pyrite by defined mixed cultures of moderately thermophilic acidophiles in pH-controlled bioreactors: Significance of microbial interactions. *Biotechnology and Bioengineering*, 2004.

[169] D. B. JOHNSON. Biodiversity and interactions of acidophiles: Key to understanding and optimizing microbial processing of ores and concentrates. *Transactions of Nonferrous Metals Society of China (English Edition)*, 2008.

[170] M. B. Stott, D. C. Sutton, H. R. Watling, and P. D. Franzmann. Comparative leaching of Chalcopyrite by selected acidophilic Bacteria and Archaea. *Geomicrobiology Journal*, 2003.

[171] Mark Dopson and E. Börje Lindström. Potential role of Thiobacillus camus in arsenopyrite bioleaching. *Applied and Environmental Microbiology*, 1999.

[172] J. Barrett, D. K. Ewart, M. N. Hughes, and R. K. Poole. Chemical and biological pathways in the bacterial oxidation of arsenopyrite. *FEMS Microbiology Reviews*, 1993.

A

SUPPORT MATERIAL

Table 32: DNA equation of *A. caldus*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gDNA)
Reactants		
dCTP	C00458	0.844
dATP	C00131	0.663
dGTP	C00286	1.017
dTTP	C00459	0.701
Products		
e-DNA	-	1.000
Diphosphate	C00013	3.226

Table 33: RNA equation of *A. caldus*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gRNA)
Reactants		
GTP	C00044	0.994
CTP	C00063	0.759
UTP	C00075	0.565
ATP	C00002	0.765
Products		
e-RNA	-	1.000
Diphosphate	C00013	3.084

Table 34: Protein equation of *A. caldus*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gProtein)
Reactants		
L-Leucyl-tRNA	C02047	1.056
Glycyl-tRNA(Gly)	C02412	0.725
L-Asparaginyl-tRNA(Asn)	C03402	0.225
Glutaminyl-tRNA	C02282	0.391
L-Isoleucyl-tRNA(Ile)	C03127	0.425
L-Arginyl-tRNA(Arg)	C02163	0.703
L-Aspartyl-tRNA(Asp)	C02984	0.457
L-Cysteinyl-tRNA(Cys)	C03125	0.083
L-Seryl-tRNA(Ser)	C02553	0.484
L-Tryptophanyl-tRNA(Trp)	C03512	0.158
L-Valyl-tRNA(Val)	C02554	0.63
L-Histidyl-tRNA(His)	C02988	0.235
L-Glutamyl-tRNA(Glu)	C02987	0.52
L-Lysyl-tRNA	C01931	0.272
L-Methionyl-tRNA	C02430	0.198
L-Threonyl-tRNA(Thr)	C02992	0.419
L-Phenylalanyl-tRNA(Phe)	C03511	0.325
L-Prolyl-tRNA(Pro)	C02702	0.49
L-Alanyl-tRNA	C00886	1.044
L-Tyrosyl-tRNA(Tyr)	C02839	0.233
Products		
tRNA(Thr)	C01651	0.419
tRNA(Ser)	C01650	0.484
tRNA(Lys)	C01646	0.272
tRNA(Glu)	C01641	0.52
tRNA(Met)	C01647	0.198
tRNA(Gln)	C01640	0.391
tRNA(Phe)	C01648	0.325
tRNA(Tyr)	C00787	0.233
tRNA(Trp)	C01652	0.158
tRNA(Gly)	C01642	0.725
tRNA(Ala)	C01635	1.044
tRNA(Arg)	C01636	0.703
tRNA(His)	C01643	0.235

Continued on next page

Table 34 – continued from previous page

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gProtein)
tRNA(Asn)	C01637	0.225
tRNA(Pro)	C01649	0.49
tRNA(Val)	C01653	0.63
tRNA(Asp)	C01638	0.457
tRNA(Ile)	C01644	0.425
tRNA(Leu)	C01645	1.056
tRNA(Cys)	C01639	0.083
e-Protein	-	1
H_2O	C00001	9.074

Table 35: Cofactor equation of *A. caldus*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gCofactor)
Reactants		
Ubiquinone-8	C17569	0.145
NADH	C00004	0.145
NAD+	C00003	0.145
CoA	C00010	0.145
S-Adenosyl-L-methionine	C00019	0.145
FAD	C00016	0.145
Heme	C00032	0.145
Glutathione	C00051	0.145
Thiamin diphosphate	C00068	0.145
FMN	C00061	0.145
Riboflavin	C00255	0.145
Tetrahydrofolate	C00101	0.145
Products		
e-Cofactor	-	1.000

Table 36: Carbohydrate equation of *A. caldus*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gCarbohydrate)
Reactants		
Starch	C00369	0.76
Products		
e-Carbohydrate	-	1

Table 37: Cell Wall equation of *A. caldus*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gCellWall)
Reactants		
KDO2-lipid A	C06026	0.197
ADP-L-glycero-beta-D-manno-heptose	C06398	0.138
Undecaprenyl-diphospho-N-acetylmuramoyl-(N-acetylglucosamine)-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alanine	C05898	0.246
Products		
e-CellWall	-	1.000

Table 38: Lipid equation of *A. caldus*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gLipid)
Reactants		
Phosphatidylglycerol	C00344	0.665
Phosphatidylethanolamine	C00350	0.270
Phosphatidyl-N-methylethanolamine	C01241	0.773
Cardiolipin	C05980	0.090
Products		
e-Lipid	-	1.000

Table 39: Fatty acid equation of *A. caldus*.

Metabolite	KEGG ID	Stoichiometry Coefficient (gPrecursor/gFatty acid)
Reactants		
Dodecanoic acid	C02679	0.047
Octadecanoic acid	C01530	0.421
Hexadecanoic acid	C00249	0.457
Tetradecanoic acid	C06424	0.074
Products		
e-FattyAcid	-	1.000

Table 40: DNA equation of *A. ferrooxidans*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gDNA)
Reactants		
dCTP	C00458	1.105
dATP	C00131	0.514
dGTP	C00286	1.105
dTTP	C00459	0.512
Products		
e-DNA	-	1.000
Diphosphate	C00013	3.236

Table 41: RNA equation of *A. ferrooxidans*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gRNA)
Reactants		
ATP	C00002	0.585
UTP	C00075	0.420
GTP	C00044	1.089
CTP	C00063	0.994
Products		
Diphosphate	C00013	3.087
e-RNA	-	1.000

Table 42: Protein equation of *A. caldus*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gProtein)
Reactants		
L-Leucyl-tRNA	C02047	1.014
Glycyl-tRNA(Gly)	C02412	0.832
L-Asparaginyl-tRNA(Asn)	C03402	0.148
L-Methionyl-tRNA	C02430	0.146
Glutaminyl-tRNA	C02282	0.239
L-Isoleucyl-tRNA(Ile)	C03127	0.383
L-Cysteinyl-tRNA(Cys)	C03125	0.057
L-Seryl-tRNA(Ser)	C02553	0.515
L-Lysyl-tRNA	C01931	0.120
L-Glutamyl-tRNA(Glu)	C02987	0.572
L-Histidyl-tRNA(His)	C02988	0.206
L-Tryptophanyl-tRNA(Trp)	C03512	0.134
L-Valyl-tRNA(Val)	C02554	0.926
L-Threonyl-tRNA(Thr)	C02992	0.504
L-Arginyl-tRNA(Arg)	C02163	0.815
L-Aspartyl-tRNA(Asp)	C02984	0.508
L-Phenylalanyl-tRNA(Phe)	C03511	0.254
L-Prolyl-tRNA(Pro)	C02702	0.511
L-Alanyl-tRNA	C00886	1.288
L-Tyrosyl-tRNA(Tyr)	C02839	0.169
Products		
tRNA(Trp)	C01652	0.134
tRNA(His)	C01643	0.206
tRNA(Ile)	C01644	0.383
tRNA(Leu)	C01645	1.014
tRNA(Phe)	C01648	0.254
tRNA(Thr)	C01651	0.504
tRNA(Ala)	C01635	1.288
tRNA(Pro)	C01649	0.511
tRNA(Arg)	C01636	0.815
tRNA(Asn)	C01637	0.148
tRNA(Asp)	C01638	0.508
tRNA(Tyr)	C00787	0.169
tRNA(Cys)	C01639	0.057

Continued on next page

Table 42 – continued from previous page

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gProtein)
tRNA(Gln)	C01640	0.239
tRNA(Ser)	C01650	0.515
tRNA(Glu)	C01641	0.572
tRNA(Gly)	C01642	0.832
tRNA(Lys)	C01646	0.120
tRNA(Val)	C01653	0.926
tRNA(Met)	C01647	0.146
H_2O	C00001	9.341
e-Protein	-	1.000

Table 43: Cofactor equation of *A. ferrooxidans*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gCofactor)
Reactants		
NADH	C00004	0.144
NAD+	C00003	0.144
CoA	C00010	0.144
S-Adenosyl-L-methionine	C00019	0.144
Pyridoxal phosphate	C00018	0.144
FAD	C00016	0.144
Heme	C00032	0.144
Glutathione	C00051	0.144
Thiamin diphosphate	C00068	0.144
FMN	C00061	0.144
Riboflavin	C00255	0.144
Tetrahydrofolate	C00101	0.144
Menaquinone-9	-	0.144
Products		
e-Cofactor	-	1.000

Table 44: Carbohydrate equation of *A. ferrooxidans*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gCarbohydrate)
Reactants		
alpha,alpha-Trehalose	C01083	2.921
Products		
e-Carbohydrate	-	1.000

Table 45: Cell wall equation of *A. ferrooxidans*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gCellWall)
Reactants		
Lipoteichoic acid	C20898	0.137
Undecaprenyl-diphospho-N-acetylmuramoyl-(N-acetylglucosamine)-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alanine	C05898	0.167
Products		
e-CellWall	-	1.000

Table 46: Fatty acid equation of *A. ferrooxidans*.

Metabolite	KEGG ID	Stoichiometry Coefficient (gPrecursor/gFatty acid)
Reactants		
Octadecanoic acid	C01530	0.092
Hexadecanoic acid	C00249	0.899
Tetradecanoic acid	C06424	0.009
Products		
e-FattyAcid	-	1.000

Table 47: Lipid equation of *A. ferrooxidans*.

Metabolite	KEGG ID	Stoichiometry Coefficient (mmol/gLipid)
Reactants		
Phosphatidylglycerol	C00344	0.865
Phosphatidylethanolamine	C00350	0.324
Phosphatidylserine	C02737	0.031
Cardiolipin	C05980	0.008
Products		
e-Lipid	-	1.000

Table 48: Consumption and production rate of the *A. caldus* model within the community.

	Reaction ID	Metabolite name	Consumption rate (|mmol/gDW/h|)
Consumption	M_C00007__extr	Oxygen	28.76333
	M_C00001__extr	H2O	25.942588
	M_C00087__extr	Sulfur	21.195146
	M_C00011__extr	CO2	0.545833
	M_C00288__extr	HCO3-	0.203908
	M_C00014__extr	Ammonia	0.149759
	M_C00283__extr	Hydrogen sulfide	0.071021
	M_C00013__extr	Diphosphate	0.00128
	Reaction ID	**Metabolite name**	**Production rate (|mmol/gDW/h|)**
Production	M_C00059__extr	Sulfate	21.19234
	M_C00080__extr	H+	8.95115
	M_C00669__extr	gamma-L-Glutamyl-L-cysteine	0.073829
	M_C00249__extr	Hexadecanoic acid	0.008083
	M_C00344__extr	Phosphatidylglycerol	0.002000
	M_C00993__extr	D-Alanyl-D-alanine	0.000772
	M_C00350__extr	Phosphatidylethanolamine	0.000509
	M_C02737__extr	Phosphatidylserine	0.000048
	M_C00416__extr	Phosphatidate	0.000012

Table 49: Consumption and production rate of the *A. ferrooxidans* model within the community.

	Reaction ID	Metabolite name	Consumption rate (|mmol/gDW/h|)
Consumption	M_Co0011__extr	CO_2	0.164767
	M_Co0669__extr	gamma-L-Glutamyl-L-cysteine	0.073829
	M_Co0014__extr	Ammonia	0.030018
	M_Co0009__extr	Orthophosphate	0.01193
	M_Co0249__extr	Hexadecanoic acid	0.008083
	M_Co0344__extr	Phosphatidylglycerol	0.002000
	M_Co0993__extr	D-Alanyl-D-alanine	0.000772
	M_Co0350__extr	Phosphatidylethanolamine	0.000509
	M_Co0007__extr	Oxygen	0.000367
	M_C14818__extr	Fe2+	0.00015
	M_Co2737__extr	Phosphatidylserine	0.000048
	M_Co0416__extr	Phosphatidate	0.000012
	Reaction ID	**Metabolite name**	**Production rate (|mmol/gDW/h|)**
Production	M_Co0288__extr	Bicarbonate	0.20391
	M_Co0080__extr	H+	0.10658
	M_Co0283__extr	Hydrogen sulfide	0.07102
	M_e_Biomass__extr	e-Biomass	0.02063
	M_Co0013__extr	Diphosphate	0.00128
	M_Co5198__extr	5′-Deoxyadenosine	0.00015
	M_Co0237__extr	CO	0.00015

Table 50: Consumption and production rate of the *A. caldus* model within the community while maximizing both organisms' growth rate.

	Reaction ID	Metabolite name	Consumption rate (|mmol/gDW/h|)	Difference
Consumption	M_Co0011__extr	CO2	0.63629	0.09046
	M_Co0007__extr	Oxygen	28.83424	0.07091
	M_Co0087__extr	Sulfur	21.23885	0.04370
	M_Co0001__extr	H2O	25.96892	0.02633
	M_Co0014__extr	Ammonia	0.17286	0.02310
	M_Co0009__extr	Orthophosphate	0.00992	0.00992
	M_C14818__extr	Fe2+	0.00005	0.00005
	M_Co0013__extr	Diphosphate	0.00000	-0.00128
	M_Co0283__extr	Hydrogen sulfide	0.03420	-0.03682
	M_Co0288__extr	HCO3-	0.09472	-0.10919
	Reaction ID	**Metabolite name**	**Production rate (|mmol/gDW/h|)**	**Difference**
Production	M_Co0080__extr	H+	9.20390	0.25275
	M_Co0059__extr	Sulfate	21.23569	0.04335
	M_e_Biomass__cytop_acaldus	e-Biomass	0.00976	0.00976
	M_C05198__extr	5′-Deoxyadenosine	0.00005	0.00005
	M_Co0237__extr	CO	0.00005	0.00005
	M_Co0160__extr	Glycolate	0.00005	0.00005
	M_Co0416__extr	Phosphatidate	0.00001	-0.00001
	M_C02737__extr	Phosphatidylserine	0.00002	-0.00003
	M_Co0350__extr	Phosphatidylethanolamine	0.00024	-0.00027
	M_Co0993__extr	D-Alanyl-D-alanine	0.00037	-0.00041
	M_Co0344__extr	Phosphatidylglycerol	0.00095	-0.00105
	M_Co0249__extr	Hexadecanoic acid	0.00364	-0.00444
	M_Co0669__extr	gamma-L-Glutamyl-L-cysteine	0.03553	-0.03830

Table 51: Consumption and production rate of the *A. ferrooxidans* model within the community while maximizing both organisms' growth rate.

	Reaction ID	Metabolite name	Consumption rate (|mmol/gDW/h|)	Difference
Consumption	M_C00416__extr	Phosphatidate	0.00001	-0.00001
	M_C02737__extr	Phosphatidylserine	0.00002	-0.00003
	M_C14818__extr	Fe2+	0.00007	-0.00008
	M_C00007__extr	Oxygen	0.00017	-0.00019
	M_C00350__extr	Phosphatidylethanolamine	0.00024	-0.00027
	M_C00993__extr	D-Alanyl-D-alanine	0.00037	-0.00041
	M_C00344__extr	Phosphatidylglycerol	0.00095	-0.00105
	M_C00249__extr	Hexadecanoic acid	0.00364	-0.00444
	M_C00009__extr	Orthophosphate	0.00443	-0.00750
	M_C00014__extr	Ammonia	0.01299	-0.01703
	M_C00669__extr	gamma-L-Glutamyl-L-cysteine	0.03553	-0.03830
	M_C00011__extr	CO2	0.07431	-0.09046
	Reaction ID	**Metabolite name**	**Production rate (|mmol/gDW/h|)**	**Difference**
Production	M_C05198__extr	5′-Deoxyadenosine	0.00007	-0.00008
	M_C00237__extr	CO	0.00007	-0.00008
	M_C00013__extr	Diphosphate	0	-0.00128
	M_e_Biomass__cytop_aferrooxidans	e-Biomass	0.00976	-0.01087
	M_C00283__extr	Hydrogen sulfide	0.03420	-0.03682
	M_C00080__extr	H+	0.04868	-0.05790
	M_C00288__extr	HCO3-	0.09472	-0.10919