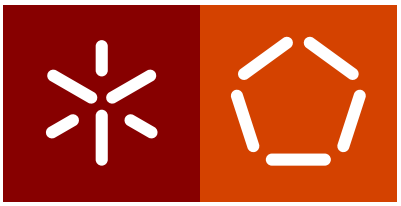
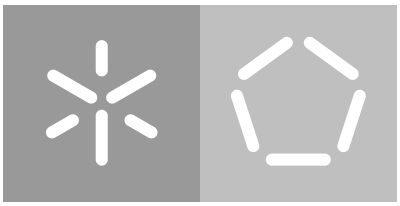


Universidade do Minho
Escola de Engenharia
Departamento de Informática

Emanuel Rodrigues da Cunha

Genome-scale reconstruction of the metabolic model of the probiotic bacterium *Lactobacillus acidophilus*

October 2019



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Emanuel Rodrigues da Cunha

Genome-scale reconstruction of the metabolic model of the probiotic bacterium *Lactobacillus acidophilus*

Master dissertation
Master Degree in Bioinformatics

Dissertation supervised by
Oscar Manuel Lima Dias
Ahmad Adel Zeidan

October 2019

Despacho RT - 31 /2019 - Anexo 3

DIREITOS DE AUTOR E CONDIES DE UTILIZAO DO TRABALHO POR TERCEIROS

ACKNOWLEDGEMENTS

First of all, it was a pleasure to perform this work as my master's degree dissertation. This thesis allowed me to apply different skills acquired through my academic path, as well as develop new ones, contributing also for my personal and professional growth.

I would like to acknowledge my supervisors for all the help. A special thanks to Doctor Oscar Dias for the vast knowledge transmitted and for all the support, availability and proximity at any hour. Another special word to Doctor Ahmad Zeidan for his guidance and support, even considering the distance and the few opportunities to meet.

I would also like to thank Davide Lagoa for the readiness for technical support with *merlin*. Another word to my colleagues in the BioSystems group for the sympathy and kindness demonstrated. A special word to Diogo, José Miguel, Óscar, and Miguel for the companionship, and the talks and jokes in the breaks.

A special thank you to my best friends Andreia and Carlos for their friendship, ridiculous jokes, and remarkable moments. That whole week with you in the summer was awesome! I hope we can repeat something like that in further years.

Finally, a big thank you to my family! Without your support, none of my achievements would be possible. Thank you for your patience when I didn't spend so much time at home as you would like.

This work is a result of the project 22231/01/SAICT/2016: Biodata.pt Infraestrutura Portuguesa de Dados Biológicos, supported by the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

Despacho RT - 31 /2019 - Anexo 4

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.

RESUMO

Nos últimos anos, a população tem vindo a adquirir mais consciência sobre a sua saúde, mudando os seus hábitos alimentares, e procurando alimentos capazes de prevenir problemas associados a doenças. Por isso, a utilização de microorganismos probióticos em alimentos e em suplementos alimentares tem vindo a ganhar a atenção da indústria alimentar. A indústria dos laticínios usa frequentemente organismos probióticos em produtos fermentados, especialmente espécies pertencentes ao grupo das bactérias lácticas.

Lactobacillus acidophilus é uma bactéria láctica probiótica, frequentemente utilizada em alimentos e suplementos alimentares. Algumas estirpes, por exemplo La-14, têm demonstrado efeitos benéficos para a saúde, contribuindo para a vitalidade do trato gastrointestinal, modulando o sistema imunitário e prevenindo problemas renais. Portanto, a biologia de sistemas, nomeadamente modelos metabólicos à escala genómica, irá permitir uma utilização mais eficiente desta espécie na industria alimentar, aprofundando o conhecimento sobre o metabolismo deste organismo e contribuindo também para o desenvolvimento de estirpes desta espécie para fins industriais.

Este trabalho teve como objetivo a obtenção de um modelo metabólico à escala genómica para a bactéria *L. acidophilus* La-14, utilizando as ferramentas *merlin* e COBRApy e a plataforma OptFlux.

A rede obtida em trabalho anterior foi curada manualmente, a composição da biomassa foi determinada com base em informação disponível na literatura, e o modelo foi validado por comparação de simulações com dados experimentais. O modelo metabólico contém 527 genes, 952 reações, e 608 metabolitos, sendo capaz de prever o comportamento da bactéria em diferentes condições ambientais.

A reconstrução do modelo metabólico confirmou as necessidades de *L. acidophilus* por aminoácidos, ácidos gordos e vitaminas. Esta espécie apresenta um metabolismo de carbono simples baseado na glicólise, com ácido láctico como produto final. Este trabalho apresenta ainda uma análise ao metabolismo de *L. acidophilus*, incluindo as vias para o catabolismo de açúcares, incorporação de ácidos gordos, e síntese de nucleótidos e aminoácidos. O potencial para produzir compostos com interesse industrial, como acetaldeído e exopolissacáridos, foi também identificado e analizado.

ABSTRACT

In recent years, people have been acquiring more awareness about their health, changing their nutritional habits, and searching for food able to prevent health disorders. Hence, the addition of probiotic microorganisms to food and their utilization as nutritional supplements has been receiving attention from food industry. Dairy products industry frequently uses probiotic organisms in its fermented products, especially species and strains belonging to the lactic acid bacteria group.

Lactobacillus acidophilus is a probiotic lactic acid bacterium, used in food and dietary supplements for many years. Some strains, such as La-14, had demonstrated health-promoting effects, contributing to the wellness of the gastrointestinal tract, modulating the immune response, and preventing kidney disorders. Therefore, systems biology, specifically genome-scale metabolic models, will allow more efficient utilization of this species by food industry, enhancing the knowledge of the metabolic capabilities of this organism, and contributing to the strain development for industrial purposes.

This work aimed at obtaining a genome-scale metabolic model for the *L. acidophilus* La-14 strain, using *merlin*, together with COBRApy and OptFlux.

The draft network assembled in previous work was manually curated, the biomass composition was determined based on available information, and the model was validated by comparing *in silico* simulations with experimental data. The *GSM* model contains 527 genes, 952 reactions, and 608 unique metabolites, and is able to predict the phenotype of the bacterium under different environmental conditions.

The reconstruction of the metabolic model has confirmed the fastidious requirements of *L. acidophilus* for amino acids, fatty acids, and vitamins. This species presents a simple carbon metabolism based on the EMB pathway, with lactic acid as the final product. This work presents an overview of the metabolism of *L. acidophilus*, including the pathways for carbohydrate catabolism, incorporation of fatty acids, and nucleotide and amino acid biosynthesis. Moreover, the potential to produce relevant compounds for the industry, such as acetaldehyde and exopolysaccharides, was also identified and analyzed.

Keywords: Genome-scale metabolic model, Lactic acid bacteria, *Lactobacillus acidophilus* La-14, *merlin*, *Optflux*, Systems biology

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

A

ADP Adenosine diphosphate.

AI Autoinducer.

ATP Adenosine triphosphate.

B

BIGG Biochemical, Genetic and Genomic.

BLAST Basic Local Alignment Search Tool.

BRENDA Braunschweig Enzyme Database.

C

CDM Chemically Defined Medium.

CHEBI Chemical Entities of Biological Interest.

COBRA Constraint-Based Reconstruction and Analysis.

D

DNA Deoxyribonucleic Acid.

E

EC Enzyme Commission.

EMB Embden–Meyerhof–Parnas.

EPS Exopolysaccharide.

EXPASY Expert Protein Analysis System.

F

FBA Flux Balance Analysis.

FVA Flux Variability Analysis.

G

GAM Growth associated maintenance.

GOLD Genomes Online Database.

GPR Gene-Protein-Reaction.

GSM Genome-Scale Metabolic.

K

KEGG Kyoto Encyclopedia of Genes and Genomes.

L

LTA Lipoteichoic acid.

M

MERLIN Metabolic Models Reconstruction Using Genome-Scale Information.

MFA Metabolic Flux Analysis.

MOMA Minimization of Metabolic Adjustment.

N

NADH Nicotinamide adenine dinucleotide.

NCBI National Center for Biotechnology Information.

NGAM Non-Growth associated maintenance.

NPV Negative Predictive Value.

P

PTS Phosphotransferase system.

R

RNA Ribonucleic Acid.

ROOM Regulatory On/Off Minimization of Metabolic Fluxes.

S

SBML Systems Biology Markup Language.

SIB Swiss Institute of Bioinformatics.

T

TA Teichoic acid.

TC Transporter Classification.

TCDB Transporter Classification Database.

TPP Thiamine diphosphate.

TRANSYT Transport Systems Tracker.

U

UNIPROTKB Universal Protein Resource Knowledgebase.

W

WTA Wall teichoic acid.

1

INTRODUCTION

1.1 CONTEXT AND MOTIVATION

Each year, products fermented by Lactic Acid Bacteria (LAB) generate an economic value exceeding 100 billion €[1]. This market is constantly expanding, as new biotechnological techniques are improving the different properties of dairy products, as well as the way of manufacturing them [2].

Lactobacillus acidophilus is a gram-positive lactic acid bacterium used by the food industry in products like yogurt, sweet acidophilus milk, and cheese. Some *L. acidophilus* strains are known for their health-promoting effects, particularly decreasing upset stomach period, and have been related with the recolonization of the intestinal microbiota. Moreover, members of this species are able to produce natural antimicrobial agents, like bacteriocins, allowing to extend the food conservation period. In addition, they contribute to the flavor, taste, and texture of fermented food products. These properties provide a significant industrial and economic advantage to *L. acidophilus* in the food industry [3].

In 2013 the complete genome sequence of the *L. acidophilus* La-14 strain was published by *Stahl and Barrangou* [4]. Several health-promoting properties have been found using *in vivo* and *in vitro* studies, including oxalate-degradation capability (preventing nephrolithiasis) [5], and beneficial contribution to the immune system [6], in addition to the ordinary cooperation to prevent and treat gastrointestinal disorders. Therefore, improving the knowledge at the metabolic level on this strain, can help optimizing its beneficial properties.

Genome-Scale Metabolic (GSM) models and constraint-based modeling are increasingly important tools in systems biology. These models contain all known metabolites, reactions, and pathways of a target organism, allowing to look at the cell from a global perspective. Hence, the metabolic mechanisms that lead to the final phenotype can be better understood, accelerating the industrial development of biological processes and decreasing not only the costs but also the time required for this kind of exploration [7].

The reconstruction of GSM models involves a large number of steps, including genome annotation, metabolic network assembly and curation, conversion into a stoichio-

metric model, and model validation [8]. The first step for building a GSM for *L. acidophilus*, *i.e.* functional genome annotation, was already performed in previous studies. Hence, this work will continue such process and proceed with the next steps. Due to the complexity and time required for this process, it is highly recommended the utilization of a user-friendly tool designed for this purpose, like *Metabolic Models Reconstruction Using Genome-Scale Information (merlin)* [9, 10]. Besides this tool, it is also necessary to obtain and analyze information available in literature and online databases.

1.2 GOALS

This work aims at obtaining a genome-scale metabolic model of the probiotic bacterium *L. acidophilus*. To achieve this goal, the mid-term goals are as follows:

1. Perform manual curation and refinement of the automatically generated draft metabolic network and functional annotation using information from literature and experimental data on the metabolism and physiology of the organism;
2. Assemble a biomass equation for this organism using available data;
3. Convert the metabolic network into a stoichiometric model;
4. Validate the model using data available in literature/wet-lab experiments.

1.3 DOCUMENT ORGANIZATION

This document is organized as follows:

1. **Chapter** 2: State of the art
   a) A brief contextualization on systems biology;
   b) Overview of the theoretical concepts, tasks, methodologies, and tools for GSM models reconstruction;
   c) Description of lactic acid bacteria, with focus on *L. acidophilus*, including its taxonomic, genomic, morphological and metabolic peculiarities, as well as its industrial and medical applications.
2. **Chapter** 3: Materials and Methods
   a) Description of the methods used for the manual curation of the network;
   b) Formulation of the biomass and energy requirements;
   c) Description of the available tools and developed approaches to identify and correct gaps;

d) Enumeration and explanation of the tools used to validate the model

3. **Chapter** 4: Results and Discussion
    a) Results regarding genome annotation;
    b) Changes in the network through the manual curation;
    c) Biomass and energy requirements formulation;
    d) Comparison of *in silico* simulations with experimental data;
    e) Overview of the main pathways included in the model;
    f) Summary of the metabolic model.
4. **Chapter** 5: Conclusion and Applications
    a) Summary of the developed work;
    b) Possible applications of the model;
    c) Future perspectives.

2

STATE OF THE ART

2.1 SYSTEMS BIOLOGY

In 1915, Walter Cannon described the human body as a control system, making the first association between *system* and *biology* [11]. In 1948, Norbert Weiner used the word *cybernetic* for the first time, introducing the system-level understanding as a relevant theme in biological sciences [12]. Two years later, Ludwig von Bertalanffy stated the general system theory [13], suggesting a new approach that accounts for all components as well as the interactions between them, to predict biological behavior. At that moment, molecular biology was still an emerging field, limiting the development of systems biology [14]. Hence, the methodologies used for strain development were based on random mutagenesis, and screening/selection of a desired phenotype[7]. In mid-1990's, a new opportunity for systems biology emerged when automated genome sequencing and high-throughput measurements become available, allowing more complex analysis of comprehensive datasets. Moreover, the Internet evolution, the development of biological databases and bioinformatics tools started the postgenomic era [7, 14]. From this moment, it became possible to see cells not only as a set of biomolecules but also quantify and analyze the interactions between them, thought mathematical models. These models must be able to simulate cell behavior, as well as predict such behavior in different environmental conditions, and must be subject to validation, comparing simulations with experimental data [7].

Hence, systems biology aims at understanding biological systems at systems-level through four key properties [14]:

1. Systems structure: gene regulation, biochemical networks, and physiological structures;
2. Systems dynamics: how a system changes over time under different conditions;
3. The control method: mechanisms related with the systematic control of the cell;
4. The design method: strategies to modify and construct biological systems having desired properties.

Metabolic models provided by systems biology usually follow two different approaches: stoichiometric modeling and dynamic (kinetic) modeling [15].

Stoichiometric models require less information, representing only the system structure: reactions, compounds, stoichiometry, and reversibility. It uses algebraic methods and constraint based modelling, with the assumption of a (pseudo) steady-state: for all intracellular metabolites, the fluxes leading to production of a given metabolite are balanced with fluxes leading to his consumption, resulting in no net accumulation of metabolites. This makes stoichiometric models and their analysis linear, parameter-free, applicable to metabolic networks of any species with sequenced genome, and easier to simulate [16]. All these advantages allow using stoichiometric models at the genome-scale, originating genome-scale metabolic models [15].

Dynamic models take into account not only structural details, but also concentrations of metabolites over time, reaction fluxes and details of different types of inhibition, and respective parameters. Since it uses differential equations and the number of parameters is usually high, it is hard to perform simulations. So, these models are frequently targeted to a small number of metabolic pathways [15].

Both stoichiometric and dynamic models, together with bioinformatics tools, allow identifying high probability genetic targets, accelerating the industrial development of biological processes [7].

2.2 Genome-Scale Metabolic Models

2.2.1 *Background*

Metabolic network reconstructions existed long before annotated genome sequences were available. Data retrieved from primary literature and biochemical characterization of enzymes were the main sources of information for the reconstruction process for well studied organisms, such as *Clostridium acetobutylicum* [17], *Bacillus subtilis* [18] and *Escherichia coli* [19].

Currently, the availability of whole-genome sequences and biochemical data in biological databases allows reconstructing metabolic networks at genome-scale, even for less characterized organisms. A genome-scale metabolic network can be defined as the set of biochemical reactions inferred from enzymes encoded in the target organism's genome. For each enzymatic reaction in the network, it is necessary to know which substrates and products are involved, its stoichiometry, and the reversibility and location of the reaction [20]. This information allows the description of the set of biochemical reactions and the relation between them. Although *GSM* networks allow inferring some physiological and biochemical properties, only *GSM* models can be used to predict full capabilities of the

metabolic system. As kinetic and regulatory information is only available for a few set of well-studied organisms, these models only add biomass composition and energy requirements to the network [21].

GSM model's reconstruction involves a large number of steps, that have been well described by several authors [21, 8]. This process (summarized in Figure 1) is usually divided into four main stages: genome annotation; metabolic network assembly; conversion of the metabolic network into a stoichiometric model; model validation [7]. However, each one of these stages can include several steps, which can be repeated, as this is an iterative process. In recent years, several tools have been developed specifically to support and accelerate this process. Nevertheless, is still necessary to retrieve information from biological databases and primary literature. Once the process is over, the *GSM* model can be used for different purposes in fields like biotechnology, industry, and biomedicine [22].

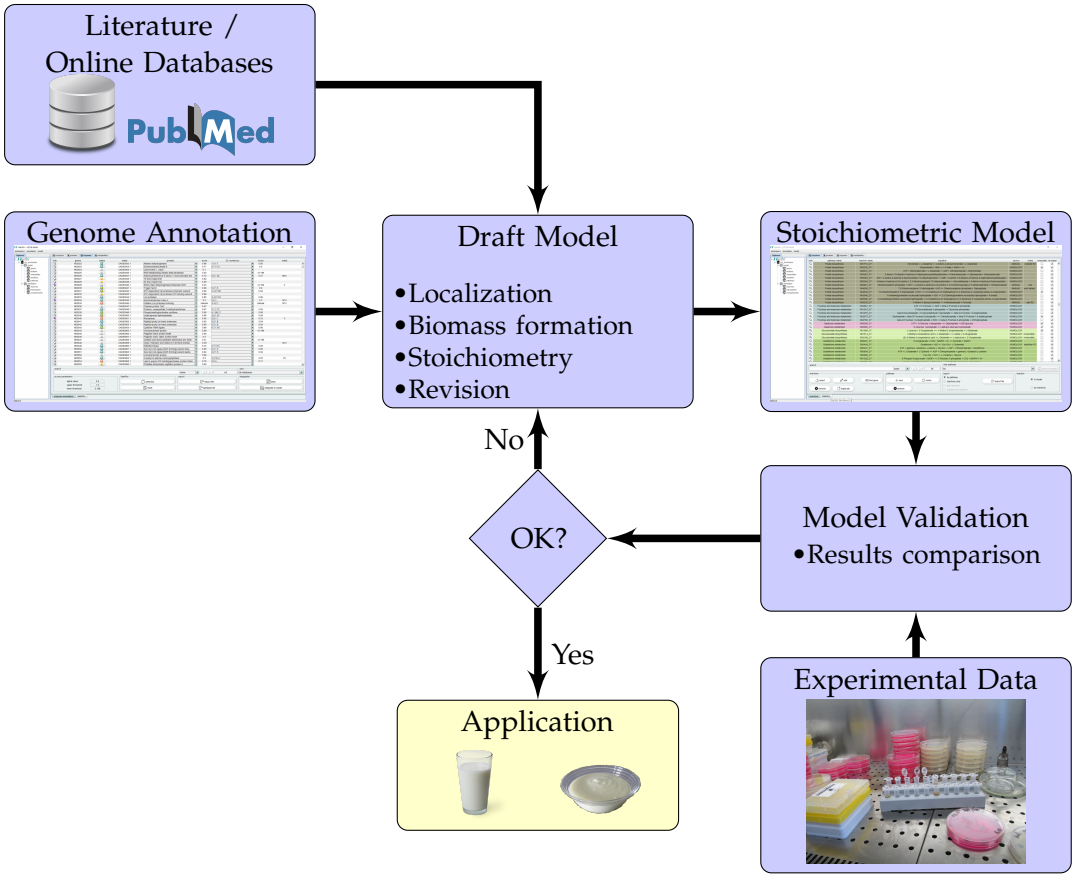


Figure 1.: Schematic representation of the *GSM* model reconstruction process. After obtaining the genome annotation, literature and online databases are used to assemble the metabolic network, who must be refined. Then, the network is converted into a stoichiometric model, that is subject to validation, through comparison of results obtained from simulations and experimental data. If the model is not correct, it must be reviewed. Otherwise, the model is ready to be applied. Adapted from *Russel et al* [7].

2.2.2 *Online resources*

Through all stages, the *GSM* model reconstruction process is supported by information available in different online databases. These resources provide information about genome sequences, proteins, transporters, and metabolic pathways. The most important online resources for *GSM* model's reconstruction are listed in Table 1.

Table 1.: Useful online bioinformatics resources for *GSM* model's reconstruction and respective data output.

Online databases and bioinformatics tools			
Resource name	**Data type**	**Web address**	**Reference**
NCBI[a]	Genomic, metabolic	http://ncbi.nlm.nih.gov/	[23]
GOLD[b]	Genomic	https://gold.jgi.doe.gov/	[24]
KEGG[c]	Genomic, metabolic	http://www.kegg.jp/	[25]
BioCyc	Genomic, metabolic	https:// biocyc.org/	[26]
ExPASy[d]	Genomic, proteomic, metabolic	https:// www.expasy.org/	[27]
UniProtKB[e]	Genomic, metabolic	http:// www.uniprot.org/	[28]
BRENDA[f]	Metabolic	http://www.brenda-enzymes.org/	[29]
BiGG[g]	Metabolic	http://bigg.ucsd.edu/	[30]
TCDB[h]	Metabolic	http://www.tcdb.org/	[31]
TransportDB 2.0	Metabolic	http://www.membranetransport.org/	[32]
ChEBI[i]	Chemical	http://https://www.ebi.ac.uk/chebi/	[33]
PSORTb 3.0	Protein Location	http://www.psort.org/psortb/	[34]
LocTree3	Protein Location	https://rostlab.org/services/loctree3/	[35]
TargetP 1.1	Protein Location	http://www.cbs.dtu.dk/services/TargetP/	[36]

[a]NCBI: National Center for Biotechnology Information, [b]GOLD: Genomes OnLine Database, [c]KEGG: Kyoto Encyclopedia of Genes and Genomes, [d]ExPASy: Expert Protein Analysis System, [e]UniProtKB: Universal Protein Resource Knowledgebase, [f]BRENDA: Braunschweig Enzyme Database, [g]BiGG: Biochemical, Genetic and Genomic Models, [h]TCDB: Transporter Classification Database, [i]ChEBI: Chemical Entities of Biological Interest.

Most resources presented in Table 1 provide more than just one type of data. These resources can be used for *GSM* model's reconstruction, and a brief description of them is performed next.

The *National Center for Biotechnology Information (NCBI)* is one of the most important repositories of biological databases, presenting several resources that provide analysis and visualization for biomedical, genomic, taxonomic and other biological data. The Assembly database [23] is a resource integrated in the NCBI repository that provides assembled genomes, which can be used for *GSM* model's reconstructions.

The *Genomes Online Database (GOLD)* is an online database that provides information related to genome and metagenome sequencing projects, as well as their associated metadata. GOLD is organized into four distinct levels: Study, Biosample/Organism, Sequencing Project and Analysis Project [24].

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collection of databases comprising information of the biological system, including the cell, the organism, and the ecosystem. It provides genomic and metabolic information, including genome sequences, reactions, pathways and chemical compounds relevant to cellular processes [25].

BioCyc is a collection of 13075 Pathway/Genome Databases. Each database provides information and tools to explore the genome and metabolic pathways of a single organism. It presents curated data from thousands of publications, as well as data retrieved from other databases (regulatory networks, gene essentially, protein features, and Gene Ontology annotations). BioCyc also provides computationally predicted metabolic pathways and operons [26].

Expert Protein Analysis System (ExPASy) is an extensible and integrative portal launched by the *Swiss Institute of Bioinformatics (SIB)* [37], that provides databases and software tools regarding different domains of life sciences, including genomics, proteomics, transcriptomics, and phylogenetics [27].

The *Universal Protein Resource Knowledgebase (UniProtKB)* results from the collaboration between the SIB [37], the European Bioinformatics Institute [38], and the Protein Information Resource [39]. According to the revision state of the data provided, this database can be divided into two distinct components: the Swiss-Prot, containing manually annotated records; and the UniProtKB/TrEMBL, whose automatically analyzed records are waiting for full manual annotation [28].

Braunschweig Enzyme Database (BRENDA) is the main repository of enzyme functional data available. It provides functional and molecular information of enzymes, such as kinetics, substrates/products, inhibitors/activators, and cofactors, using the *Enzyme Commission (EC)* system [40]. Each dataset on a classified enzyme is manually checked, which allows using BRENDA as a primary source of enzymatic information [29].

Biochemical, Genetic and Genomic (BiGG) Models is a biochemical, genetic and genomic knowledge database, containing more than 85 high-quality manually-curated *GSM* models. This resource allows users to browse, search and visualize models connected to genome annotations and external databases [30].

The *Transporter Classification Database (TCDB)* incorporates both functional and phylogenetic information of membrane transport proteins to provide the *Transporter Classification (TC)* [41], an extensive approved classification method for transporter proteins (analogous to the EC system for classification of enzymes) [31].

TransportDB 2.0 is a relational database describing automatically predicted cytoplasmic membrane transport proteins, which are classified into protein families according to the TC classification system. Functional and substrate predictions are also provided by this resource for any organism whose genome was already sequenced [32].

Chemical Entities of Biological Interest (ChEBI) is a database of molecular entities focused on "small" chemical compounds. The molecular entities in question are either natural or synthetic products present in biological processes of living organisms. ChEBI contains structure and nomenclature information along with hyperlinks to many well-regarded databases. This database uses a carefully developed ontological classification, whereby the relationships between molecular entities or classes of entities and their parents and/or children are precisely specified [33].

PSORTb 3.0 is a high-precision protein location prediction tool that belongs to the PSORT family. PSORTb 3.0 has remained the most precise bacterial protein subcellular location predictor since it was first made available in 2003, and handles with Gram-positive, Gram-negative and *archaea* sequences [34].

LocTree3 and TargetP are other tools designed for the same purpose of PSORTb 3.0. LocTree3 can be used to assign subcellular protein localization, both for eukaryotes and prokaryotes, through homology and machine learning based inference. [35].

TargetP is a neural network-based tool designed for large-scale subcellular location prediction of newly identified proteins. Using N-terminal sequence information only, it discriminates between proteins destined for the mitochondrion, the chloroplast, the secretory pathway, and "other" subcellular locations, predicting also protein cleavage sites [36].

2.2.3 *Genome annotation*

The first step of any *GSM* model's reconstruction is obtaining the genome annotation, which can be found in organism-specific databases or in databases with collections of genome annotations, like GOLD, KEGG or NCBI. However, if such annotation is not available, it can be performed using specific tools [42]. This step can be divided into two main steps: structural annotation and functional annotation. Structural annotation involves the

search and identification of all features present in the genome, such as protein encoding sequences, promoters, and different types of RNA. This step is usually performed using bioinformatics tools, due to the high cost and time consuming of experimental verification. Genome functional annotation involves assign functions to genes in an unambiguous way through EC and TC numbers. Although only enzyme and transporter encoding genes are included in the model, regulatory and signaling genes can also be integrated in further stages [21].

As mentioned before, enzymes annotation involves the assignment of EC numbers to enzyme encoding genes. EC numbers are a sequence of numbers that allow classifying an enzyme according to the reaction catalyzed. The first number goes for the class of the enzyme; the second goes for the subclass, specifying the type of reaction; the third indicates the sub-subclass, determining the metabolites involved in the reaction; the fourth goes for the serial number of the enzyme in the sub-subclass [40]. Transporters annotation uses TC numbers, in a similar way of EC numbers for enzymes, including also phylogenetic information. TC numbers are composed by five components, corresponding to the transporter class (number), subclass (letter), family (number), subfamily (number), and transported substrate(s) (number) [41].

Different bioinformatics methods have been developed to perform genome annotation, including gene-finding algorithms, such as GLIMMER [43], GlimmerM [44], and GENSCAN [45]. For functional annotation, homology searches are usually performed, using the *Basic Local Alignment Search Tool (BLAST)* [46] or HMMER [47]. Moreover, non-homology-based algorithms like gene neighbor [48], gene cluster [49] and phylogenetic profiles [50] are used to assign function to genes, according to patterns across genomes [51].

EC and TC numbers can be assigned, according to the results provided by the method used for annotation. Bioinformatics tools, such as *merlin* [9, 10] and SuBliMinaL Toolbox [52], allow performing this task semi-automatically. However, this annotation must be manually curated, classifying each assignment with a confidence level, which can be useful to decide the inclusion or exclusion of reactions in the model [21]. The genome functional annotation is a critical step for the *GSM* model reconstruction, as it is usually performed only once and is the basis of all process reconstruction [7].

2.2.4 *Metabolic network assembly*

As soon as the genome annotation is available, the metabolic network can be assembled: EC and TC numbers assigned in the previous stage are converted into enzymatic/transport reactions, which can be performed using databases like KEGG, BRENDA or ExPASy. Spontaneous and non-enzymatic reactions must be included in the model using the same databases. Then, the network must be curated, correcting its stoichiometry, re-

versibility, and subcellular localization of proteins. Moreover, a biomass equation must be assembled and the energetic requirements must be determined. At the end of this stage, the network must be refined, aiming at identifying and correcting gaps [21].

2.2.4.1 *Genes, Proteins, and Reactions*

All annotated metabolic genes must be associated with proteins and reactions through the so-called *Gene-Protein-Reaction (GPR)* associations. These associations provide a mechanistic link between the genotype and phenotype of an organism. The simplest genetic mechanism is one gene encoding one protein that is involved in one reaction: one gene-one protein-one reaction. However, at the genome-scale the presence of enzyme complexes (multiple genes-one protein), isoenzymes (multiple proteins-one reaction), and promiscuous enzymes (one protein-multiple reactions) is very common. These cases must be verified, using the genome annotation and appropriate literature [8]. GPR associations usually implement Boolean rules to link genes, proteins, and reactions in a less complex way. Once again, biological databases can be used to perform these associations, using identifiers for genes, proteins, and reactions [53].

2.2.4.2 *Stoichiometry and Reversibility*

Usually, reactions retrieved from online databases do not present the correct stoichiometry. The commonest errors are missing water molecules and protons, polymerization reactions (monomer + polymer $\rightarrow$ polymer), and missing or wrong metabolites formula. The stoichiometry of these reactions must be corrected, using data available in online databases, like BRENDA, BioCyc or ChEBI. Reactions that cannot be corrected must be removed from the model [21].

Directionality of reactions must be determined at this stage, using the standard Gibb's free energy associated with the reaction, and leading to the definition of appropriated thermodynamic model constraints [54].

2.2.4.3 *Substrate and cofactor usage*

In databases like KEGG, each enzyme is associated with all identified reactions for that enzyme in any organism. Cases whose enzymes are associated with multiple reactions, with different substrates and cofactors, must be manually verified. Organism specific databases and primary literature can be used as a source of information to determine the substrate and cofactor specificity of each enzyme. Reactions with generic compounds, such as "nucleotide" or "primary alcohol", can be refined specifying the respective metabolites, if such information is available. It is important to avoid false inclusions since it can dramatically change the *in silico* behavior [8].

2.2.4.4 *Compartments and localization*

The cellular localization of proteins and reactions is an important step in a *GSM* model reconstruction since it determines the organelles in which the reaction occurs [7]. This step allows to distinguish similar reactions that occur in different compartments and respective metabolites. These reactions and metabolites must be considered different reactions/metabolites, and therefore, they must be replicated in each compartment [21].

In prokaryotic organisms, cellular compartments are usually limited to the cytosol, periplasmic space, and extracellular space. On the other hand, in eukaryotes several organelles can be accounted, including the Golgi apparatus, lysosome, mitochondrion, or chloroplast (in plants) [21]. Information about compartmentalization can often be found in online databases and literature. However, there are bioinformatics tools, like PSORTb 3.0 and TargetP 1.1 (Table 1), which are able to predict the subcellular location of proteins based on amino acid sequence and physiological properties of the organism. If such information cannot be determined, the reaction should be assigned to the cytosol [21].

2.2.4.5 *Biomass formation and energy requirements*

The next step in *GSM* model's reconstruction is to add a set of reactions describing the biomass formation. The biomass reaction includes all known macromolecules as well as their fractional contribution to the overall cellular biomass. These components usually comprises DNA, *Ribonucleic Acid (RNA)*, proteins, lipids, and carbohydrates. However, other cellular constituents like cell wall components must be included in the biomass reactions, depending on the organism's physiology. For each component, a reaction describing its formation form the respective precursors must be included. Thus, the biomass formation can be expressed as:

$$\sum_{k=1}^{p} c_k X_k \rightarrow Biomass,$$

where c_k are the coefficients of the metabolite or macromolecule X_k. The flux of this reaction represents the growth rate of the organism. The biomass composition of the target organism must be experimentally determined for cells in logarithmic phase [55]. If such assays cannot be performed, bioinformatics methods and primary literature can be used. According to some authors [56], the biomass equation of a closely related organism can be used, without introducing significant errors in simulations. DNA, RNA and protein contents can be estimated from the genome of the target organism, using the codon usage bias [57], which provides better results than using the biomass equation from a related species [58].

Macromolecules whose content cannot be inferred from the genome, like lipids and carbohydrates, require experimental data. The lipid composition is more difficult to deter-

mine, as each lipid has fatty acids whose chains can differ in length and saturation degree. In this case, the model will not present all possible combinations, but only a compound representing the average composition that is consistent with experimental data [8]. Although cofactors represent a small portion of the biomass composition, they are crucial for cell growth. Hence, they must be included in the biomass equation [59].

Besides the already mentioned macromolecules, the biomass formulation must include *Growth associated maintenance (GAM)* requirements in terms of *Adenosine triphosphate (ATP)* molecules needed per gram of biomass synthesized. These requirements are associated with the energy required to replicate the cell (*e.g.* DNA, RNA, and protein synthesis). *Non-Growth associated maintenance (NGAM)* requirements should also be included in *GSM* models. These requirements account for the energy necessary to maintain cell conditions, like turgor pressure and membrane potential, and are included in the model in a simple hydrolysis of ATP into *Adenosine diphosphate (ADP)* and inorganic phosphate [60]. Both GAM and NGAM can be determined in chemostat experiments, plotting the ATP production rate *vs* biomass growth rate, in which the y-intercept and slope represent the NGAM and GAM, respectively. Alternately, these requirements can be estimated by determining the energy required for macromolecular synthesis or obtained from an organism related data [8, 61]. For aerobic organisms, the amount of ATP formed per oxygen atom (P/O ratio) must be determined, to obtain the ATP stoichiometry of oxidative phosphorylation [54]. Before proceed with the network refinement, growth medium requirements must be collected, and a detailed growth medium must be defined, accounting a carbon, phosphorus, nitrogen and sulfur sources, auxotrophies, and other environmental requirements of the target organism.

2.2.4.6 *Network refinement*

Despite all the benefits of automatic methods for *GSM* model reconstruction, a manual refinement of the metabolic network must be performed. At this stage, missing and unconnected reactions must be identified, as well as inconsistencies of proteins and function identifiers, and incomplete EC numbers. Reactions known to be present in the target organism, but are absent in the network must also be included. The charge of each metabolite must be determined, according to the pH of the organelle [7].

2.2.5 *Conversion into a stoichiometric model*

The third stage of the reconstruction process is to convert the metabolic network into a constraint-based mathematical format, through a stoichiometric matrix. Here, pseudo-steady state is assumed, with the metabolites concentrations remaining constant, and the

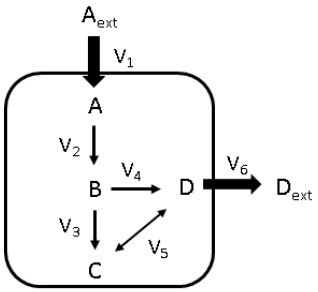
consumption rates equal production rates for each metabolite, as shown in the following equation:

$$S \cdot v = 0 \tag{1}$$

where S is the stoichiometric matrix (columns represent reactions and rows represent metabolites), and v is the flux vector. Usually, the number of fluxes distributions is much greater than the number of mass balance constraints. Hence, metabolic networks are undetermined systems, with an infinite number of solutions, the so-called *null space* of S. An infinite number of solutions blocks the computation of a single solution, restraining the information that can be retrieved from the model [7]. To reduce the null space of S, constraints must be added to the model. There are four different types of constraints that must be considered: physicochemical, topological, regulatory and environmental conditions constraints. *GSM* models usually rely on physiochemical and environmental constraints (flux and energy balances, enzyme and transporter capacity, and thermodynamics) [61]. These constraints can be established by setting boundaries to reactions and transport fluxes. Reactions boundaries determine their reversibility: a reversible reaction is constrained between minus infinity and plus infinity, while the minimum or maximum of irreversible reactions must be fixed to zero. Transport fluxes follow a similar approach: the maximum uptake of nutrients must be limited to a specific uptake rate. Other usual constraints are related with the oxygen availability, metabolites unavailable in the medium and excreted metabolites, depending on the target organism and growth medium [7]. Constraints can be introduced in the network as inequalities, as shown in equation 2.

$$\alpha_j \leq \mathrm{v}_j \leq \beta_j, \qquad j = 1, ..., N \tag{2}$$

As mentioned before, these constraints reduce the number of solutions, providing a set of feasible flux distributions. However, additional constraints related with gene regulation and kinetic limitations can also be considered. Figure 2 presents a simple example of a metabolic network with hypothetical compounds and fluxes. Here, the key aspects of converting a *GSM* network into a stoichiometric model are summarized.



(1)

$$
\begin{aligned}
A &: -V_2 + V_1 = 0 \\
B &: -V_3 - V_4 + V_2 = 0 \\
C &: -V_5 + V_3 = 0 \\
D &: -V_6 + V_4 + V_5 = 0
\end{aligned}
$$

(2)

$$
\begin{aligned}
&0 \leq V_1 \leq a \\
&0 \leq V_2 \leq \infty \\
&0 \leq V_3 \leq \infty \\
&0 \leq V_4 \leq \infty \\
&-\infty \leq V_5 \leq \infty \\
&0 \leq V_6 \leq \infty
\end{aligned}
$$

(3)

$$
\begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 & -1 & 0 \\
0 & 0 & 0 & 1 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6
\end{bmatrix}
=
\begin{bmatrix}
0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix}
$$

(4)

Figure 2.: Illustration of a metabolic network with four metabolites (A to D) and six fluxes (V_1 to V_6). The stoichiometry of all metabolites is defined as 1 and -1. Section (1) represents the reaction scheme. V_1 and V_6 are the exchange fluxes of the metabolites A and D, respectively. Double arrows are for reversible reactions and irreversible reactions are indicated with forward arrow. Section (2) shows the steady-state mass balances, and section (3) represents the boundaries of flux values (a is the specific uptake rate for the metabolite A). Section (4) shows the mass balance in matrix format (S matrix and flux vector). Adapted from *Rocha I., Frster J. and Nielsen* [21].

At the end of this stage, the stoichiometric model is saved in a *Systems Biology Markup Language (SBML)* file [62], a simple format, readable by the most of specialized tools designed for simulation, such as OptFlux [63] or *Constraint-Based Reconstruction and Analysis (COBRA)* [64].

2.2.6 *Validation of the Metabolic Model*

The final stage of the *GSM* model reconstruction consists on the verification, evaluation, and validation of the model. The metabolic model created must be tested for its ability to produce biomass precursors. Usually, this analysis leads to the identification of missing metabolic functions, which are added by repeating part of the second and third stages. It is expected that the network presents a significant number of gaps. Therefore, dead-end metabolites and missing reactions and functions must be identified first. Then, a deep literature search is required, using biochemical textbooks, KEGG maps, or other biochemical maps, to identify the metabolic context of the dead-end metabolite. This allows identifying the enzyme(s) involved in the production of the dead-end metabolite, providing a starting point for literature search, and if necessary, the revision of the genome annotation [7]. Fill-

ing a gap requires caution, as it may generate new gaps or enable the model to carry out a function that the target organism is not able to do. However, in some cases, gap-filling is needed to ensure the model's feasibility, like guarantying the biomass precursors biosynthesis. As a rule, if a gap reaction does not present information that sustains its existence, it should not be added to the model, unless it is mandatory for the model's usefulness [7].

Different methodologies can be followed to perform model simulation and flux distributions prediction. If measurements of external exchange fluxes are available, total flux distributions can be calculated using *Metabolic Flux Analysis (MFA)* [65]. However, MFA is fallible in two situations:

1. when the number of measured fluxes is not enough;
2. when a single solution does not exist due to the presence of different pathways with the same overall stoichiometry (parallel and cyclic pathways) [54].

Nevertheless, several algorithms have been developed to explore *GSM* models in metabolic engineering. These algorithms can be divided into: linear programming; quadratic programming; mixed integer linear programming; and evolutionary programming [61].

Flux Balance Analysis (FBA) [56, 66] is the most frequently used approach for analysis of biochemical networks [67]. This method is based on linear programming to calculate optimal flux distribution, using a linear objective function and constraints. Therefore, an FBA problem can be formulated as:

$$\begin{aligned} &\text{Maximize} && Z \\ &\text{subject to} && S \cdot v = 0 \\ &&& \alpha_j \leq \mathrm{v}_j \leq \beta_j, \qquad j = 1, ..., N \end{aligned} \tag{3}$$

In metabolic engineering, the objective function Z can be related to different objectives: maximize the cell growth rate (biomass formation); maximize/minimize the ATP or *Nicotinamide adenine dinucleotide (NADH)* production; and maximize the production of target metabolites [61]. FBA provides a single solution for flux distribution, which is not necessarily unique since other optimal flux distributions can exist. Figure 3 illustrates the conceptual basis of constraint-based modeling and FBA, showing the solution space through the process.

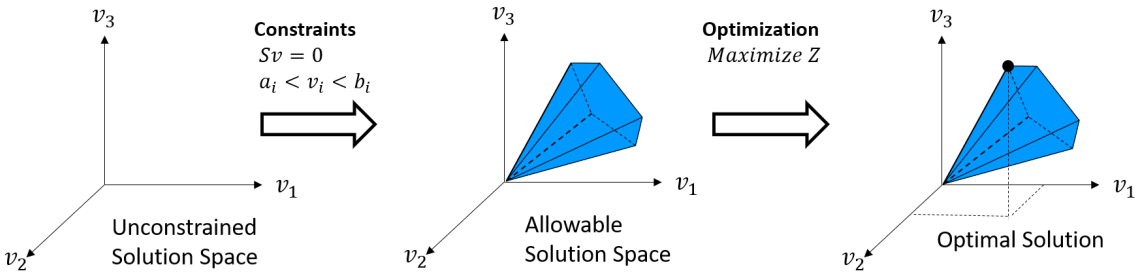


Figure 3.: Illustration of the conceptual basis of FBA. Without constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints (1) and capacity constraints (2) are applied to the network, an allowable solution space is defined. The network may acquire any flux distribution within this space. Through optimization of an objective function, FBA can identify a single optimal flux distribution that lies on the edge of the allowable solution space. Adapted from *Orth et al* [67]

However, FBA has its limitations. Once it does not use kinetic parameters, it can not predict metabolites concentrations, being only suitable for determining fluxes at steady-state. FBA does not account for enzyme regulation by protein kinases or regulation of gene expression, which can mislead predictions [67].

Besides FBA there are available several other algorithms and approaches for *GSM* model's analysis. *Flux Variability Analysis (FVA)* [68] is an approach for evaluating the possible range of each reaction flux satisfying the constraints. The *Minimization of Metabolic Adjustment (MOMA)* [69] is based on the assumption that the metabolism of knockout strains is close to metabolism operating in wild-type strains. MOMA employs quadratic programming without the assumption of optimal growth, and has been displaying better results than FBA [69, 61]. *Regulatory On/Off Minimization of Metabolic Fluxes (ROOM)* [70] is an algorithm similar to MOMA, but it uses mixed integer linear programming. OptGene [71] is an evolutionary algorithm developed to find deletion targets in microorganisms, using Simulated Annealing [72, 73] and Set-based Evolutionary Algorithms [72].

The analysis of strategic fluxes, like specific growth rate or by-product formation, for different growth conditions is another usual approach for model's evaluation. These fluxes can be determined by implementing environmental conditions reported in literature as model constraints. The comparison of the generated results can be used for model validation or to calibrate the model, adjusting, for example, the GAM and NGAM parameters. Finally, simulations with deletion mutants must be performed, providing perceptions of the prediction capability of the model. Regardless of the approach used for validation, if the model does not satisfy the expectations, it must be verified and corrected, usually from the second stage [7].

2.2.7 *Computational tools*

The large number of steps required for reconstructing *GSM* models, lead to the development of bioinformatics tools that perform these steps (semi-)automatically. The increasing number of available tools provides several features that make the reconstruction process faster and simpler (Table 2).

Table 2.: List of relevant available metabolic modeling tools used for *GSM* model's reconstruction, and respective features. Adapted from *Dias et al* [9]

Automatic metabolic modeling tools							
Software	*MEMOsys 2.0*	*CoReCo*	*Pathway Tools*	*Model SEED*	*SuBliMinaL Toolbox*	*RAVEN*	*merlin*
Enzymes annotation		✓		✓	✓	✓	✓
Transporters annotation			✓	✓	✓		✓
Compartments prediction	*a*			✓	✓	✓	✓
Biomass equation		*b*		✓	✓	*b*	*b*
Export to SBML	✓	✓		✓	✓	✓	✓
Runs locally	✓	✓	✓		✓	✓	✓
Graphical interface for manual curation			✓				✓
Pathways visualization			✓	✓		✓	✓
GPR rules				✓	✓		✓
Highlight metabolic dead-ends			✓			✓	✓
Reactions stoichiometry validation	*c*	✓	*c*	*c*	✓		✓
Prokaryotic models	✓	✓	✓	✓	✓	✓	✓
Eukaryotic models		✓	✓		✓	✓	✓

a Allow to manual assign compartments to reactions

b Biomass inserted manually

c Model SEED and Pathway tools use their own metabolite database

Tools like *MEMOsys 2.0* [74] and *CoReCo* [75] are specialized in particular stages of the reconstruction procedure, offering a few features to assist this process. *Pathway tools* presents more features, through not allowing to export the model in the SBML format. *Model SEED* [76] can only be used for prokaryotics. Other tools, such as the SuBliMinal Toolbox, provide a great number of features but do not offer a graphical interface, which can be an obstacle for less advanced users. *RAVEN* [77] requires a commercial license for its utilization. *merlin* offers a panoply of features for all stages of the *GSM* model reconstruction for both eukaryotic and prokaryotic beings. This open-source tool presents a graphical interface, allowing also export the model in the SMBL format.

2.2.7.1 *merlin*

The reconstruction process using *merlin* starts by downloading the GenBank file containing the genome of the target organism, which can be performed automatically, inserting the taxonomic ID. Then, two modules are provided. The first one allows performing the genome annotation (enzymes and transporters annotation and compartmentalization). The second module allows the model's assembly and curation. Usually, the genome annotation starts with the enzymes annotation. *merlin* allows to use both BLAST (against several remote databases) and HMMER for similarity searches. For every possible annotation of each gene, a score comprising two factors is calculated, according to equation 4.

$$Score = \alpha \cdot Score_{frequency} + (1 - \alpha) \cdot Score_{taxonomy} \tag{4}$$

The score can range between 0 and 1, and is used as a confidence level of the function automatic assignment for each gene. The $Score_{frequency}$ is associated with the frequency each EC number is found within the homologous gene record annotation. The $Score_{taxonomy}$ is related with the taxonomy of the organisms to which these records belong. After the manual annotation of a small sample of genes, *merlin* automatically determines the α value, as well as an upper threshold and a lower threshold using the embedded tool SamPler. EC number assignments above and below these thresholds are automatically accepted and rejected, respectively. The remaining ones must be manually annotated. The *Transport Systems Tracker (TranSyT)* [78] is a tool implemented in *merlin* able to identify transport systems, and generate the corresponding transport reactions. The information present in TranSyT database is automatically extracted from TCDB, and annotated using BioSynth with TC numbers. The assignment of compartments to reactions can be achieved using protein localization prediction tools, such as LocTree or Psortb. The obtained file can be loaded into *merlin*, and the compartmentalization is then integrated into the model, assigning a compartment to each reaction.

The second module allows performing the next steps in *GSM* model reconstruction. The genome annotation can be integrated and combined with KEGG metabolic data, to assembly a draft network. *merlin* allows determining the reversibility of reactions automatically, using the MODEL SEED or Zeng[79] databases. Unbalanced and unconnected reactions can be highlighted, and must be corrected manually. A reaction describing the biomass formation (named "e-biomass") can be added to the model, together with reactions discriminating the assembly of DNA, RNA, and proteins. These reactions ("e-DNA","e-RNA", and "e-Protein") describe the components of each one of these building blocks, as well as the respective coefficients, determined by the codon usage bias. Lipids, carbohydrates, and other macromolecules must be added manually, as well as their coefficients and mass fraction. Another important feature provided by *merlin* is the determination of

the GPR rules, using KEGG. This task is automatic, but manual curation can be done yet. Drains (exchange reactions) can also be added for external metabolites, setting a lower boundary to 0 and an upper boundary to 999999. Then, drains of metabolites that compose the growth media must be identified and changed if necessary.

merlin does not allow performing simulations and validation. However, the model can be exported in the SBML format with the desired information (metabolites formulae, biomass reaction name, level, and version of SMBL file). The SBML file can be validated online if necessary. This file can be used for model validation, using appropriate tools for this purpose [9].

2.2.7.2 *Simulation and Validation tools*

The SBML file containing the *GSM* model can be imported into several tools like OptFlux and COBRA to perform simulations and validate the model.

Over the past decade, COBRA methods have been largely used to simulate, analyze and predict a variety of metabolic phenotypes using *GSM* models. COBRA toolbox [80], a MATLAB package for implementing COBRA methods, is currently running with version 3.0. This version presents some useful features, such as: network gap filling, ^{13}C analysis, and omics-guided analysis and visualization. COBRA toolbox 3.0 allows performing common approaches for *GSM* model simulation, such as FBA, FVA, MOMA, OptGene, and OptKnock. It allows visualization and map manipulation of metabolic models, using Cell Designer [81, 82]. Another package implementing the COBRA methods is COBRApy [83]. This Python package is designed in an object-oriented way, simplifying the representation of the complex biological processes, such as metabolism and gene expression. COBRApy allows to perform FBA, FVA, and strain design (accessible through the COBRA Toolbox).

OptFlux in an open-source and modular software, with a graphical user interface, designed for *in silico* metabolic engineering tasks. It was the first tool to use Evolutionary Algorithms/Simulated Annealing metaheuristics or the OptKnock [84] algorithm for strain optimization, and presents several methods for model simplification, pre-processing, and visualization (compatible with the layout information of Cell Designer). It also allows the utilization of stoichiometric metabolic models for:

1. phenotype simulation of both wild-type and mutant organisms, using the methods of FBA, MOMA or ROOM;
2. Metabolic Flux Analysis, using a given set of measured fluxes to compute the admissible flux space;
3. pathway analysis through the calculation of Elementary Flux Modes;

This software is compatible with the SBML format, allowing also importing and exporting models to several flat file formats. [63].

2.2.8 *Current State and Applications of GSM models*

For more than 15 years, *GSM* models have been applied in different fields, such as food biotechnology, biomedicine, biofuels industry, and biological discovery. The utilization of these models is based on their comprehensive description of genotype-phenotype associations and simple mathematical formulation [22], which allows using *GSM* models in different ways: *in silico* simulation of the phenotypic behaviour of the microorganisms under different environmental and genetic conditions; metabolic engineering design; assignment of functions to unknown genes; and identification of drug targets [85, 86, 87, 88].

Until 2016, almost 200 *GSM* models for organims belonging to different domains (*Bacteria*, *Eukarya*, and *Archea*) were released. However, the number of *GSM* models released per year has been decreasing since 2014, which can be attributed to two reasons [87]:

1. the most important, scientifically interesting, and better-studied microorganisms have already been reconstructed;

2. the reconstruction process has become a routine job, so the work is no longer published.

Food biotechnology has been applying different mathematical approaches, especially for design and optimization of interesting strains [89, 90, 91]. However, the full potential of constraint-based modeling applied together with *GSM* models remains largely unexplored in this field [92]. These methods provide opportunities for strain and culture development, and optimization of bioprocesses for production of microbial food cultures. Lactic acid bacteria has been an important group of organisms for this field.

Genome-scale metabolic models are already available for several LAB, including *Lactococcus lactis* [93, 94], *Lactobacillus plantarum* [89], *Streptococcus thermophilus* [95], *Lactobacillus casei* [96, 97], *Oenococcus oeni* [98], and *Leuconostoc mesenteroides* [99]. These models have been providing relevant information for food industry, including knowledge about carbohydrate utilization, amino acid auxotrophies, nutritional requirements, and metabolic secretions [92].

GSM models can also be used for *in silico* strain design. The food industry has been searching for strains with improved functional, technological, and safety properties [92], as well as the ability to produce functional food ingredients (low-calorie sweeteners, vitamins, amino acids and bioactive peptides [100, 101, 102]). Once *GSM* models present a view of the target organism as a whole, the process of identifying targets for gene deletion, for example, is easier and sometimes allows identifying unexpected targets [103]. An overview of the contributions of *GSM* models of LAB is presented in Table 3.

Table 3.: Main contributions of available *GSM* models for lactic acid bacteria.

***GSM* models of Lactic Acid Bacteria**		
Specie	**Contribution**	**Reference**
*L. lactis IL*1403	•Basis for other *GSM* model reconstructions •Design of enhanced metabolic engineering strategies	[94]
L. lactis MG1363	•Identification of reactions directly or indirectly involved in flavor formation	[93]
L. plantarum WCFS1	•Comparison between the traditional view of ATP production from lactate and acetate, and based on experimental constraints. •Identification of catabolic reactions for ATP production	[89]
S. thermophilus LMG18311	•Identification of a unique pathway for acetaldehyde production and the absence of a complete pentose phosphate pathway	[95]
L. casei 12A, ATCC334	•Understanding of glutamate requirements for this species •Confirm the hypothesis of gene decay during adaption to nutrient rich environments	[96]
L. casei LC2W	•Analysis of the oxygen effect on flavor compound biosynthesis.	[97]
O. oeni PSU-1	•Relation between consumption rates of fructose, amino acids, oxygen, and malic acid and production rates of erythritol, lactate, and acetate	[98]
L. mesenteroides iLME620	•Reported new hypothesis on the malolactic fermentation mechanism	[99]

Usually, food cultures are based in microbial communities, instead of a single strain. Therefore, interactions between strains must be evaluated, extending the constraint-based modeling approach used for single organisms for microbial communities [92]. Different frameworks, like mixed-bag modeling, species compartmentalization or multi-species dynamic modeling [104, 105, 106], can be used to build and analyze community metabolic models [92].

Although *GSM* models present a huge potential in different fields, there are some limitations for its reconstruction and application. First, for less studied organisms, lack of knowledge and experimental data presents an obstacle in the reconstruction process [87]. Second, this process is still slow and requires a substantial amount of knowledge and effort, despite the recent tools that automated part of the process. Finally, the negative

perception of the general public on genetically modified organisms, together with the tight regulamentation imposed by many European countries [107] are a major barrier, especially for food biotechnology [92].

2.3 LACTIC ACID BACTERIA

2.3.1 *Background*

The term "lactic acid bacteria" was used for the first time at the beginning of the 20th century to describe "milk souring organisms" [108]. The criteria used back then (cellular morphology, mode of glucose fermentation, temperature ranges of growth, and sugar utilization patterns) served as the basis for the current classification of LAB. Modern taxonomic tools, especially molecular biology methods, allowed also tracking the origin of this group, as well as refine its taxonomy [109].

Apparently, LAB represent an adaptation of ancient *Bacillus*-like soil and plant organisms to the gut of herbivorous animals. The adaptation process induced the loss of many metabolic activities, that are compensated by efficient fermentation systems, energy recycling, transport mechanisms, adhesion to intestinal cells, and acid production and tolerance. These adaptive advantages provide to LAB the means to compete with other microorganisms in their environment [109, 110].

LAB belong to the phylum *Firmicutes*, class *Bacilli*, and order *Lactobacillales*. Currently, LAB comprises seven genera: *Lactococcus*, *Lactobacillus*, *Enterococcus*, *Pediococcus*, *Streptococcus*, *Leuconostoc* and *Oenococcus* [111, 112]. A phylogenetic tree of LAB and some related species is shown in Figure 4.

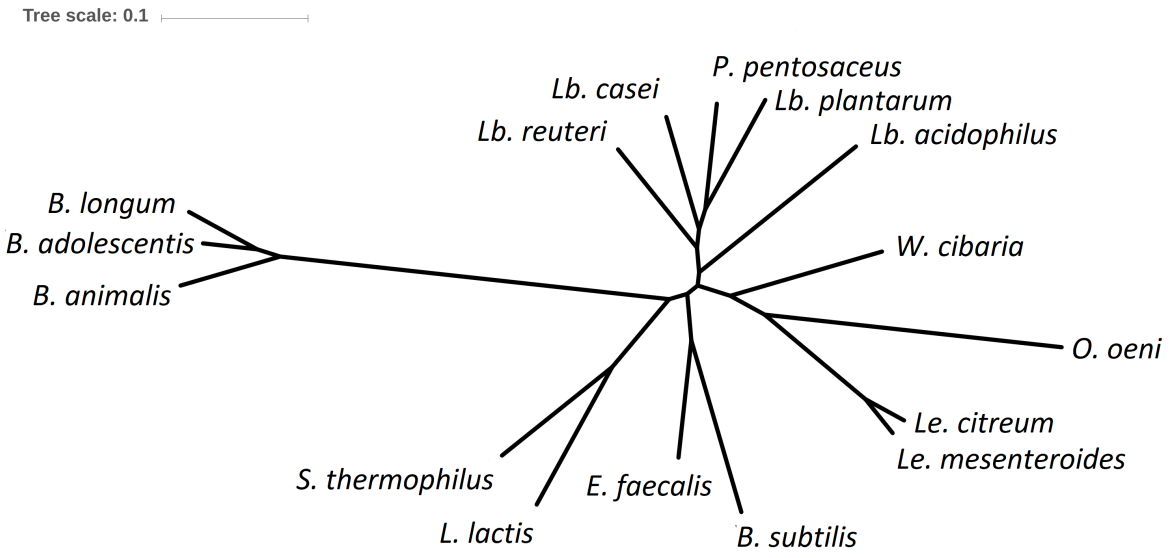


Figure 4.: Phylogenetic tree of different species belonging to the LAB group. The tree was created by applying a neighbor-joining method to an alignment of 16S rRNA gene sequences obtained from NCBI, using the following GenBank accession numbers: M58802 (*Lactobacillus acidophilus*), NR_115534 (*Lactobacillus casei*), NR_119069 (*Lactobacillus reuteri*), M58827 (*Lactobacillus plantarum*), M58834 (*Pediococcus pentosaceus*), NR_036924 (*Weissella cibaria*), NR_040810 (*Oenococcus oeni*), NR_041727 (*Leuconostoc citreum*), M23035 (*Leuconostoc mesenteroides*), NR_112116 (*Bacillus subtilis*), NR_040789 (*Enterococcus faecalis*), NR_040954 (*Lactococcus lactis*), NR_042778 (*Streptococcus thermophilus*), NR_145535 (*Bifidobacterium longum*), NR_040867 (*Bifidobacterium animalis*), M58729 (*Bifidobacterium adolescentis*).

From the physiological and metabolic perspective, LAB are gram-positive, nonsporulating, and nonrespiring (but aerotolerant) rod or cocci bacteria, which produce lactic acid as the major final product of fermentation processes. There are two basic fermentative pathways for sugars in LAB: the homofermentative pathway (glycolysis or *Embden–Meyerhof–Parnas (EMB)* pathway), which produces only lactic acid; and the heterofermentative pathway (6-phosphogluconate pathway), which produces CO_2, ethanol, and acetic acid, besides lactic acid. Pentoses can only be fermented by heterofementative LAB, entering in the respective pathway as ribulose-5-phosphate or xylulose-5-phosphate [109].

LAB are generally considered beneficial and probiotic microorganisms, contributing also to the biopreservation of fermented products, due to the production of bacteriocins. Additionally, acetaldehyde and diacetyl produced by LAB provide an appealing flavor to these products [110]. Therefore, some species are used by food industry in fermented foods and beverages, such as cheese, yogurt, and other fermented milk; sourdough and

other breads; alcoholic beverages, such as wine, cider and beer [113]. However, some genera (*Streptococcus*, *Lactococcus*, *Enterococcus*, *Carnobacterium*) present species or strains recognized as pathogens [109].

2.3.2 *L. acidophilus*

2.3.2.1 *Taxonomy*

L. acidophilus belongs to the *Lactobacilli* genus, the largest and the most diverse genus of LAB, comprising more than 200 species and subspecies [114]. This species belongs to a heterogeneous group of *Lactobacilli* called *L. acidophilus* complex, comprising six different species: *L. acidophilus*, *Lactobacillus crispatus*, *Lactobacillus johnsonii*, *Lactobacillus gallinarum*, *Lactobacillus amylovorus*, and *Lactobacillus gasseri*. The recognition of this group was performed in 1980 by DNA hybridization studies [115, 116]. The phylogenetic relationship of the *L. acidophilus* complex, as well as other relevant close related species, are presented in Figure 5.

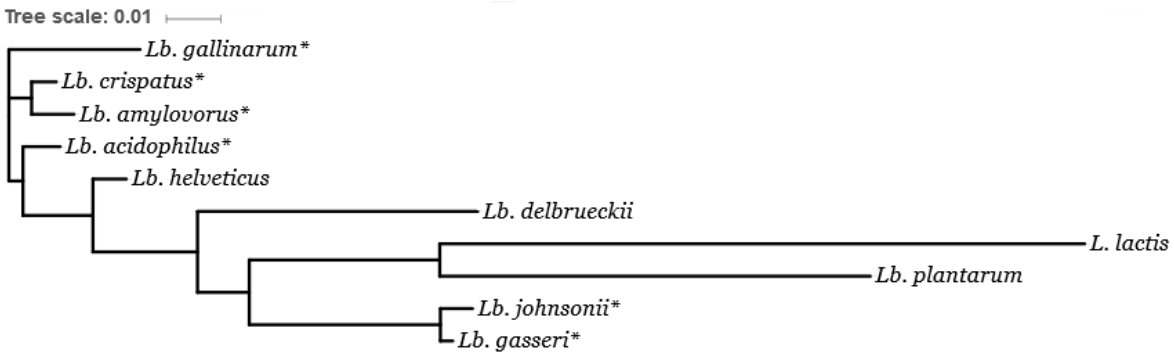


Figure 5.: Phylogenetic tree showing the relationship between members of the *L. acidophilus complex* (marked with "*"), other *Lactobacilli*, and *Lactococcus lactis*. The tree was created by applying a neighbor-joining method to an alignment of 16S rRNA gene sequences obtained from NCBI, using the following GenBank accession numbers: AM088020 (*L. lactis*),M58805(*Lb. amylovorus*),Y17362 (*Lb. crispatus*),X61138 (*Lb. acidophilus*),X61141 (*Lb. helveticus*), AJ242968 (*Lb. gallinarum*),AJ002515 (*Lb. johnsonii*),X52654 (*Lb. delbrueckii*), M58820 (*Lb. gasseri*), X52653 (*Lactobacillus plantarum*).

L. acidophilus is most closely related to *Lactobacillus helveticus* (a milk-fermenting *Lactobacillus*), *L. crispatus*, and *L. amylovorus*. Surprisingly, *L. acidophilus* is more phylogenetically distant of some species of the *L. acidophilus* complex, like *L. gasseri*, than other milk-fermenting *Lactobacilli*, such as *L. helveticus*. The genetic relationship between members of the *L. acidophilus* complex, and other *Lactobacilli* (especially *L. helveticus* and *L. delbrueckii*), has been elucidated by genome sequencing of these species, and consequent comparative genomics. These methods allowed to find relevant disparities between species belonging to the *L. acidophilus* complex that were not found previously [117].

2.3.2.2 *Genomic features*

Currently, 16 *L. acidophilus* strains have been sequenced, with 6 genomes available [4, 118, 119, 120, 121]. The genome of the La-14 strain was published in 2013 by *Stahl and Barrangou* [4], presenting a 1,991,830-bp length, in agreement with other *L. acidophilus* strains. Genomes of this species have a relatively small size (1.25-2.05 Mbp) [111], which represents an outcome of the evolutionary adaption to nutrient-rich niches, like the mammalian intestinal tract. These strains lack amino acid biosynthetic pathways, which is compensated with the expression of several proteases and oligopeptide transporters. Moreover, genomes of *L. acidophilus* strains encode multiple saccharide transporters, allowing the catabolism of different carbohydrates present in the gastrointestinal tract. In total, about 13-18% of these genomes encode amino acid and sugar transport proteins [117].

Despite biochemical and physiological analysis show some diversity between *L. acidophilus* strains, genotypic analysis indicates less disparity within the genomes of this species [4, 111]. The alignment of genomes of the La-14 and NCFM strains showed high similarity (98%) [4]. The relative amount of GC is 34.7% for both strains, and a deletion in an ABC transporter ATP binding protein was found in the La-14 strain. In total, 16 single-base-pair indels, and 95 single-nucleotide-polymorphisms were found, affecting 52 genes [4].

2.3.2.3 *Morphology and Growth conditions*

Like other LAB, *L. acidophilus* is a gram-positive microorganism. This species is a homofermentative anaerobic bacterium, presenting a rod shape that ranges in size from 2 to 10 μm. It grows at a temperature of 30-45 °C, and a pH of 4-5 [3], requiring not only carbohydrates, but also nucleotides, amino acids, vitamins, calcium pantothenate, folic acid, niacin, and riboflavin. Pyridoxal, thiamine, thymidine, and vitamin B12 are not essential [122].

Alanine, lysine, asparagine, glutamine, and glycine are the only non-essential amino acids in *L. acidophilus*. However, this species is able to synthesize some amino acids if specific precursors are available on the growth medium: Leucine (α-ketoisovalerate or α-ketoisocaproate); Serine (3-phosphoserine); tyrosine (hydroxyphenylpyruvate); cysteine (sulfide); and phenylalanine (chorismate or phenylpyruvate) [123].

L. acidophilus presents resistance to several antibiotics: kanamycin, chloramphenicol, erythromycin, tetracycline, penicillin, and vanomycin [124, 125, 126]. However, some drugs were recognized as inhibitors of *L. acidophilus* growth, such as amoxicillin, amiodarone, and paracetamol [127].

The cell wall of *L. acidophilus* is composed by four major components: peptidoglycan, *Teichoic acid (TA)* (wall teichoic acids and lipoteichoic acids), S-layer, and polysaccharides (Figure 6) [128].

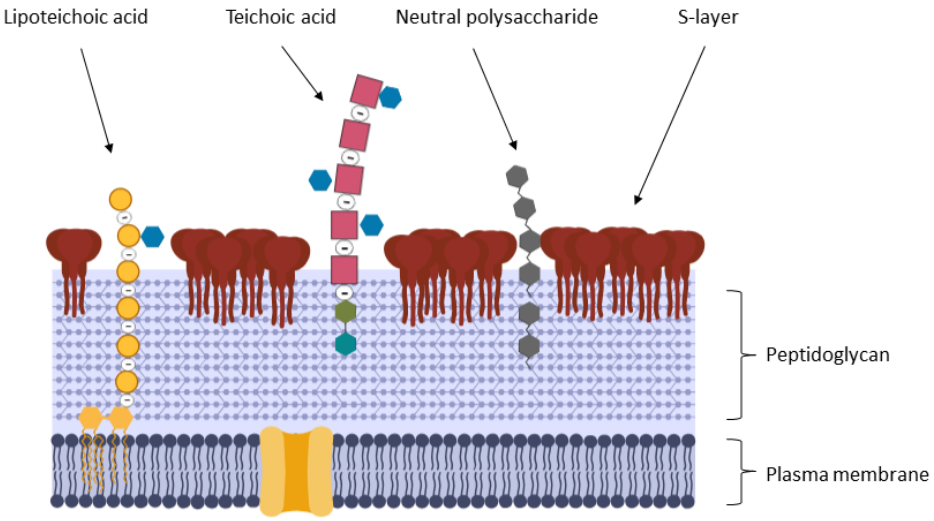


Figure 6.: Cell wall structure of *L. acidophilus*. The plasma membrane is covered by peptidoglycan, decorated with teichoic acids and neutral polysaccharides, surrounded by S-layer proteins. Adapted from *Delcour et al* [128]. Created with BioRender.com.

Peptidoglycan is a polymer composed by sugars, N-acetylglucosamine(GlucNAC) and N-acetylmuramic acid (MurNAC) cross-linked to a peptide, forming a three-dimensional mesh-like layer. The amino acid portion of the peptidoglycan of lactic acid bacteria is composed by a pentapeptide. The amino acid linked to the MurNAC is usually L-alanine, followed by D-glutamate. The other three positions can differ, depending on the species [128]. In *L. acidophilus*, the pentapeptide is composed by L-alanine/D-glutamate/L-lysine/D-alanine/D-alanine, cross-linked by D-aspartate [129, 122]. Peptidoglycan major function is to maintain cell integrity, but also gives support to proteins and teichoic acids, and defines the cell shape. Additionally, it can be a target for recognition by the eukaryotic innate immune system [130].

Teichoic acids can comprise more than 50% of the weight of the cell wall [130]. In *L. acidophilus*, two types of teichoic acids are present: *Wall teichoic acid (WTA)*, and *Lipoteichoic acid (LTA)*. The structures and abundance of these macromolecules are quite diverse, depending on the strain, stage of growth, pH of the medium, carbon source, etc. WTA consists of a polymer (usually composed by glycerol), linked to peptidoglycan by a linkage unit (a disaccharide). The glycerol units can have hydroxyl groups (OH) substituted with glucose or D-alanine. Wall teichoic acids are involved in protection of the bacteria from lysozyme [131] and antibiotics like vanomicyn [132], and play important roles in cell division [133]. In *L. acidophilus*, LTA is a polyglycerophosphate chain (with D-alanine substitutions), anchored to the membrane by a glycolipid. LTAs are related with scavenging of cations required for enzyme functions (especially Mg^{2+}) [134], adhesion to epithelial cells

[130], and interaction with bacteriophages [135]. The structures of both WTA and LTA are presented in Figure 7 [135].

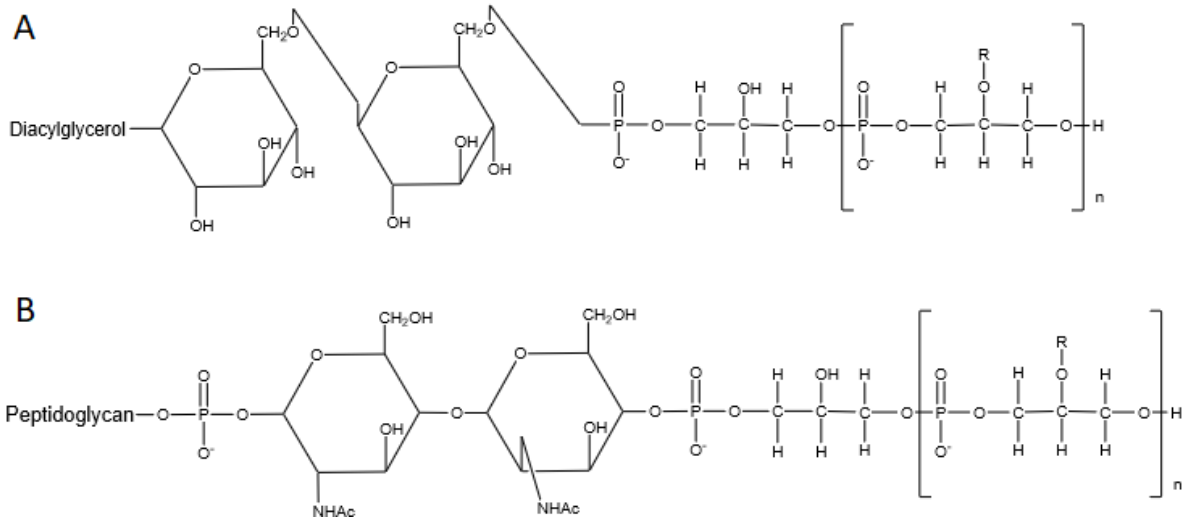


Figure 7.: General structure of TA. A: LTA with a poly-glycerol-phosphate chain and a glycolipid anchor to plasmatic membrane. B: WTA with a poly-glycerol-phosphate chain, and a dissacharide anchor to peptidoglycan; R indicate potential substitute groups of polyols chains (e.g., D-Ala, Glc, Gal, GlcNAc). Adapted from *Chapot-Chartier et al* [135]

L. acidophilus strains are reported to present an S-layer [136]. S-layers are two dimensional arrays of proteins and glycoproteins, which form a symmetric, porous, and lattice-like structure. Their major functions are related with protection from environmental hostile conditions, adhesion to mucus and epithelial cells, and scaffolding extracellular proteins, such as enzymes [137].

The term *Exopolysaccharide (EPS)* refers to bacterial polysaccharides found outside of the cell wall [138], either associated with the cell surface or secreted to the extracellular space. EPS can have different chemical compositions, molecular weight, and charge, according to the environmental conditions at their biosynthesis, and the microorganism. EPS are thought to protect the cell against desiccation, phagocytosis, antibiotics, toxic compounds, osmotic stress, and biofilm formation [139].

2.3.2.4 *Metabolic features*

Free sugars can be imported from the extracellular space through two ways: permease systems, and specific *Phosphotransferase system (PTS)*. *L. acidophilus* can only use glycolysis to ferment hexoses, making of this species a homofermentative LAB. Different carbohydrates can be used by this organism to obtain lactic acid (D and L isomers), including fructose, galactose, lactose, cellobiose, amygdalin, maltose, glucose, and stachyose.

Lactic acid is an anti-microbial compound secreted by *L. acidophilus*. Additionally, hydrogen peroxide and different bacteriocins are also produced by this species. Bacteriocins are cationic proteins produced at ribosomes, with a molecular weight ranging from 2.5 to 10 kDa, with the ability to inhibit and kill other bacteria living in the same ecological niche [140], without any repercussion to the parent strain. *L. acidophilus* bacteriocins were already studied by *Todorov et al* [127]. The spectrum of this anti-bacterial compound includes activity against different bacteria, such as *Listeria monocytogenes*, and *Listeria innocua*, two pathogenic species. Activity against cephalosporin resistant *Escherichia coli* was also identified [140]. However, no activity was detected against *Staphylococcus aureus*, *Lactobacillus sakei* and *Bacillus cereus* [127].

2.3.2.5 *Probiotic Properties and Industrial Applications*

Several LAB species, mainly from the genera *Lactobacillus* and *Bifidobacterium* are considered to be probiotic. The probiotic effects on human health of *L. acidophilus* have been well described, and are a consequence of its morphology (especially cell wall components), and production of anti-microbial drugs, such as bacteriocins. These effects include anti-inflammatory activity [141], treatment of constipation and diarrhea [142], anti-*Helicobacter pylori* infection (responsible for gastritis, peptic ulcers, and gastric cancer) [143], immune-enhancing and anticarcinogenic [144, 145], antidiabetic and antioxidant [146], cholesterol-lowering [144], antifungal and antibacterial activity [147, 148, 149], and prevention and treatment of hyperoxaluria [150].

L. acidophilus strains have been used for the production of several fermented dairy products. The most common are sweet acidophilus milk (mostly consumed by individuals with lactose intolerance) and yogurt. This species is also used in several types of cheese, usually together with other probiotics (*L. casei*, *Lactobacillus paracaisei* or *Bifidobacterium spp.*). Moreover, *L. acidophilus* is also used in non-dairy products, such as soymilk, and fermented tomato, carrot, cabbage and beet juice [122]. Besides the mentioned products, *L. acidophilus* can be obtained from nutritional supplements, usually available as freeze-dried powders, granules, and capsules [3].

3

MATERIALS AND METHODS

The *L. acidophilus* La-14 genome sequence was retrieved from the GenBank repository, with the accession number CP005926 (assembly ID ASM38967v2). A draft metabolic network for this strain was obtained in previous work. Based on the assembled network, the reconstruction process has continued until a validated *GSM* model was obtained. *Merlin* was used for the genome annotation, manual refinement of the network, and conversion of the network into a stoichiometric model. In the model validation stage, COBRApy and Optflux were used for gap identification, *in silico* simulations, and analysis of the *GSM* model. Online resources (described in subsection 2.2.2) and primary literature were used as a source of information.

3.1 PREVIOUS WORK

Enzymes annotation was performed in previous work, using *SamPler* [151], a tool included in *merlin* that allows obtaining a semi-automatic annotation. This process was based on two BLAST searches performed in March 2018, against Swiss-Prot and UniProtKB, setting an e-value threshold of 10^{-30}. Genes requiring manual annotation (according to the scoring system described in subsubsection 2.2.7.1) were annotated based on the workflow available in support material (figure S1).

A draft network was assembled by integrating the enzymes annotation and loading KEGG's metabolic data (including spontaneous reactions).

Unbalanced reactions were identified using the tool "Unbalanced reactions" available in *merlin*. The stoichiometry of these reactions was corrected, using BRENDA, BioCyc, and ChEBI to retrieve information on metabolites formula and reaction stoichiometry. Reactions whose stoichiometry was not possible to correct were removed from the model. Metabolites with variable formula but necessary to keep the model functionality, such as "Fatty acid" or "Acyl-carrier-protein", were maintained in the network. Approaches used to correct unbalanced reactions are presented in Table 4.

Table 4.: Approach used to correct the stoichiometry of unbalanced reactions.

Problem	Correction
Missing water molecules or protons	Add water molecules/protons[a], in *merlin*'s "Reactions" panel
Metabolites with missing/wrong formula	Add/correct the formula[a], in *merlin*'s "Metabolites" panel
Polymerization reactions (polymer + monomer $\rightarrow$ polymer)	Remove the polymer from the reactants, assuring that the stoichiometry is correct (monomer $\rightarrow$ polymer)[b]
Generic reactions	Remove the reaction

[a] These alterations were performed after checking information available in BRENDA, BioCyc or ChEBI.

[b] Depolymerization reactions were corrected similarly, by removing the polymer from the products.

The sub-cellular protein location was obtained using PSORTb 3.0, and the results were then loaded into *merlin*. Proteins with "Unknown" location were automatically assigned as "CYTOPLASMATIC". The integration of the predicted protein locations has doubled reactions and metabolites present in different cellular compartments.

3.2 MANUAL CURATION OF THE NETWORK

3.2.1 *Reversibility and Directionality*

Since all reactions loaded from KEGG are reversible by default, it is necessary to refine the respective reversibility and direction. This step was performed automatically using the tool available in *merlin* for this purpose, setting the *Ma H. and Zeng A.* [79] database as the data source.

Through the reconstruction process, the reversibility and direction of some reactions were manually corrected, which was performed according to the following workflow:

1. BRENDA: Information for the reaction associated with *L. acidophilus* at BRENDA was consulted, if available.

2. *L. acidophilus* La-14 database at BioCyC: Information retrieved from here was used if the reversibility and direction of the reaction are manually curated.

3. BiGG: Information from reactions present in *GSM* models of closely related organisms was preferred.

However, if no information was found in these databases, the e-quilibrator tool (www.equilibrator.weizmann.ac.il) was used. This tool determines the free Gibbs energy (ΔG^0) for a given reaction, allowing to infer its reversibility and direction. Cases in which the ΔG^0 could not be determined by this tool, such as reactions with metabolites with non-specific formula, the reaction was considered to be reversible, except for reactions involving ATP hydrolysis, and reactions associated with enzymes that can use both NAD^+ and $NADP^+$ as cofactor.

3.2.2 *Transport Reactions*

Transport reactions were automatically obtained using TranSyT. This tool retrieves information from TCDB and performs the annotation using BioSynth through TC numbers. An internal database is then created with the originated transport reactions. The automatic annotation is obtained through a BLAST of the organism's genome, and the originated reactions are included in a pathway labeled as "Transporters Pathway". Nevertheless, additional transport reactions were added in the following cases:

1. A metabolite is present in the growth medium, but does not have any uptake reaction associated;
2. It is expected the production of a given metabolite, but there are no excretion reactions associated;
3. A transport mechanism is described in the literature, but was not automatically included by TranSyT;
4. A metabolite is known to cross the membrane through simple diffusion.

In these cases, the TCDB, BiGG, and TransportDB 2.0 databases were used to retrieve information for substrates, mechanisms, and genes associated with each transport reaction. If no information was found, the transport reaction was added without a gene association to assure the model functionality.

3.2.3 *Substrate and cofactor usage*

Enzymes associated with more than one reaction required manual refinement to determine the substrate and cofactor specificity of these enzymes. Special attention was taken to oxidoreductases and enzymes whose reactions include carbohydrates.

Oxidoreductases often use NAD^+ and/or $NADP^+$ as an acceptor. The specificity of these enzymes was determined using the BRENDA, BioCyc, and UniProt databases, and

GSM models of closely related species (*B. subtilis* 168 [152], *L. Lactis ssp. lactis* IL1403 [94], and *L. plantarum* WCFS1 [89]).

Reactions including monosaccharides (glucose, fructose, galactose) and its derivates were analyzed to select the correct conformation (α and/or β) used by the corresponding enzyme. If an enzyme can catalyze a reaction using both conformations, the respective reaction was duplicated. This step was performed retrieving information from BRENDA, BioCyc, and literature.

3.2.4 *Pathway-by-pathway analysis*

Pathways loaded from KEGG were globally analyzed. Generic pathways were removed from the model, since their visualization on KEGG maps is not straightforward, and the large majority of the associated reactions are included in other pathways. Redundant pathways (those which all reactions are associated with other pathways), and pathways with less than three associated reactions were also removed. This step allowed to decrease the number of pathways, simplifying the next stages of the reconstruction process. When a pathway is removed, the associated reactions are kept in the model, not associated with any pathway.

The remaining pathways were analyzed according to the following workflow:

1. Comparison of the pathway in the model with the respective pathway on KEGG maps for *L. acidophilus* La-14;
2. If differences were identified, the genome annotation was reviewed and compared with information present in the BioCyc database for *L. acidophilus* La-14;
3. In case of doubt, primary literature was consulted.

3.2.5 *GPR associations*

The GPR associations were obtained using *merlin*, which retrieves information from KEGG BRITE. For every EC number present in the model, *merlin* searches for the respective subunits and stoichiometry of the protein complex. This information is then processed, allowing to identify the GPR rules for each protein. If errors in the GPR rules were identified through the model validation stage, the KEGG, BioCyc, and BiGG databases were used.

3.3 CONVERSION INTO A STOICHIOMETRIC MODEL

3.3.1 *Biomass and Energy Requirements*

As mentioned before, the biomass equation represents the biomass formation from complex macromolecules. The complex macromolecules included in the biomass equation were inferred from the *L. plantarum* WCFS1 model [89]: DNA, RNA, protein, lipids, exopolysaccharides, peptidoglycan, lipoteichoic acids, teichoic acids, and cofactors/vitamins.

Merlin presents the tool *e-BiomassX* [153] that generates a template for the biomass formulation. In a pathway labeled as "Biomass Pathway", this tool creates reactions for the assembly of DNA, RNA, protein, and cofactors, as well as a reaction representing the biomass equation. The biomass equation was formulated as:

$$\sum_{k=1}^{p} c_k X_k \rightarrow \textit{e-biomass},$$

where c_k are the coefficients (in g/gDW) of the macromolecule X_k, and "e-biomass" represents the biomass of the organism. The content of each macromolecule and respective precursors was determined using three different sources: experimental data, *GSM* models of closely related organisms, and *merlin*.

The mass fraction (g/g DW) of each macromolecule was determined using information retrieved from the *L. plantarum* WCFS1 *GSM* model, except for exopolysaccharides, lipoteichoic acids, and cofactors. The exopolysaccharide content was inferred from LAB *GSM* models (*L. plantarum* WCFS1, *L. casei* LCW2, and *L. lactis ssp. lactis* IL1403), as well as experimental data for *L. helveticus* and *L. rhamnosus*. The content of lipoteichoic acids and cofactors was inferred from the *L. lactis ssp. lactis* IL1403 model.

The DNA, RNA and protein precursors and respective content were automatically determined by the *e-BiomassX* tool, using the codon usage bias. The DNA assembly reaction from deoxyribonucleotides ($dXTP_k$) was formulated as:

$$\sum c_i dXTP_k \rightarrow \textit{e-DNA} + \textit{Orthophosphate},$$

where c_i represent the stoichiometry of dATP, dCTP, dTTP, and dGTP. The RNA assembly reaction from ribonucleotides (XTP_k) was formulated as:

$$\sum c_i XTP_k \rightarrow \textit{e-RNA} + \textit{Orthophosphate},$$

where c_i represent the stoichiometry of ATP, CTP, UTP, and GTP. The protein assembly reaction was formulated as:

$$\sum c_i AA - trna_i \rightarrow e\text{-}Protein + c_i trna_i + H_2O,$$

where c_i represent the stoichiometry of each aminoacyl-trna (AA-$trna_i$). The amino acid composition in the protein fraction was corrected for amino acids present in peptidoglycan, LTA, and WTA.

The precursors of lipids and peptidoglycan were determined using *L. acidophilus* experimental data. The EPS composition was inferred from experimental data of *L. acidophilus* and *L. helveticus*. The precursors of teichoic acids were included according to the *B. subtillis* 168 model [152]. For lipoteichoic acids, experimental data of *L. helveticus* was used. The cofactors and vitamins included in the model were obtained from *merlin* (universal cofactors), *L. plantarum* WCFS1 model, and primary literature.

Although fatty acids are not directly included in the biomass reaction, they are components of lipids and lipoteichoic acids. Hence, a metabolite representing the average fatty acid composition (”Fatty acid”) was created and included in a reaction named as ”R-Fatty_acid”. The molecular weight of this metabolite allows determining the molecular weight of both lipids and lipoteichoic acids.

The GAM energy requirement was determined using information retrieved from the *GSM* model of *L. plantarum* WCFS1, and was added to the ”R_e-biomass” reaction. The NGAM energy requirement initial flux was obtained from the same metabolic model and was then adjusted to experimental data. The adjustment was performed by plotting the growth rate *vs* ATP maintenance value. The NGAM was included in the model in a simple ATP hydrolysis reaction, constraining the lower and upper bounds according to the defined value.

3.3.2 *Growth medium and exchange reactions*

A literature search for experimental data of *L. acidophilus* was performed, allowing to determine a *Chemically Defined Medium (CDM)* for this species. A minimal growth medium was defined [123], containing all the compounds needed to allow the growth of *L. acidophilus*: carbon, nitrogen, sulfur, and phosphorus sources, and all the auxotrophies of this species. This medium was used for the model troubleshooting stage, where the model was tested to produce all the biomass precursors. Experimental procedures usually apply a rich medium, containing growth-promoting metabolites, such as non-essential amino acids, nucleotides and vitamins. This medium was defined [123] and used to access the growth rate by comparison with experimental data. The minimal and rich CDM are available in support material (Table S3).

Exchange reactions represent the growth medium in a metabolic model, including both consumed and produced metabolites. *Merlin* allows creating these reactions automat-

ically based on the transport reactions present in the model. The lower bound of exchange reactions for metabolites present in the CDM was changed from 0 to -999999, and the upper bound was maintained as 999999. The carbon source (glucose or other carbohydrates) was selected as the growth-limiting factor, so the corresponding lower bound was settled according to experimental data. *In silico* simulations in anaerobiosis were performed by constraining the lower bound of the oxygen exchange reaction to zero. Aerobiosis was defined by unconstraining the oxygen uptake.

3.3.3 *Model troubleshooting*

A tool available in *merlin* (BioISO) was used to identify and solve gaps blocking the synthesis of biomass precursors. This tool uses the COBRApy *toolbox* to retrieve information regarding metabolites and reactions, allowing also the creation of exchange, transport, sink and demand reactions. Whenever a biomass precursor was not being produced (or available in the medium), a traceback was performed to identify gaps, namely missing enzymatic and transport reactions, errors in the genome annotation, and reactions with wrong reversibility or direction. After the identification of the gap, online databases (KEGG, BRENDA, BioCyc, and BiGG) and primary literature were consulted to solve it. This iterative process was repeated until a positive flux through the biomass reaction was achieved.

3.4 GAP FILLING

Dead-end metabolites and blocked reactions were automatically identified by the "Blocked reactions" tool, available in *merlin*. The respective pathway was analyzed using the "Draw in Browser" button, which allows highlighting both metabolites and enzymes on KEGG maps. Figure 8 shows an example of a view on KEGG maps for the "Glycine, serine and threonine metabolism" pathway.

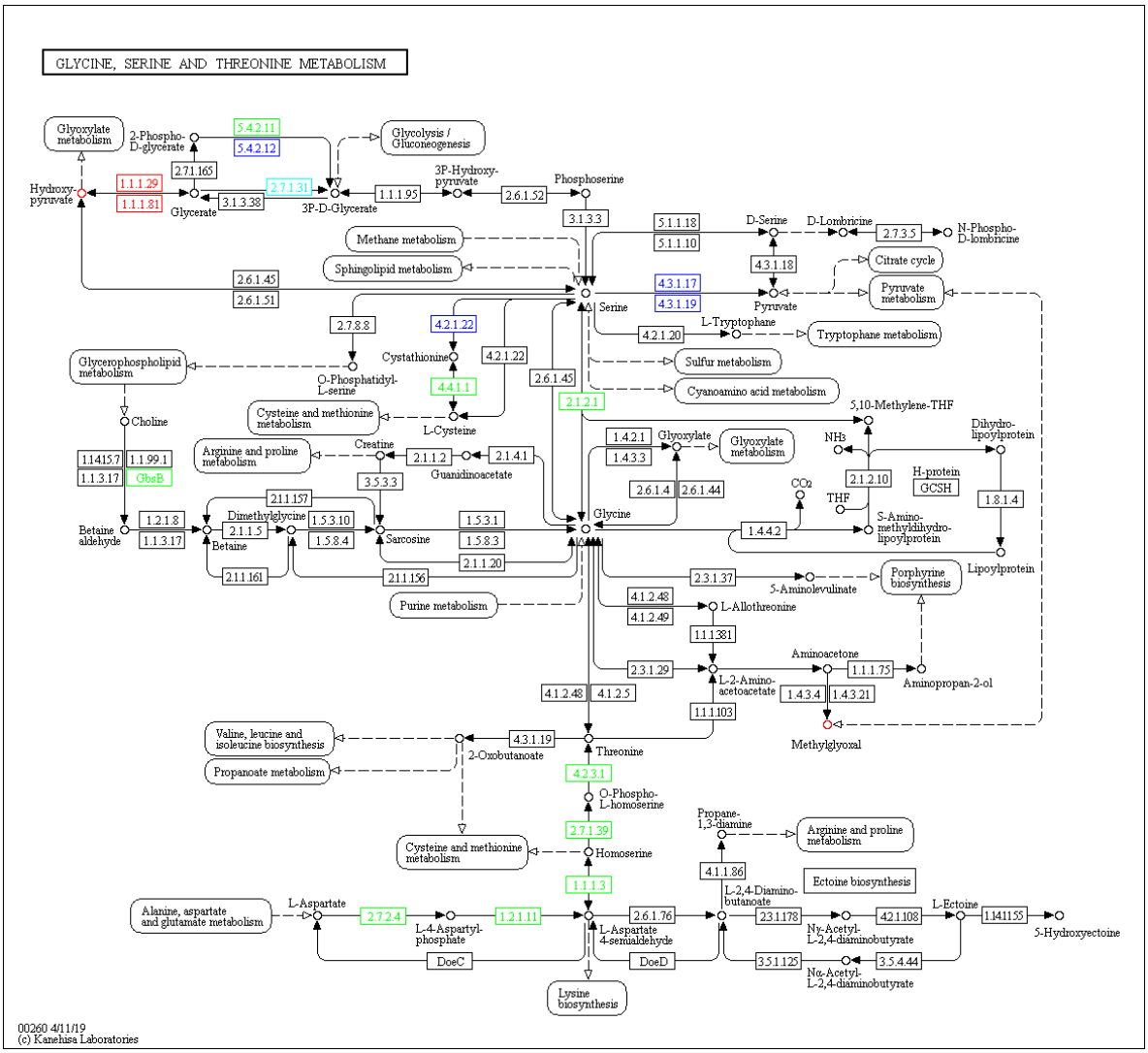


Figure 8.: Example of a KEGG metabolic map (”Glycine, serine and threonine metabolism”) colored by *merlin*.

The color scheme represented above has the following key:

1. Green: The enzyme and reaction are present in the model, and the reaction is not blocked;

2. Dark Blue: The reaction is present in the pathway, but the correspondent enzyme is missing; however, another enzyme present in another pathway might be associated with the reaction;

3. Cyan Blue: The enzyme is present in the model, but the reaction is connected to a dead-end;

4. Red: The enzyme is present in the network, but the reaction is a blocked and associated with a dead-end metabolite;

5. Colorless (Black): The enzyme is not present in the network.

Each blocked reaction and dead-end metabolite was analyzed, using information present on BioCyc, BiGG, BRENDA, and primary literature. Blocked reactions for which no information was found were maintained in the model.

3.5 MODEL VALIDATION

The model validation stage was performed using Optflux v3.4.0 for simulation and analysis of the metabolic model, followed by comparison with experimental data and literature. This process allowed to identify and correct inconsistencies between the model and available information on the metabolism of *L. acidophilus*. Moreover, the validation of the reconstructed metabolic model was compared with the semi-automatic AGORA model for *L. acidophilus* NCFM (available at www.vmh.life/#microbe/Lactobacillus_acidophilus_NCFM) [154]. The same environmental conditions were applied to both models if possible.

Optflux presents several tools for simulation and analysis of *GSM* models. Using an available plug-in, the model was automatically loaded from *merlin* to Optflux without generating an SBML file, simplifying and accelerating the process.

Different environmental conditions were settled by manipulating the lower and upper bound of reactions, including not only drains but also transport and enzymatic reactions. This allows changing the growth medium characteristics, remove or change the reversibility of reactions, directly in Optflux without introducing alterations in the original model.

In silico simulations were performed using pFBA maximizing the biomass production, by setting the biomass reaction as the objective function. The originated panel presents several relevant information, like the flux of all reactions, and consumed and produced metabolites. Additionally, Optflux also allows simulating knockout, and over and underexpression of genes and reactions. Knockout simulations were performed using LMOMA as simulation method. The determination of critical genes and reactions was also performed with Optflux, which uses FBA to perform this task.

Various tests were applied to validate the metabolic model, including:

1. Spontaneous growth;

2. Auxotrophies;

3. Growth rate assessment;

4. Secreted products;
5. Alternative element sources;
6. Gene essentiality.

Spontaneous growth was tested by setting the lower bound of all drains to zero.

Auxotrophies were verified by the single omission of amino acids, pyrimidines, purines, and cofactors from the CDM. The requirement for these metabolites was verified and compared with information described in literature.

To assess the growth rate, the rich medium was used for comparison and adjustment to experimental data, in anaerobic conditions. *In silico simulations* in aerobic conditions were also performed for qualitative comparison with experimental data.

Literature was consulted to collect information about expected secreted compounds by *L. acidophilus*, namely fermentation end-products. The production of the expected metabolites was tested *in silico*, using different environmental conditions.

The growth with different carbon sources and the amino acid requirements were tested and compared qualitatively with available data. The performance measure calculations used to evaluate the ability of the model to present growth using different carbohydrates were determined as follows:

1. Sensitivity: TPR = TP / (TP + FN);
2. Specificity: SPC = TN / (FP + TN);
3. *Negative Predictive Value (NPV)*: NPV = TN / (TN + FN);
4. Accuracy: ACC = (TP + TN) / (TP + TN + FN + FP).

, where TP represents true positive, FN false negative, TN true negative, and FP false positive.

Critical genes and critical reactions were identified using the tool present in Optflux for that purpose, under different environmental conditions: minimal and rich medium, and aerobiosis/anaerobiosis.

The main metabolic pathways (carbohydrate and pyruvate metabolism, biosynthetic capabilities) were analyzed according to *in silico* simulations and information available in literature.

4

RESULTS AND DISCUSSION

4.1 GENOME ANNOTATION

4.1.1 *Enzymes annotation*

The enzymes annotation provided 716 genes accepted as enzyme encoding genes, resulting in 433 different EC numbers. Throughout the manual curation and model validation stages, the annotation was updated to correct the model, according to experimental data and available information. These changes, together with the removal of generic reactions, resulted in a reduction of the number of genes and enzymes present in the model.

Thus, the model presents 315 different EC numbers, four of which are incomplete: Fatty acid kinase (EC 2.7.-.-); Major cardiolipin synthase ClsA (EC 2.7.8.-); Glycosyltransferases (EC 2.4.-.-/2.4.1.-). The fatty acid kinase, described in different *Lactobacillales* [155], has been associated with the phosphorylation of exogenous fatty acids. However, a complete EC number was not found for this enzyme. The Major cardiolipin synthase CLsA catalyzes the synthesis of cardiolipin from phosphatidylglycerol and has just a preliminary EC number assigned at BRENDA. The glycosyltransferases with the EC numbers 2.4.-.- and 2.4.1.-, belonging to the *eps* gene cluster, were included in the biomass reaction for exopolysaccharides. Table 5 shows the final distribution of EC number classes in the model.

Table 5.: Enzymes distribution in the model according to the EC number classification system.

EC class	Percentage of EC numbers (%)
EC 1: Oxidoreductases	14.3
EC 2: Transferases	37.8
EC 3: Hydrolases	17.1
EC 4: Lyases	9.5
EC 5: Isomerases	6.4
EC 6: Ligases	14.6
EC 7: Translocases	0.3

Oxidoreductases represent 14.3% of the enzymes included in the model. These enzymes are associated with the reduction and oxidation of numerous substrates, using mainly NAD^+ and $NADP^+$ as a coenzyme. Lactate dehydrogenase, an important enzyme of the *L. acidophilus* metabolism, belongs to this class. Genes encoding both L and D-lactate dehydrogenases were identified in the genome, in agreement with the production of both isomers of lactic acid by *L. acidophilus* reported in the literature [114].

Transferases are the most representative class in the model, comprising 37.8% of total enzymes. From these, 50% are phosphate transferring enzymes (EC 2.7), mainly involved in nucleotide metabolism and carbohydrate phosphorylation.

Hydrolases included in the model (17.1%) are predominantly associated with hydrolysis reactions acting on ester (EC 3.1), sugar (EC 3.2), and carbon-nitrogen (EC 3.5) bonds. These enzymes allow the catabolism of lipids, carbohydrates, and amino acids.

Most of lyases included in the model are decarboxylases (EC 4.1.1), including the oxalyl-CoA carboxy-lyase. This enzyme plays a role in the catabolism of oxalate and has been extensively studied in *L. acidophilus* La-14, due to the oxalate-degrading capability of this strain [156, 157, 158, 159].

Isomerases represent just 6.4% of the enzymes, including racemases, epimerases, and mutases. Racemases participate in the conversion of L-amino acids to the respective D isomer. D-alanine and D-Aspartate are especially important since these amino acids are included in the structure of biomass components, like peptidoglycan and teichoic acids.

The model contains only two ligases subclasses: ligases forming carbon-oxygen bonds (EC 6.1), and ligases forming carbon-nitrogen bonds (EC 6.3). The first subclass includes only amino acid-tRNA ligases, present in the "Aminoacyl-tRNA biosynthesis" pathway. Ligases forming carbon-nitrogen bonds are mainly associated with anabolic pathways to produce macromolecules containing peptide bonds, like the peptidoglycan.

The Translocases class (EC 7) is a new class representing a group of enzymes that catalyze the movement of ions or molecules across membranes. These enzymes were previously classified as ATPases (EC 3.6.3.-), although the hydrolytic reaction is not the major function of these enzymes. *L. acidophilus* presents the F-type H^+-transporting ATPase (EC 7.1.2.2), which has been reported to play an important role in maintaining the pH gradient across the membrane, by exporting protons to the extracellular space, with ATP hydrolysis. This translocase activity enhances the survival ability of this species in low pH environments [160].

4.1.2 *Compartmentalization*

Four different compartments were identified using the PSORTb 3.0 tool: cytoplasm (865 proteins), cytoplasmic membrane (551 proteins), cell wall (49 proteins) and extracel-

lular space (54 proteins). The localization of 355 proteins was not possible to determine, so these were assumed as cytoplasmatic. Nevertheless, reactions and metabolites present in the model are only located in the cytoplasm (enzymatic and spontaneous reactions), cytoplasmatic membrane (transport reactions) and extracellular space (exchange reactions).

4.1.3 *Transport Reactions*

The metabolic model comprises 213 transport reactions, 172 of which were automatically created and included in the model by the *TranSyt* tool. Despite these, 41 transport reactions were manually added, including 18 simple diffusion reactions, 20 transport reactions associated with genes, and 3 reactions not associated with any gene. Table 6 shows a sample of the most relevant transport reactions manually added to the model.

Table 6.: Sample of manually added transport reactions.

Reaction name	**Transport reaction**	**Gene(s) associated**	**Source/ Reference**
T_PEP_Sucrose	Sucrose (EXTR) + Phosphoenolpyruvate (CYTOP) => Sucrose 6-phosphate (CYTOP) + Pyruvate (CYTOP)	LA14_0397, LA14_1706	[161] / KEGG
T_alpha-D-Galactose	α-D-Galactose (EXTR) + H^+ (EXTR) <=> α-D-Galactose (CYTOP) + H^+ (CYTOP)	LA14_1374, LA14_1115	[161]
T_aminobutanoate_glutamate_antiport	L-Glutamate (EXTR) + 4-Aminobutanoate (CYTOP) <=> L-Glutamate (CYTOP) + 4-Aminobutanoate (EXTR)	LA14_0057	[162, 163]
T_aspartate_alanine_antiport	L-Aspartate (EXTR) + L-Alanine (CYTOP) => L-Aspartate (CYTOP) + L-Alanine (EXTR)	LA14_1697	TransportDB 2.0 / [164]
T_Oxalate	Oxalate (CYTOP) <=> Oxalate (EXTR)	—	[157]
T_Thymine	Thymine (EXTR) + H^+ (EXTR) => Thymine (CYTOP) + H^+ (CYTOP)	—	BiGG

The sucrose uptake through the PTS by *L. acidophilus* is described in the literature [161]. Transport reactions through this mechanism for cellobiose, maltose, N-acetylglucosamine, and N-acetylmuramate were also manually added, based on information retrieved from KEGG, TransportDB 2.0, and literature.

Like the F-ATPase, the aminobutanoate-glutamate and aspartate-alanine antiporters have been reported to contribute to the proton gradient across the membrane, enhancing the

ability of *L. acidophilus* to survive in low pH environments [162, 163, 164]. These reactions were created based on literature and information in the TransportDB 2.0 database.

In *Oxalobacter ormigenes*, the first oxalate-degrading obligate anaerobe to be described in humans, the oxalate transport is associated with an oxalate-formate antiporter [165]. However, the gene associated with this transport protein was not found in the *L. acidophilus* genome. *Azcarate et al* [157] proposed that oxalate crosses the membrane through diffusion, so a simple diffusion transport reaction for this compound was created.

No transport protein was found for thymine. However, this pyrimidine is present in the CDM, thus a transport reaction was created. The mechanism of the reaction (symport with H^+) was inferred from the *GSM* models for *L. lactis* MG1363 and *B. subtilis* 168, available at the BiGG database.

LAB usually present four different transport mechanisms, according to the TC number classification system: class 1, channels; class 2, secondary carrier-type facilitators (uniport, symport, and antiport); class 3, primary active transporters (namely the ABC-binding cassette system); class 4, group translocators (PTS). Since LAB do not present transmembrane electron carriers, class 5 of transport mechanisms is usually absent in these bacteria. Figure 7 shows the distribution of each one of these mechanisms in the model.

Table 7.: Distribution of transport mechanisms in the model.

TC class	Percentage of transport mechanism (%)
TC 1: Channels	7.7
TC 2: Secondary transporters	37.4
TC 3: ABC-binding cassette	49.7
TC 4: Group translocators (PTS)	5.2
TC 5: Electron carriers	0

Transport through channels represents just 7.7% of the transport reactions included in the model. The Large-conductance mechanosensitive channel (LA14_0418) is responsible for the transport of several intracellular inorganic compounds, such as orthophosphate or bicarbonate, to the extracellular space in response to osmotic pressure. Although ammonia can cross the membrane via diffusion, *L. acidophilus* presents a channel responsible for its transport.

The second transport mechanism represents 37.4% of the transport reactions. This mechanism can be divided into symport (49 reactions), antiport (8 reactions), and uniport (16 reactions). Antiport and symport reactions use mainly protons as co-substrate, highlighting the relevance of the proton gradient across the membrane for the successful growth and survival of this species.

About half of the transport reactions included in the model (49.7%) use the ABC-binding cassette mechanism. Amino acids, sugars, nucleotides, and several other metabolites can be transported through this system with ATP consumption.

The transport of free sugars can occur through the PTS using phosphoenolpyruvate as a phosphate donor. The resulting sugar is converted to glucose 6-phosphate, following the EMB pathway. Although the percentage of reactions through the PTS (5.2%) is lower than the remaining ones, this transport system is essential for *Lactobacilli*, since it is the main transport mechanism for the uptake of carbohydrates.

4.2 MANUAL CURATION OF THE NETWORK

4.2.1 *Substrate and cofactor usage*

Reactions using NADH or NADPH were analyzed to determine the cofactor specificity of the respective enzyme. Three enzymes were wrongly associated to reactions using NADH: glutathione reductase (EC 1.8.1.7), UDP-N-acetylmuramate dehydrogenase (EC 1.3.1.98), and dihydrofolate reductase (EC 1.5.1.3). These enzymes use NADPH with high affinity, presenting an insignificant/inexistent activity with NADH [166, 167]. Hence, reactions associated with these enzymes using NADH were removed (R03191, R00094, R00936, R00937, R02235).

The specificity of enzymes for the glucose, fructose, and galactose conformations was determined. Hence, only α and β-monosaccharides are present in the model. Table 8 shows a sample of the alterations performed in this step.

Table 8.: Alterations performed to refine the specificity of enzymes for glucose, galactose, and fructose conformations.

Reaction	EC number	Enzyme	Alteration
R00028	3.2.1.20	α-glucosidase	D-Glucose $\rightarrow$ α-D-Glucose
R00765	3.5.99.6	glucosamine-6-phosphate deaminase	D-Fructose 6-Pi $\rightarrow$ β-D-Fructose 6-Pi
R00768	2.6.1.16	glucosamine 6-phosphate synthase	D-Fructose 6-Pi $\rightarrow$ β-D-Fructose 6-Pi
R00803	2.4.1.7	sucrose phosphorylase	D-Fructose $\rightarrow$ β-D-Fructose
R00837	3.2.1.93	phosphotrehalase	D-Glucose $\rightarrow$ α-D-Glucose
R00838	3.2.1.122	phospho-α-glucosidase	D-Glucose $\rightarrow$ α-D-Glucose
R00839	3.2.1.86	phospho-β-glucosidase	D-Glucose $\rightarrow$ β-D-Glucose
R01555	2.4.1.8	maltose phosphorylase	Reaction duplicated
R01718	3.2.1.10	oligo-1,6-glucosidase	D-Glucose $\rightarrow$ α-D-Glucose
R01791	3.2.1.10	oligo-1,6-glucosidase	D-Glucose $\rightarrow$ α-D-Glucose
R02410	3.2.1.26	β-fructofuranosidase	D-Fructose $\rightarrow$ β-D-Fructose
R03635	3.2.1.26	β-fructofuranosidase	D-Fructose $\rightarrow$ β-D-Fructose
R03921	3.2.1.26	β-fructofuranosidase	D-Fructose $\rightarrow$ β-D-Fructose
R01101	3.2.1.22	$\alpha - galactosidase$	D-Galactose $\rightarrow$ α-D-Galactose
R01103	3.2.1.22	$\alpha - galactosidase$	D-Galactose $\rightarrow$ α-D-Galactose

Both α and β-conformations of glucose and galactose are present in the model. On the other hand, only the β-conformation of fructose (and fructose 6-phosphate) was included, mainly in the EMB pathway. Reaction R01555 is catalyzed by maltose phosphorylase (EC 2.4.1.8), which transfers phosphate to maltose, originating β-D-glucose 1-phosphate, and α or β-D-Glucose. Hence, this reaction was duplicated, to account for both glucose conformations.

4.2.2 *Pathway-by-pathway analysis*

Before proceeding to a pathway-by-pathway analysis, generic and redundant pathways were removed. From the 108 pathways initially loaded from KEGG, 58 were removed (support material S1). When a pathway is removed, the associated reactions are kept in the model, in the remaining pathways or not associated with any pathway. Table 9 presents the pathways present in the model and the respective number of associated reactions.

Table 9.: Pathways available in the model and respective number of reactions.

Pathway	Number of reactions
Alanine, aspartate and glutamate metabolism	15
Amino sugar and nucleotide sugar metabolism	23
Aminoacyl-tRNA biosynthesis	24
Arginine and proline metabolism	08
Arginine biosynthesis	04
Biomass Pathway	12
Butanoate metabolism	05
Carbon fixation pathways in prokaryotes	10
Chloroalkane and chloroalkene degradation	03
Citrate cycle (TCA cycle)	03
Cyanoamino acid metabolism	03
Cysteine and methionine metabolism	27
D-Glutamine and D-glutamate metabolism	05
Drains pathway	164
Drug metabolism - other enzymes	12
Fatty acid biosynthesis	16
Folate biosynthesis	07
Fructose and mannose metabolism	06
Galactose metabolism	27
Glutathione metabolism	09
Glycerolipid metabolism	14
Glycerophospholipid metabolism	15
Glycine, serine and threonine metabolism	14
Glycolysis / Gluconeogenesis	21
Glyoxylate and dicarboxylate metabolism	07
Isoquinoline alkaloid biosynthesis	03
Limonene and pinene degradation	04
Lysine biosynthesis	13
Methane metabolism	08
Monobactam biosynthesis	06
Nicotinate and nicotinamide metabolism	11
One carbon pool by folate	13
Pantothenate and CoA biosynthesis	09
Pentose phosphate pathway	19
Peptidoglycan biosynthesis	12

Continued on next page

Table 9 – continued from previous page

Pathway	Number of reactions
Porphyrin and chlorophyll metabolism	04
Propanoate metabolism	05
Purine metabolism	69
Pyrimidine metabolism	58
Pyruvate metabolism	16
Riboflavin metabolism	06
Selenocompound metabolism	07
Spontaneous	19
Starch and sucrose metabolism	19
Sulfur metabolism	04
Terpenoid backbone biosynthesis	12
Thiamine metabolism	06
Transporters pathway	213
Valine, leucine and isoleucine biosynthesis	03
Valine, leucine and isoleucine degradation	06
Vitamin B6 metabolism	08
Non-Associated to Pathway	70

Several pathways regarding the metabolism of carbohydrates, such as ”Galactose metabolism” or ”Starch and sucrose metabolism”, were included. The EMB pathway is complete, and several reactions of the pentose phosphate pathway are also available. Only three reactions of the citrate cycle are present (R00341, R01082, R02164).

Pathways related to the metabolism of amino acids have a low number of reactions associated, reflecting the high requirements of *L. acidophilus* for amino acids. Nevertheless, the ”Cysteine and methionine metabolism” is an exception, presenting 27 associated reactions.

At this stage, 105 generic and glycan-associated reactions were removed (Table S2). Generic reactions include metabolites such as ”Protein” or ”acceptor”, and are general representations of biochemical processes. Reactions including KEGG Glycans were also removed since these reactions are just duplicates of reactions containing KEGG Compounds.

4.3 CONVERSION OF THE NETWORK INTO A STOICHIOMETRIC MODEL

4.3.1 *Biomass composition and Energy Requirements*

The biomass equation includes nine different entities, labelled as "e-Metabolites", representing the complex macromolecules found in *L. acidophilus*. The protein, DNA, RNA, lipid, and teichoic acid contents were obtained from the *L. plantarum* WCFS1 model. Since the LTA's and cofactors contents were not experimentally determined in that study, the stoichiometry of these two entities was inferred from the *L. lactis ssp. lactis* IL1403 GSM model. The exopolysaccharide content and composition present high variability between species and strains. Therefore, the content of EPS was determined using GSM models of different LAB (*L. plantarum* WCFS1, *L. lactis* ssp. lactis IL1403, *L. casei* LC2W), and experimental data available for *L. delbrueckii* [168] and *L. rhamnosus* [169]. The stoichiometry (g/g DW) of the biomass macromolecules in *L. acidophilus*, *L. lactis*, and *L. plantarum* are presented in Table 10.

Table 10.: Biomass composition of *L. acidophilus* La-14, *L. plantarum* WCFS1, and *L. lactis ssp. lactis IL*1403. The stoichiometry is represented in grams of macromolecule per gram of biomass (g/gDW).

e-Metabolite	*L. plantarum* WCFS1	*L. lactis ssp. lactis IL*1403	*L. acidophilus* La-14	Reference
e-Protein	0.261[a]	0.460	0.288[a]	[89]/*merlin*
e-DNA	0.019	0.023	0.019	[89]/*merlin*
e-RNA	0.090	0.107	0.090	[89]/*merlin*
e-Lipid	0.063	0.034	0.063	[89]/[170]/[171]
e-Peptidoglycan	0.145	0.118	0.145	[89]/[129]
e-Exopolysaccharide	0.099	0.120	0.119	[89]/[94]/[97]/ [168]/ [169]
e-Teichoic acid	0.138	—	0.138	[89]/[152]
e-Lipoteichoic acid	0.041	0.080	0.080	[94]/[172]
e-Cofactor	—	0.058	0.058	[94]/[173]/*merlin*

[a] the original protein content is 29.9 g/gDw. These values account for the content of amino acids in other macromolecules.

Lactobacilli present a low protein content (around 30% of the biomass) when comparing to other gram-positive bacteria [89, 97, 152]. The 28.8% inferred from the *L. plantarum model* is in agreement with this characteristic.

The EPS content can be very different among different LAB species and even strains. In *L. casei* this value was reported as 8.1% [97], but *L. delbrueckii* presents almost twice EPS

content (15.5%) [168]. The 11.9% determined for *L. acidophilus* is between these two values and is similar to the 12% of *L. lactis*.

Wall teichoic acids and lipoteichoic acids account for 46% of the total cell wall weight, which is close to the approximately 50% described in the literature for gram-positive bacteria [128].

The amino acid contents (Table 11) were automatically obtained by *merlin* using the *e-BiomassX* tool.

Table 11.: Protein composition of *L. acidophilus*.

Precursor	Formula	Stoichiometry (g/g)
L-Alanyl-tRNA	$C13H22NO11PR2(C5H8O6PR)n$	0.031
L-Arginyl-tRNA(Arg)	$C21H33N9O11PR(C5H8O6PR)n$	0.052
L-Asparaginyl-tRNA(Asn)	$C14H23N2O12PR2(C5H8O6PR)n$	0.083
L-Aspartyl-tRNA(Asp)	$C14H22NO13PR2(C5H8O6PR)n$	0.056
L-Cysteinyl-tRNA(Cys)	$C18H26N6O11PSR(C5H8O6PR)n$	0.005
L-Glutamyl-tRNA(Glu)	$C20H28N6O13PR(C5H8O6PR)n$	0.058
Glutaminyl-tRNA	$C20H29N7O12PR(C5H8O6PR)n$	0.044
Glycyl-tRNA(Gly)	$C12H20NO11PR2(C5H8O6PR)n$	0.033
L-Histidyl-tRNA(His)	$C16H24N3O11PR2(C5H8O6PR)n$	0.025
L-Isoleucyl-tRNA(Ile)	$C21H32N6O11PR(C5H8O6PR)n$	0.080
L-Leucyl-tRNA	$C21H32N6O11PR(C5H8O6PR)n$	0.093
L-Lysyl-tRNA	$C16H29N2O11PR2(C5H8O6PR)n$	0.085
L-Methionyl-tRNA	$C20H30N6O11PSR(C5H8O6PR)n$	0.032
L-Phenylalanyl-tRNA	$C19H26NO11PR2(C5H8O6PR)n$	0.057
L-Prolyl-tRNA	$C15H24NO11PR2(C5H8O6PR)n$	0.031
L-Seryl-tRNA	$C13H22NO12PR2(C5H8O6PR)n$	0.046
L-Threonyl-tRNA	$C14H24NO12PR2(C5H8O6PR)n$	0.053
L-Tryptophanyl-tRNA	$C26H31N7O11PR(C5H8O6PR)n$	0.018
L-Tyrosyl-tRNA	$C24H30N6O12PR(C5H8O6PR)n$	0.058
L-Valyl-tRNA	$C20H30N6O11PR(C5H8O6PR)n$	0.061
		Total: 1

Alanine, glutamate, aspartate, and lysine were included in other biomass macromolecules (peptidoglycan, WTA, and LTA). Thus, the amounts of those amino acids in these biomass components were accounted for the protein content.

The ribonucleotide and deoxyribonucleotide composition, represented in the triphosphate form, was determined automatically by *merlin*. The DNA and RNA structural units and respective stoichiometry are presented in Table 12.

Table 12.: DNA and RNA composition of *L. acidophilus*.

Macromolecule	***Precursor***	Formula	Stoichiometry (g/g)
DNA	dTTP	$C10H17N2O14P3$	0.323
	dGTP	$C10H16N5O13P3$	0.184
	dATP	$C10H16N5O12P3$	0.330
	dCTP	$C9H16N3O13P3$	0.163
			Total: 1
RNA	ATP	$C10H16N5O13P3$	0.260
	GTP	$C10H16N5O14P3$	0.309
	CTP	$C9H16N3O14P3$	0.215
	UTP	$C9H15N2O15P3$	0.216
			Total: 1

Although fatty acids are not directly represented in the biomass equation, they are present in the lipids and lipoteichoic acids composition. The fatty acid content of *L. acidophilus* was determined experimentally by *Veerkamp et al* [170]. Hence, a compound representing the average fatty acid composition, named as "Fatty_acid", was created and included in the model.

Table 13.: Fatty acid composition of *L. acidophilus* according to *Veerkamp et al* [170].

Precursor	Formula	Stoichiometry (g/g)
Tetradecanoic acid	$C14H28O2$	0.021
Hexadecanoic acid	$C16H32O2$	0.368
Hexadecenoic acid	$C16H30O2$	0.082
Octadecanoic acid	$C18H36O2$	0.032
Octadecenoic acid	$C18H34O2$	0.378
Lactobacillic acid	$C19H36O2$	0.119
		Total: 1

Hexadecanoic and octadecenoic acids are the fatty acids presenting higher amounts in *L. acidophilus*. Lactobacillic acid (11R,12S-Methylene-octadecanoic acid) is a cyclopropane fatty acid, produced from octadecenoic acid and S-adenosylmethionine. This fatty acid,

identified in several *Lactobacilli*, has been associated with the stabilization of the membrane and resistance to acid shock, by decreasing the permeability of the membrane to protons [174].

Table 14 shows the lipid composition of *L. acidophilus* based on available experimental data [171]. The Fatty acid compound described above was used to determinate the molecular weight of each phospholipid.

Table 14.: Lipid composition of *L. acidophilus*.

Precursor	Formula	Stoichiometry (g/g)
Phosphatidylglycerol	$C8H13O10PR2$	0.888
3-O-L-Lysyl-1-O-phosphatidylglycerol	$C14H25N2O11PR2$	0.036
Cardiolipin	$C13H18O17P2R4$	0.077
		Total: 1

Lipids included in the model are essentially phospholipids. Part of the lipoteichoic acid structure has lipidic nature, but it was included in the "R_e-lipoteichoic_acid" reaction. Phosphatidylglycerol is the most abundant phospholipid in *L. acidophilus*, comprising almost 90% of the lipid composition. 3-O-L-Lysyl-1-O-phosphatidylglycerol and cardiolipin present low percentages and are obtained from phosphatidylglycerol, through a lysyltransferase (EC 2.3.2.3), and cardiolipin synthase (EC 2.7.8.-), respectively.

The peptidoglycan composition (Table 15) was determined according to its reported structure [129].

Table 15.: Peptidoglycan composition of *L. acidophilus*.

Precursor	Formula	Stoichiometry (g/g)
L-Alanine	$C3H7NO2$	0.067
D-Glutamate	$C5H9NO4$	0.121
UDP-N-acetyl-α-D-glucosamine	$C17H27N3O17P2$	0.191
L-Lysine	$C6H14N2O2$	0.121
D-Alanine	$C3H7NO2$	0.134
D-Aspartate	$C4H7NO4$	0.107
UDP-N-acetylmuramate	$C20H31N3O19P2$	0.259
		Total: 1

Similarly to *L. lactis*, the peptidoglycan of *L. acidophilus* presents L-lysine, instead of meso-2,6-Diaminopimelate. The difference between these two structures is based on the presence of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase (EC 6.3.2.7)

or UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase (EC 6.3.2.13). The EC 6.3.2.13 was assigned to the gene *murE* (LA14_1809) in the genome annotation stage. Hence, this annotation was reviewed and corrected, by assigning the EC number 6.3.2.7 to this gene.

The EPS composition can be very different depending on the species, strain, and growth conditions. Thus, experimental data for both *L. acidophilus* [175, 176] and *L. helveticus* [177, 178, 179, 180, 181, 182] was accounted, resulting in the EPS composition presented in Table 16.

Table 16.: Exopolysaccharide composition of *L. acidophilus*.

Precursor	Formula	Stoichiometry (g/g)
UDP-glucose	$C15H24N2O17P2$	0.554
UDP-N-acetyl-α-D-glucosamine	$C17H27N3O17P2$	0.050
UDP-α-D-galactose	$C15H24N2O17P2$	0.396
		Total: 1

Unlike more phylogenetically distant *Lactobacilli*, like *L. rhamnosus* and *L. paracasei*, *L. acidophilus* and *L. helveticus* do not seem to present rhamnose in the composition of EPS. Instead, two *L. acidophilus* strains were reported to include N-acetyl-glucosamine, in lower amounts than glucose and galactose [175, 176]. The gene cluster for the synthesis of EPS was associated with the "R_e-Exopolysaccharide" reaction. This cluster is composed of 14 genes, including *epsA-epsF*, *epsI*, *epsJ*, and five variable proteins, including glycosyltransferases and polysaccharide polymerases [118]. The gene *epsH*, responsible for the inclusion of rhamnose in the EPS of *L. lactis* [183], was not identified in the *L. acidophilus* La-14 genome, explaining the absence of rhamnose in the EPS produced by this species.

L. plantarum and *L. acidophilus* present different wall teichoic acids structures: the polymer of the first species has a ribitol-phosphate chain, while *L. acidophilus* presents glycerol-phosphate. Hence, the structure and chain length of WTA were inferred from the *B. subtillis* 168 GSM model (Table 17).

Table 17.: Teichoic composition of *L. acidophilus*.

Precursor	Formula	Stoichiometry (g/g)
UDP-glucose	*C*15*H*24*N*2*O*17*P*2	0.226
UDP-N-acetyl-α-D-glucosamine	*C*17*H*27*N*3*O*17*P*2	0.015
Glycerol 3-phosphate	*C*3*H*9*O*6*P*	0.645
D-Alanine	*C*3*H*7*NO*2	0.099
UDP-N-acetyl-D-mannosamine	*C*17*H*27*N*3*O*17*P*2	0.015
		Total: 1

The WTA structure presents an anchor composed of N-acetyl-glucosamine and N-acetyl-mannosamine, and a glycerol-phosphate chain with 44 units of glycerol-phosphate, coupled to glucose and D-alanine.

The LTA of *L. acidophilus* consists of a poly-glycerol phosphate chain linked to peptidoglycan by a glycolipid. Information regarding the structure and composition of LTA's of *L. helveticus* was used since data for *L. acidophilus* was not found.

Table 18.: Lipoteichoic composition of *L. acidophilus*.

Precursor	Formula	Stoichiometry (g/g)
sn-Glycerol 3-phosphate	*C*3*H*9*O*6*P*	0.628
D-Alanine	*C*3*H*7*NO*2	0.165
Glycolipid	*C*23*H*36*O*20*R*2	0.153
		Total: 1

In LAB like *L. lactis*, LTA are composed of a disaccharide liked to two fatty acids and a poly-glycerol phosphate chain. However, it was reported in several *Lactobacilli* the presence of a trisaccharide in the LTA structure [184]. In *L. helveticus* the glycolipid anchor is composed of glucose-galactose-glucose-diacylglycerol. The poly-glycerol phosphate chain consists of 24 glycerol-phosphate monomers, coupled to D-alanine, but not to glucose like in the WTAs [172]. The "Glycolipid" compound (D-Glc-(1 $\rightarrow$ 6)-α-D-Gal-(1 $\rightarrow$ 2)-α-D-Glc-(1 $\rightarrow$ 3)-1,2-diacylglycerol) was manually created since it was not found in KEGG, although it is present in ChEBI (CHEBI:63785). The final step of its synthesis is also absent in KEGG, so a reaction representing the transfer of D-glucose from UDP-glucose to α-D-Gal-(1 $\rightarrow$ 2)-α-D-Glc-(1 $\rightarrow$ 3)-1,2-diacylglycerol was created (R_Glycolipid).

Merlin includes automatically universal cofactors in the model, through the "R_e-Cofactor" reaction. A literature review was performed to confirm the inclusion of these

compounds and identify additional cofactors and vitamins, specific for *L. acidophilus*. Each cofactor was considered to be present in the same g/g DW amounts (Table 19).

Table 19.: Cofactor composition of *L. acidophilus*, formula, stoichiometry (g/gDW), and respective source/reference.

Precursor	Source/ Reference	Formula	Stoichiometry (g/g)
CoA	*merlin*	$C21H36N7O16P3S$	0.071
FMN	*merlin*	$C17H21N4O9P$	0.071
Riboflavin	*merlin*	$C17H20N4O6$	0.071
$NADP^+$	*merlin*	$C21H29N7O17P3$	0.071
NAD^+	*merlin*	$C21H28N7O14P2$	0.071
Pyridoxal phosphate	*merlin*	$C8H10NO6P$	0.071
S-Adenosyl-L-methionine	*merlin*	$C15H22N6O5S$	0.071
FAD	*merlin*	$C27H33N9O15P2$	0.071
Tetrahydrofolate	*merlin*	$C19H23N7O6$	0.071
Pantothenate	*merlin*	$C9H17NO5$	0.071
Undecaprenyl diphosphate	[89]	$C9H17NO5$	0.071
Spermidine	[185]	$C7H19N3$	0.071
Putrescine	[185]	$C4H12N2$	0.071
			Total: 1

Cofactors included by *merlin* are universal organic compounds essential for prokaryotic organisms. These compounds are usually included in GSM models but still require manual confirmation. Undecaprenyl diphosphate is necessary for the biosynthesis of peptidoglycan and its inclusion was inferred from the *L. plantarum* WFCS1 model. Spermidine and putrescine are polyamines whose presence was identified in *L. acidophilus* [185]. The presence of ornithine decarboxylase allows the production of putrescine from ornithine, while spermidine is available in the growth medium. Genes encoding lysine decarboxylase and arginine decarboxylase were not found [162], so cadaverine and agmatine are not produced and were not included in the biomass cofactors.

Other cofactors and vitamins were considered, but not included. Based on the enzymes annotation, the genome of *L. acidophilus* appears to not encode thiamine-phosphate synthase, which catalyzes the last reaction of the thiamine synthesis pathway. The absence of this enzyme was also verified in the work of *Altermann et al* [118]. Hence, *Thiamine diphosphate (TPP)* cannot be produced, is not required for growth of this species [186, 187], and is not present in the defined growth medium. Consequently, TPP was not included in the biomass cofactors. Ubiquinone (coenzyme Q10) is associated with respiration processes,

and is not present in *L. acidophilus*. Heme is associated with aerobic organisms, which is not the case of *L. acidophilus*. Glutathione is a protective agent against reactive oxygen species. Although this metabolite is present in the model, the genes associated with its production were not found, and the growth medium does not contain glutathione. Biotin participates in several reactions involving the transfer of carbon dioxide. It is not necessary for the growth of *L. acidophilus* and is not present in growth media. After consulting BRENDA, there was not found any reaction in the model requiring biotin as a cofactor. The same result was obtained for cobalamin, menaquinone, ascorbic acid, and tetrahydrobiopterin.

Information regarding the energetic requirements of *L. acidophilus* was not found. Hence, a comparison of the GAM and NGAM energy requirements among different LAB was performed (Table 20).

Table 20.: GAM and NGAM energy requirements in *GSM* models of LAB and *B. subtillis*.

Organism	GAM ($mmol\ gDW^{-1}$)	NGAM ($mmol\ h^{-1}\ gDW^{-1}$)	Reference
L. plantarum WCFS1	27.40	0.36	[89]
L. lactis MG1363	39.40	0.92	[93]
L. lactis ssp. lactis IL1403	18.15	1.00	[93]
L. casei LC2W	41.15	1.52	[97]
L. acidophilus La-14	27.40	1.48	—

The growth-associated maintenance energy requirement was inferred from data available for *L. plantarum* WCFS1 (27.4 $mmol\ gDW^{-1}$), due to its phylogenetic proximity to *L. acidophilus* (comparing to *L. lactis* and *L. casei*), allowing also an *in silico* growth rate closer to the experimental value.

The NGAM energy requirement was determined by adjusting this parameter to the *in silico* growth rate (Figure 9). The determined value (1.48 $mmol\ h^{-1}\ gDW^{-1}$) is relatively close to the NGAM energy requirements found in other LAB.

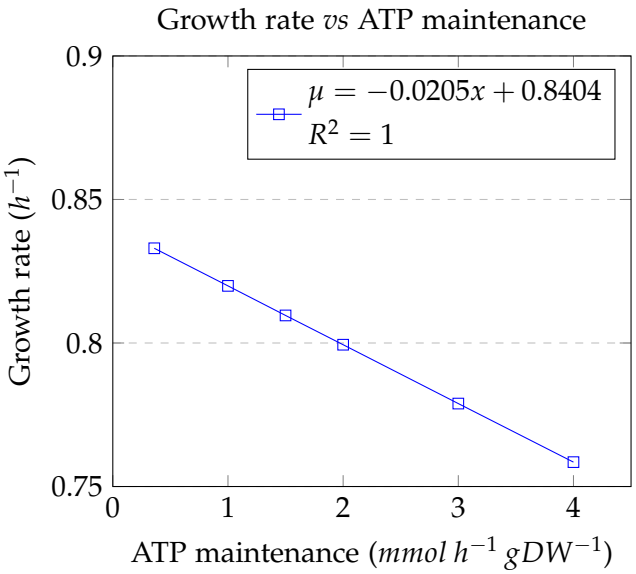


Figure 9.: *In silico* growth rate (μ) *vs* flux of the ATP_maintenance reaction (x). The equation presented in the plot was used to determine the NGAM requirements, using μ = 0.81 h^{-1}.

4.3.2 *Model troubleshooting*

Once the biomass equation was formulated and included in the network, the model was evaluated to assure the production of all the considered biomass precursors. Hence, a minimal medium, containing just the necessary compounds for growth, was defined. The complete medium composition is available in support material Table S3. The lower bounds of the exchange reactions of metabolites included in the medium were settled to -999999.

A plugin developed for *merlin* (BioISO) was used to identify and correct errors blocking the production of biomass precursors. Three main iterations were performed until a positive biomass production was achieved.

Table 21.: Alterations performed through the first iteration of the model troubleshooting stage.

Macromolecule	**Missing metabolites**
DNA	All
RNA	All
Protein	Aminoacyl-tRNAs
Lipid	All
Peptidoglycan	All, except D-Aspartate
EPS	All
WTA	All
LTA	All
Cofactors	CoA, S-adenosyl-L-methionine, FAD, Tetrahydrofolate, NAD^+, $NADP^+$, Undecaprenyl-diphosphate
Type of modification	**Gene/Reaction/Metabolite**
Annotation alterations	LA14_0880, LA14_0955: 2.6.1.- $\rightarrow$ 2.6.1.83; LA14_0690: 1.1.1.- $\rightarrow$ 1.1.1.100; LA14_1120: 1.8.1.4 $\rightarrow$ 1.8.1.7
Added reactions	R00260; R04467; R07613; R00115; R04533, R04953, R04536, R04534, R04964, R04566, R04543, R07763
Reversibility alterations	R03192; R00177; R00430; R01858; R01138

At the beginning of this stage, most of the biomass precursors were not being produced. The model was unable to produce nucleotides in its triphosphate form, inhibiting the majority of biosynthetic pathways. Lipids and LTAs were not being produced since the pathway for the lactobacillic acid biosynthesis was blocked.

Two complete EC numbers were identified at this point: LL-diaminopimelate aminotransferase (2.6.1.83), which participates in the "Lysine biosynthesis" pathway; and 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100), one of the two enzymes associated with the synthesis of fatty acids identified in *L. acidophilus* La-14. The EC 1.8.1.4 was assigned to the gene with the *locus tag* LA14_1120 at the genome annotation stage. After revising the annotation, the EC number of this gene was modified to 1.8.1.7 (glutathione-disulfide reductase).

Table 22.: Alterations performed during the second iteration of the model troubleshooting stage.

Macromolecule	**Missing metabolites**
DNA	dATP, dCTP, dTTP
RNA	UTP, GTP, CTP
Protein	Aminoacyl-tRNAs
Lipid	All
Peptidoglycan	L-lysine, UDP-N-acetylmuramate, UDP-N-acetyl-α-D-glucosamine
EPS	All
WTA	UDP-glucose, UDP-N-acetyl-glucosamine, UDP-N-acetyl-mannosamine
LTA	Glycolipid
Cofactors	CoA, NAD^+, $NADP^+$, Undecaprenyl diphosphate, S-adenosyl-methionine, tetrahydrofolate
Type of modification	**Gene/Reaction/Metabolite**
Annotation alterations	LA14_1420: 3.5.1.- $\rightarrow$ 3.5.1.19
Added reactions	R01268, R00156, R00330, R00570, R00722, R02093
Reversibility alterations	R00940, R00937, R01072, R02291, R04467, R03905

The changes performed in the first iteration allowed the model to obtain ATP. However, the synthesis of several precursors containing UDP, like UDP-N-Acetylglucosamine, require UTP. Similarly to what was found in *L. lactis* [188], the genome of *L. acidophilus* La-14 does not encode nucleoside-diphosphate kinase, an enzyme that transfers phosphate to nucleosides-diphosphate, originating nucleosides triphosphate. In *L. lactis* and a mutant *E. coli*, it was reported that adenylate kinase and pyruvate kinase are responsible for the synthesis of nucleosides triphosphate [188, 189]. Hence, reactions catalyzed by nucleoside-diphosphate kinase were included but associated with the adenylate kinase and pyruvate kinase enzymes. The completion of the EC number 3.5.1.- to 3.5.1.19, associated with nicotinamidase, allowed the synthesis of NAD^+. Reactions included here, together with the changes in the reversibility of six reactions, allowed the model to produce most of the biomass precursors (Table 23).

Table 23.: Alterations performed during the third iteration of the troubleshooting stage.

Macromolecule	**Missing metabolites**
DNA	—
RNA	—
Protein	Aminoacyl-tRNAs
Lipid	All
Peptidoglycan	—
EPS	—
WTA	—
LTA	Glycolipid
Cofactors	—
Type of modification	**Gene/Reaction/Metabolite**
Annotation alterations	LA14_0344: 6.1.1.17 → 6.1.1.24; LA14_0559: 6.3.5.- → 6.3.5.6; LA14_o887: 3.2.2.- → 3.2.2.1
Added reactions	R03411, R_autoinducer2, T_autoinducer2, R03651; R01290, R01245, R012273, R01677, R01770, R02143, R02341
Reversibility changes	R00190

Two EC numbers were updated in this iteration: aspartyl-tRNA synthase (EC 6.3.5.6), and purine nucleosidase (EC 3.2.2.1). The EC number 6.1.1.17, associated with the gene with the *locus tag* LA14_0344, was replaced by 6.1.1.24. The specificity of the enzyme associated with 6.1.1.17 is restricted to tRNA(Glu). On the other hand, the low stereospecificity of the glutamate-tRNA ligase (EC 6.1.1.24) allows the formation of both glutamyl-tRNA(Glu) and glutamyl-tRNA(Gln). In *B. subtilis*, the gene *gltX* (*locus tag* BSU00920) was reported to act on both tRNA(Glu) and tRNA(Gln) [190]. This gene is available in the BLAST results of the LA14_0344 gene (with an e-value of 0.0). This alteration was essential since glutamyl-tRNA(Gln) is not produced by any other reaction in the "Aminoacyl-tRNA biosynthesis" pathway.

As mentioned before, *L. acidophilus* can obtain lactobacillic acid from octadecenoic acid (R03411). This reaction requires S-adenosylmethionine as a methyl donor and has S-adenosylhomocysteine as a by-product. The *Autoinducer (AI)*-2 is produced at the end of this pathway (figure 10). AI-2 is a quorum-sensing signaling molecule produced by several gram-positive and gram-negative bacteria [191]. The production of this molecule was observed in *L. acidophilus* and was reported to play a role in the adherence of this species to the intestinal epithelium [192]. The non-enzymatic reaction representing the conversion of 4,5-dihydroxy-2,3-pentanedione to AI-2 (R_Autoinducer2), and the transport reaction for this compound (T_Autoinducer2) were added in this stage. These reactions

were created manually, according to available literature [192]. The transport reaction was associated with the specific transporter with the TC number 2.A.86.

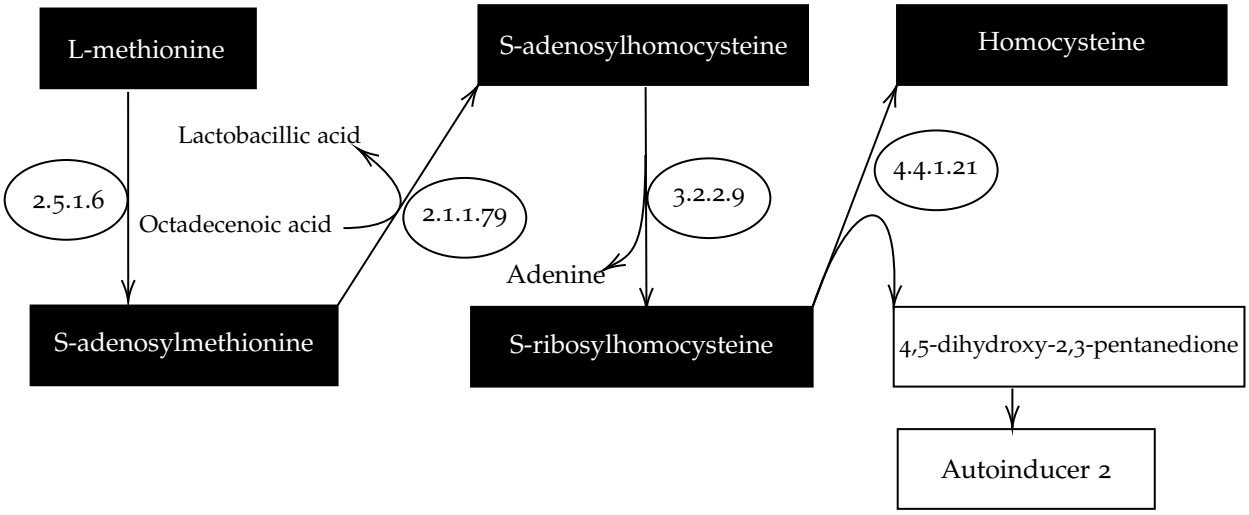


Figure 10.: Pathway for the production of Autoinducer-2 from methionine in *L. acidophilus*. The EC numbers have the following correspondence: 2.5.1.6 (methionine adenosyltransferase), 2.1.1.79 (S-adenosylmethionine methyltransferase), 3.2.2.9 (adenosylhomocysteine nucleosidase), 4.4.1.21 (S-ribosylhomocysteine lyase). The circularization of 4,5-dihydroxy-2,3-pentanedione into AI-2 is non-enzymatic. The AI-2 is secreted through the specific transporter with the TC number 2.A.86.

4.3.3 *Gap filling*

After obtaining a model able to produce all biomass precursors, the model was gap-filled. At this stage, blocked reactions and dead-end metabolites were identified and blocked reactions corrected. The reversibility/direction of 27 reactions was changed, and 21 reactions were manually added (Table 24). The gap-filling process allowed to decrease the number of blocked reactions to 185 and the number of dead-ends to 177 metabolites. The "Fatty acid biosynthesis" pathway includes several reactions associated with the *fabG* and *fabT* genes. Since all other genes regarding the synthesis of fatty acids are missing in the genome of this strain, almost all reactions associated with this pathway are unconnected. A similar situation can be found in the "Drug metabolism - other enzymes" pathway. Additionally, the synthesis of peptidoglycan is accounted for in the "Biomass pathway". Hence, the KEGG's "Peptidoglycan biosynthesis" pathway does not have an end product, turning reactions associated with this pathway directly or indirectly unconnected.

Table 24.: Alterations performed during the gap-filling stage.

Reaction ID	**Alteration**	**EC number**	Gene
R02142, R01665, R00137, R03601, R04859	To reversible	—	—
R01130, R01231, R10147, R04198, R04199, R03192, R00416, R01248, R01251, R03291, R03293, R01773, R01775, R00842, R00844, R03591, R03592, R04198, R04199, R01520, R01521	To irreversible	—	—
R03104	Backward direction	—	—
R00704	Manually added	1.1.1.28	LA14_0055
R00207	Manually added	1.2.3.3	LA14_1962
R01876	Manually added	2.4.2.3	LA14_1080, LA14_1078, LA14_1081
R01978	Manually added	2.3.3.10	LA14_0657
R01663	Manually added	3.5.4.12	LA14_0206
R05627	Manually added	3.6.1.27	LA14_1019
R01658	Manually added	2.5.1.1	LA14_1328
R02061	Manually added	2.5.1.29	LA14_1328
R06447	Manually added	2.5.1.31	LA14_1269
R03122	Manually added	3.2.1.54	LA14_1956
R01600, R01786	Manually added	2.7.1.2	LA14_0908
R00751	Manually added	4.1.2.5	LA14_0254
R03314, R04861	Manually added	non-enzymatic	—
R02806	Substrate specification	2.4.2.6	LA14_0146, LA14_1629
R07171	Manually added	1.6.3.3	LA14_1416, LA14_1419

D-Lactate dehydrogenase was not included during the genome annotation. Since *L. acidophilus* produces both L and D-lactate, the annotation was reviewed and the EC number 1.1.1.28 was assigned to the gene with the *locus tag* LA14_0055.

The presence of pyruvate oxidase has been described in *Lactobacilli* [193, 194, 195]. The EC number 2.2.1.6 (acetolactate synthase) was first assigned to the gene with the *locus tag* LA14_1962. Despite the activity of this enzyme is reported in several LAB, it seems to be absent in *L. acidophilus* [196]. After an annotation review, this gene was reannotated as encoding pyruvate oxidase, whose presence in *L. acidophilus* was confirmed in literature [193].

Glucokinase (EC 2.7.1.2) was not included through the genome annotation stage. Its absence implies that intracellular glucose cannot be phosphorylated, and therefore cannot enter glycolysis. After checking KEGG, this enzyme was included in the model, associated with the gene with the *locus tag* LA14_0908.

The activity of a threonine aldolase was demonstrated in *L. acidophilus* [197]. The enzyme associated with this activity appears to have both threonine aldolase (EC 4.1.2.5) and glycine hydroxymethyltransferase (EC 2.1.2.1) activities, as reported in *S. thermophilus* [198] and *L. lactis* [93]. Hence, after revising the annotation, the gene with the *locus tag* LA14_0254, was associated with both activities.

The enzyme nucleoside deoxyribosyltransferase (EC 2.4.2.6) is associated with the generic reaction R02806, catalyzing the transfer of the deoxyribosyl group from one purine or pyrimidine to another. The substrate specificity of this enzyme was reported to be very low, presenting activity with most of purine and pyrimidine bases in *L. lactis* [199]. Nevertheless, *Soska et al* [200] have reported that the *in vivo* activity of this enzyme in *L. acidophilus* is associated with the conversion of deoxyadenosine to adenine:

$$Deoxyadenosine + Thymine \rightleftharpoons Adenine + Thymidine$$

Generic metabolites "Deoxynucleoside" and "Base" included in the reaction R02806 were replaced by deoxyadenosine and adenine, respectively.

The presence of an NADH oxidase was not detected through the genome annotation stage. Nevertheless, the presence of this enzyme was reported in *L. acidophilus*, with hydrogen peroxide production [201, 202, 203, 204]. *Altermann et al* [118], and *Pridmore et al* [205] associated NADH oxidase with two genes in *L. acidophilus* NCFM and *L. johnsonii* NCC 533. The homologous genes available in *L. acidophilus* La-14 were identified (LA14_1416 and LA14_1419) through BLAST searches against the genome of this strain, and were associated with the NADH oxidase enzyme (H_2O_2 forming).

4.4 MODEL VALIDATION

Various tests were applied to validate the metabolic model: spontaneous growth, growth rate assessment, auxotrophies, alternative elements in the medium, and determi-

nation of critical genes. The phenotype simulation method used was pFBA, except when otherwise indicated.

4.4.1 *Growth rate assessment and in silico simulations*

First of all, the lower bound of all drains was set to zero to test spontaneous growth. As expected, no growth was observed.

The growth rate was accessed according to the study of *Lv et al* [206], which used the *L. acidophilus* KLDS strain. The glucose consumption was measured in this work (16.4 $mmol\ h^{-1}\ gDW^{-1}$), allowing to define this value as the *in silico* uptake rate for carbon sources. As mentioned before, the *in silico* growth rate was adjusted to the experimental value (0.81 h^{-1}) by amending the NGAM energy requirements. *Soska et al* [200] obtained an identical growth rate (0.83 h^{-1}), using a similar CDM (with the addition of L-asparagine), though not measuring, the glucose consumption rate. In *L. lactis*, a growth rate of 0.79 h^{-1} was measured anaerobically with a glucose consumption rate of 13.7 $mmol\ h^{-1}\ gDW^{-1}$ [94].

The consumption rate of metabolites included in the minimal CDM for *in silico* growth of *L. acidophilus* La-14 is indicated in Table 25. A growth rate of 0.72 h^{-1} was determined using this medium.

Table 25.: *In silico* consumption rates of metabolites present in the minimal medium. The simulation was performed using pFBA in anaerobic conditions.

Metabolite	**Consumption rate** ($mmol\ h^{-1}\ gDW^{-1}$)	**Metabolite**	**Consumption rate** ($mmol\ h^{-1}\ gDW^{-1}$)
α-D-Glucose	16.4	L-Proline	0.067
L-Aspartate	1.861	Hexadecanoic acid	0.055
Orthophosphate	1.559	L-Histidine	0.038
L-Serine	1.409	Spermidine	0.022
Acetate	0.404	Riboflavin	0.020
L-Glutamate	0.264	L-Tryptophan	0.020
L-Leucine	0.170	Pantothenate	0.019
L-Isoleucine	0.147	Pyridoxal	0.013
L-Valine	0.127	Hexadecenoic acid	0.012
L-Arginine	0.106	Nicotinamide	0.009
L-Phenylalanine	0.081	H+	0.009
L-Methionine	0.074	Folate	0.007
L-Tyrosine	0.073	Octadecanoic acid	0.004
Octadecenoic acid	0.067	Tetradecanoic acid	0.004

Acetate is required for both *in silico* and *in vivo* [200] growth of *L. acidophilus*. This requirement will be discussed in Section 4.6.2.

The requirements of this species for amino acids, fatty acids, and vitamins are demonstrated in the *in silico* minimal CDM: six vitamins, 13 amino acids, and fatty acids were identified as essential. On the other hand, *in silico* growth could be achieved without pyrimidine and purine nucleotides.

Compounds produced in minimal and rich media and respective production rates are shown in Table 26.

Table 26.: *In silico* secreted compounds, and respective production rates in minimal and rich media. The simulation was performed using pFBA in anaerobic conditions.

Metabolite	Production rate ($mmol\ h^{-1}\ gDW^{-1}$)	
	Minimal media	Rich media
L/D-Lactic acid	30.340	29.862
H_2O	9.794	9.016
CO_2	2.265	0.737
Succinate	0.112	0.000
Urea	0.037	0.041
H_2S	0.015	0.017
Hydroxybutanoic acid	0.001	0.017
Autoinducer-2	0.015	0.017
HCO_3^-	0.000	0.002
Malate	0.001	0.000

L. acidophilus can produce both L and D-lactic acid, usually in similar amounts [187]. Since pFBA minimizes the enzyme-associated flux, only one of the isomers is produced. The ratio between the lactic acid production and glucose consumption is in agreement with literature (1.8 $mol_{lactate}/mol_{glucose}$) [207].

The CO_2 production is associated with the activity of phosphogluconate dehydrogenase, which converts 6-phospho-gluconate to ribose 5-phosphate, releasing CO_2 in the presence of $NADP^+$. Diphosphomevalonate decarboxylase, L-ornithine carboxy-lyase, and phosphopantothenoylcysteine decarboxylase are also responsible for the production of CO_2, although in lower amounts.

Urea is produced together with ornithine from the hydrolysis of arginine. Since there are no reactions using urea as a reactant, this compound is secreted into the extracellular space.

The degradation of L-homocysteine by cystathionine-γ-Lyase originates hydrogen sulfide (which is secreted), ammonia, and 2-oxobutanoate (which is reduced to hydrox-

ybutanoic acid). Hydrogen sulfide can also be obtained from L-cysteine, in a reaction catalyzed by the same enzyme. The production of this sulfurous compound was reported in *L. lactis* [208], playing an important role in the flavor of cheddar cheese [209]. The work of *Sreekumar et al* [210] has demonstrated that *L. acidophilus* NCFM produces methanethiol (MeSH) but not H_2S. The production of MeSH is usually attributed to L-methionine-γ-lyase, which was not found in the genome of the La-14 and NCFM strains. Cystathionine-β-lyase can also convert methionine to MeSH, although less efficiently [211, 212]. This activity of cystathionine-β-lyase is available in the metabolic model.

The production of poly-β-hydroxybutyrate (polymer of hydroxybutanoic acid) was already reported in several LAB, including *L. acidophilus* [213, 214]. However, the accumulation of this compound is usually achieved when the carbon availability is excessive, and the growth is limited by the availability of nitrogen and phosphorus sources [214].

In minimal media, the production of succinate and malate was detected as a result of the purine and pyrimidine metabolism.

L. acidophilus achieves optimal growth in anaerobic and microaerophilic conditions, even though most strains are aerotolerant. The lower growth rate in aerobic conditions has been associated with the damage caused by reactive oxygen species to biomolecules like proteins and lipids [215]. Some LAB, such as *L. plantarum* and *L. casei*, produce acetate in aerobic conditions through the pyruvate oxidase pathway:

$$Pyruvate + O_2 + H_3PO_4 \rightarrow Acetylphosphate + H_2O_2 + CO_2$$

$$Acetylphosphate + ADP \rightleftharpoons Acetate + ATP$$

The rerouting of pyruvate through this pathway allows the production of additional ATP, while NAD^+ is regenerated either by NADH oxidase and NADH peroxidase. *In silico* simulations in aerobic conditions predict the production of acetate through the POX pathway and a higher growth rate. Nevertheless, *L. acidophilus* is obligatory homofermentative and does not produce acetate, even if oxygen is available [203, 216]. To access the *in silico* requirement for acetate production, FVAs were performed in anaerobic and aerobic conditions (Table 27), constraining the maximum growth rate to 0.81 h^{-1}.

Table 27.: Minimum and maximum fluxes of acetate, lactate, and ethanol, determined through FVA simulations in aerobic (oxygen uptake unconstrained) and anaerobic conditions. In both conditions, the growth rate was limited to 0.81 h^{-1}.

	Anaerobic conditions		Aerobic conditions	
Compound	Minimum flux	Maximum flux	Minimum flux	Maximum flux
Acetate	-0.27	-0.26	0	47.68
Lactate	29.88	29.89	0	29.69
Ethanol	0	0.001	0	4.98

In anaerobic conditions, acetate is consumed while lactate production is required, which is in good agreement with available information. In these conditions, lactate (originated from pyruvate) is the only by-product, and the lactate dehydrogenase activity assures the regeneration of NAD^+. Ethanol can also be produced from acetaldehyde in residual amounts, as a result of the threonine aldolase and alcohol dehydrogenase activities.

In aerobic conditions, the production of lactate, acetate or ethanol is not required to achieve the defined growth rate. No information was found explaining why *L. acidophilus* does not produce acetate in these conditions. In *L. johnsonii*, the flux through pyruvate oxidase is limited but sufficient to eliminate the requirement for acetate, without producing this compound [194]. Hence, POX may be used by *L. acidophilus* and *L. johnsonii* just to provide acetyl-CoA, and not for acetate production (Figure 11).

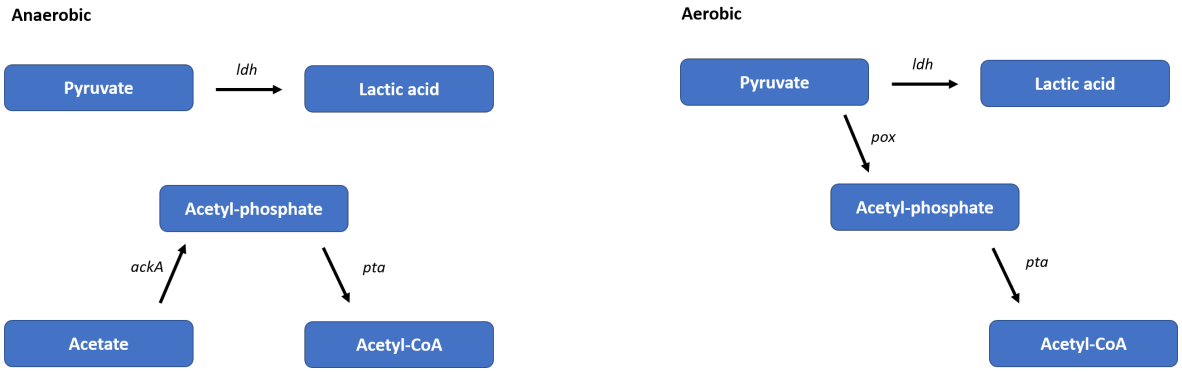


Figure 11.: Pyruvate metabolism in anaerobic (left) and aerobic (right) conditions. In anaerobic conditions pyruvate is converted to lactic acid by lactate dehydrogenase (*ldh*), while acetate forms acetyl phosphate by acetate kinase (*ackA*). Acetyl phosphate is then converted to acetyl-CoA by phosphate acetyltransferase (*pta*). In aerobic environment, pyruvate can be used as a source of acetyl phosphate due to the activity of pyruvate oxidase (*pox*).

Nevertheless, this hypothesis requires experimental confirmation in *L. acidophilus*. In fact, more quantitative experimental data for this species, preferably for the La-14 strain, in anaerobic and aerobic conditions, would be useful to increase the reliability of the metabolic

model. Experimental procedures available in literature often use complex media instead of CDM, without measuring the consumption and production rate of substrates and by-products, respectively.

4.4.2 *Fermentation pattern*

L. acidophilus can grow using hexoses and disaccharides as carbon source. The fermentation pattern of *L. acidophilus* with different carbohydrates [217, 218] was compared with *in silico* simulations using the model developed in this work, and the AGORA model for *L. acidophilus* NCFM. The genomes of the La-14 and NCFM strains are very similar (as shown in subsubsection 2.3.2.2), thus no major differences in the metabolism of these two strains are expected. Table 28 presents a confusion matrix evaluating the performance of the two metabolic models. A more detailed analysis is available in **Supplementary Table S4**.

Table 28.: Confusion matrix and respective performance measure calculations of the model in presenting growth with different carbohydrates.

	This work		AGORA model	
	Exp. positive	Exp. negative	Exp. positive	Exp. negative
Predicted Positive	14	1	12	1
Predicted Negative	1	5	3	5
Measure	**Value**		**Value**	
Sensitivity	0.93		0.80	
Specificity	0.83		0.83	
Precision	0.93		0.92	
NPV	0.83		0.63	
Accuracy	0.91		0.81	

Growth with a given carbohydrate as carbon source was considered as positive. Expected outcomes were defined according to experimental data. Predicted (P.) outcomes were determined with pFBA.

This metabolic model presents a good performance predicting growth with different carbohydrates. In general, the values of the performance measures are higher than the respective ones in the AGORA model.

L. acidophilus grows with the most frequently used carbohydrates (glucose, galactose, sucrose, lactose, and fructose). *In silico* growth matching experimental data was also observed with cellobiose, N-acetyl-glucosamine, mannose, and salicin. Although growth with stachyose was not determined in the considered studies, *in silico* growth was observed, which is in agreement with the work of *Stern et al* [219]. Growth with melibiose, trehalose, maltose and, raffinose depends on the strain in question. *In silico* simulations allow the growth using these sugars, due to the lack of knowledge regarding the specificity of glu-

cosidases and galactosidases of the *L. acidophilus* La-14 strain. In opposition to experimental data, no growth was observed using amygdalin (a cyanogenic glycoside), since this metabolite is not even present in the metabolic model.

Srinivas et al [220] have defined a growth order for *L. acidophilus* according to the carbon source: $glucose > fructose > sucrose > lactose > galactose$. This result was confirmed by other studies [219, 221, 222]. The comparison of this information with *in silico* simulations can only be performed qualitatively, as complex media were used in these experiments and information regarding the carbohydrate uptake is missing. Figure 12 shows the *in silico* growth rate, using the mentioned sugars.

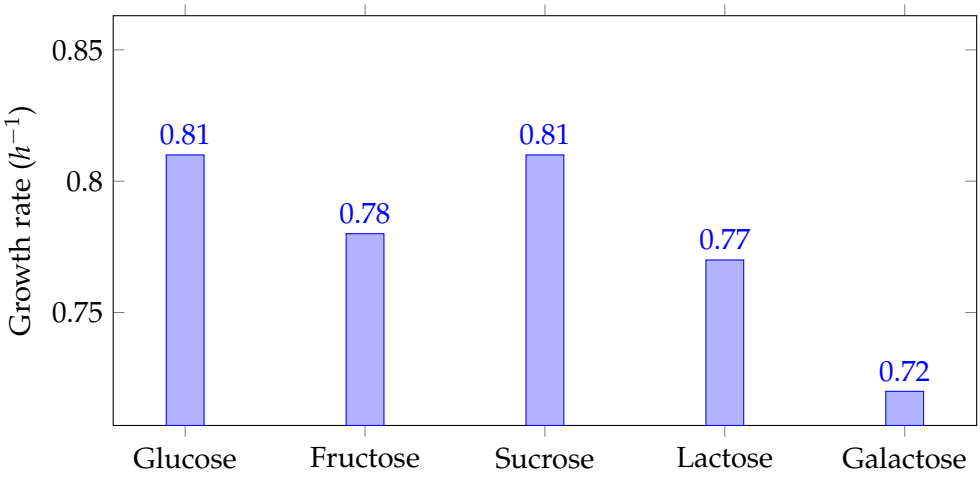


Figure 12.: *In silico* growth rate, using different carbon sources. The lower bound of each carbohydrate exchange reaction was accounted for the number of carbon atoms: glucose, fructose, galactose (16.4 $mmol\ h^{-1}\ gDW^{-1}$), lactose, sucrose (8.2 $mmol\ h^{-1}\ gDW^{-1}$).

As expected, the utilization of glucose allows a higher growth rate compared with the other carbohydrates. Growth using fructose and lactose is slightly lower, and galactose originated the lowest growth rate. These differences are related to the transport system associated with the uptake of each sugar, and the pathway followed for its degradation. Moreover, *in vivo* differences are associated with factors not accounted for in the metabolic model, namely gene expression and quantitative enzymatic activity. For example, genes associated with the uptake of glucose and fructose through the PTS are consistently expressed, regardless of the carbohydrate source, explaining the preference of *L. acidophilus* for these sugars [161]. On the other hand, genes associated with the metabolism of lactose and galactose need to be induced by the presence of these carbohydrates [161].

4.4.3 *Amino acid requirements*

Growth media for *L. acidophilus* usually contain non-essential amino acids to stimulate growth. Table 29 presents the comparison between available information and *in silico* simulations regarding the amino acid requirements of *L. acidophilus*.

Table 29.: Amino acid requirements of *L. acidophilus* according to *Morishita et al* [123] (using *L. acidophilus* ATCC11506) and *in silico* simulations.

Amino acid	*Morishita et al* [123]	*L. acidophilus* La-14 (This work)	*L. acidophilus* NCFM (AGORA model)
Alanine	NE	NE	E
Arginine	E	E	E
Asparagine	NE	NE	NE
Aspartate	E	E	NE
Cysteine	**E**	**NE**	**E**
Glutamate	E	E	E
Glutamine	NE	NE	NE
Glycine	NE	NE	NE
Histidine	E	E	E
Isoleucine	E	E	E
Leucine	E	E	E
Lysine	NE	NE	NE
Methionine	E	E	E
Phenylalanine	E	E	E
Proline	E	E	E
Serine	E	E	E
Threonine	**E**	**NE**	**NE**
Tryptophan	E	E	E
Tyrosine	E	E	E
Valine	E	E	E

The model requirements for amino acids match the available information, except for cysteine and threonine. L-cysteine can be obtained from cystathionine by cystathionine-γ-lyase. *In silico* analysis of the *L. acidophilus* NCFM genome has also indicated the potential to produce L-cysteine [118].

The reversible activity of threonine aldolase allows the production of threonine from glycine and acetaldehyde in both metabolic models. Alternatively, threonine can be produced from phospho-L-homoserine by threonine synthase. The potential to produce thre-

onine was also reported in *L. acidophilus* NCFM [118], thus the differences between experimental data and *in silico* simulations might be related with the utilization of different strains.

4.4.4 *Gene essentiality and single gene deletion*

Critical genes were determined using Optflux, under different environmental conditions (aerobiosis/anaerobiosis; minimal/rich medium). No information regarding gene essentiality of *L. acidophilus* was found. Table 30 presents the number of critical genes and critical reactions identified in the model.

Table 30.: Number of critical genes and critical reactions identified in the metabolic model, under different environmental conditions.

Environmental Conditions	Critical genes	Critical reactions
Anaerobic		
Minimal medium	133	155
Rich Medium	104	123
Aerobic		
Minimal medium	133	156
Rich Medium	104	122

The number of critical genes and reactions in aerobic and anaerobic conditions is similar. In the presence of oxygen, LA14_1398 (NADH peroxidase) becomes essential to regenerate NAD^+ and eliminate H_2O_2. On the other hand, lactate permease is not essential in aerobic conditions since POX provides an alternative fate for pyruvate. *L. acidophilus* presents both *ldhL* and *ldhD*, thus the knockout of one of these genes will not affect *in silico* growth. Nevertheless, the lactate permease and lactate dehydrogenase may be essential for *in vivo* growth of *L. acidophilus* since lactic acid is the fermentation final product in both anaerobic and aerobic conditions.

In rich media, critical genes are mainly associated with the synthesis of peptidoglycan precursors (”Amino sugar and nucleotide sugar metabolism” pathway), phospholipids (”Glycerolipid metabolism” and ”Glycerophospholipid metabolism” pathways), and cofactors (”Nicotinate and nicotinamide metabolism”, ”Pantothenate and CoA biosynthesis”, ”Terpenoid backbone biosynthesis”, and ”Riboflavin metabolism” pathways). Six genes associated with glycolysis were identified as critical, as well as 19 genes responsible for the aminoacyl-tRNA biosynthesis (gln-tRNA ligase is associated with the synthesis of both glutamyl-tRNA(Glu) and glutamyl-tRNA(Gln). The *eps* gene cluster and genes associated with the assembly of WTA and LTA ware also identified as necessary. Furthermore, the

genes with the *locus tag* LA14_0942 and LA14_0315 were also identified as critical, and are associated with the transport of several cofactors, namely riboflavin, folate, and pyridoxal.

In minimal media under both aerobic and anaerobic conditions, 29 additional genes were identified as critical. These genes are associated with the synthesis of non-essential amino acids, such as lysine and alanine, and the *de novo* synthesis of purine and pyrimidine nucleotides.

Additionally, the number of critical reactions is always higher than the number of critical genes due to the existence of reactions not associated with genes. The availability of datasets containing information on gene essentiality of this species would be useful to test the phenotypic prediction capability of the metabolic model.

A search for experimental data regarding single gene deletions in *L. acidophilus* was performed and compared with simulations. The information available in literature relies mainly on knockouts associated with the carbohydrate uptake and metabolism of *L. acidophilus* NCFM. Table 31 shows a comparison between the expected phenotype (according to available information) and *in silico* simulations results.

Table 31.: Comparison between the expected phenotype (according to available information) and results of *in silico* simulations. All knockouts were performed using LMOMA. The deleted genes have the following *locus tag* correspondence: Δ*lacS*, (LA14_1458); Δ*galA*, (LA14_1437), Δ*msmE*, (LA14_1441);Δ*frc*, (LA14_0391); Δ*treB*, (LA14_1026); Δ*treC*, (LA14_1028).

Gene deleted	**Expected phenotype**	***In silico* simulation**	**Reference**
Δ*lacS*	Non-growth on lactose	Non-growth on lactose	[223]
Δ*galA*	Non-growth on raffinose, melibiose, stachyose	Non-growth on raffinose, melibiose, stachyose	[224]
Δ*msmE*	Growth on galactose, Non-growth on raffinose	Growth on galactose, Non-growth on raffinose	[224]
Δ*treB*	Non-growth on trehalose	Non-growth on trehalose	[225]
Δ*treC*	Non-growth on trehalose	Growth on trehalose	[225]
Δ*frc*	Lost of oxalate degrading capability	Lost of oxalate degrading capability	[157]

The gene *lacS*, encoding a lactose/galactose permease, is essential for the growth on lactose, since it is the only transport mechanism for this sugar found in *L. acidophilus* [223]. The inactivation of α-galactosidase through the knockout of the *galA* gene does not allow the growth on different carbohydrates (raffinose, melibiose, and stachyose). The deletion

of the *msmE* gene, associated with the raffinose uptake through the ABC-binding cassette system, blocks the growth with this sugar.

Trehalose enters the cell through the PTS, so the knockout of *treB* is fatal for *L. acidophilus* when using this sugar as a carbon source. The *in silico* knockout of *treC*, encoding a trehalose 6-phosphate hydrolase, does not affect growth, contradicting the expected phenotype. Maltose 6-phosphate glucosidase hydrolyzes several 6-phospho-α-D-glucosides, including trehalose 6-phosphate, becoming an alternative to trehalose 6-phosphate hydrolase. No information regarding the specificity of this enzyme in *L. acidophilus* was found, although in *B. subtilis* trehalose 6-phosphate can be used as substrate [226].

L. acidophilus La-14 presents a high ability for the degradation of oxalate, a compound associated with the prevalence of kidney stones [156]. This activity is possible due to the presence of a formyl-CoA transferase and an oxalyl-CoA decarboxylase, encoded by the *frc* and *oxc* genes, respectively. The knockout of the *frc* gene blocks the oxalate degrading capability of *L. acidophilus* [157], as observed in *in silico* simulations.

Generally, results from simulations correspond to the expected phenotypes. However, the information found in literature is mainly associated with carbohydrate uptake and metabolism, thus more experimental data regarding gene knockouts associated with other pathways, such as nucleotide metabolism, should be useful to increase the phenotypic prediction capability of the GSM model.

4.5 MODEL SUMMARY

As mentioned in subsection 2.2.8, *GSM* models for six LAB species are available at the moment, in which *L. plantarum*, *L. casei*, and *L. lactis* are the most closely related species to *L. acidophilus*. An overview of the available metabolic models for these species is presented in Table 32.

Table 32.: Overview of the *GSM* models of four lactic acid bacteria.

	L. acidophilus La-14	*L. acidophilus* NCFM	*L. plantarum* WCFS1	*L. lactis* MG1363	*L. casei* LC2W
Genes	527	540	721	518	846
Gene coverage	28.1 %	29.0 %	23.5 %	19.9 %	27.7 %
Total Reactions	952	1460	762	754	969
Internal reactions	575	923	413	530	604
Transport reactions	213	132	118	119	227
Exchange reactions	164	405	113	105	139
Metabolites	802	1120	658	650	785
Internal	604	715	549	551	—*
External	164	405	113	105	—*
Unique metabolites	608	1009	554	552	604
Compartments	(c,e)	(c,e)	(c,e)	(c,e)	(c,p,e)

* data not available.

c: cytoplasm; e: extracellular space; p: periplasm.

The *GSM* model contains 527 genes, which corresponds to 28.1% of the total genes in the genome of *L. acidophilus* La-14. This percentage is similar to the one found in *L. casei* and in the *L. acidophilus* NCFM AGORA model but is slightly higher than in *L. plantarum* and *L. lactis*.

In general, the number of genes, reactions, and metabolites in the metabolic model is higher than the respective ones in the *L. plantarum* and *L. lactis* models. This may be associated with a more restrictive reconstruction approach in these metabolic models. Similar results are found in the *L. casei* LC2W model. In this model, an additional compartment (periplasm) was included, which is not usual in *GSM* models of gram-positive bacteria. The number of internal and exchange reactions available in the AGORA model is surprisingly high, which might be a result of the semi-automatic reconstruction and lack of manual curation. In fact, this model includes the fatty acid biosynthetic pathway regardless of the absence of the genes required for this pathway. Exchange reactions for several glycans and mineral ions are also available in this model.

4.6 METABOLISM OVERVIEW

4.6.1 *Carbohydrate metabolism*

The pathways used by *L. acidophilus* for the degradation of carbohydrates (glucose, fructose, galactose, lactose, and sucrose) were analyzed accounting *in silico* simulations and available information. Figure 13 shows a reconstruction of these pathways.

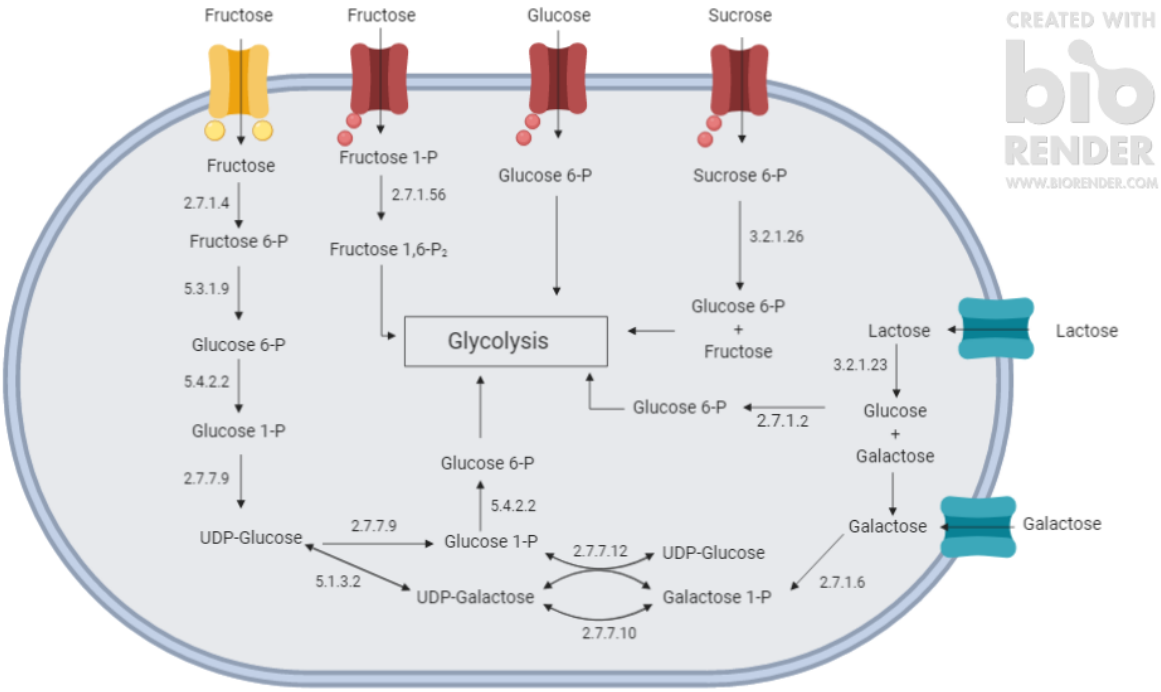


Figure 13.: Pathway reconstruction of the metabolism of sucrose, fructose, glucose, lactose, and galactose in *L. acidophilus*, based on the metabolic model and *in silico* simulations. The uptake of sucrose, fructose, and glucose is made through the PTS (red). An ABC system for the uptake of fructose is presented in yellow. Galactose and lactose enter the cell through a permease (blue). The shown EC numbers have the following correspondence: 2.7.1.2 – glucokinase; 2.7.1.4 – fructokinase; 2.7.1.6 – galactokinase; 2.7.1.56 – 1-phosphofructokinase; 2.7.7.9 – UTP-glucose-1-phosphate uridylyltransferase; 2.7.7.10 – UTP-hexose-1-phosphate uridylyltransferase; 2.7.7.12 – UTP-hexose-1-phosphate uridylyltransferase; 3.2.1.23 – β-galactosidase; 3.2.1.26 – β-fructofuranosidase; 5.1.3.2 – UDP-glucose 4-epimerase; 5.3.1.9 – glucose-6-phosphate isomerase; 5.4.2.2 – phosphoglucomutase. Created with BioRender.com.

The glucose uptake through the PTS allows this sugar to enter the cell as glucose 6-phosphate, which is mainly directed to glycolysis. However, a minor amount is converted to UDP-glucose, required for the synthesis of EPS and WTA.

The sucrose uptake is also performed through the PTS. The β-fructofuranosidase activity allows the degradation of sucrose 6-phosphate to glucose 6-phosphate and fructose. Once again, most of the glucose 6-phosphate follows to glycolysis, and the remaining is con-

verted to UDP-glucose. Fructose obtained from the sucrose hydrolysis is phosphorylated to fructose 6-phosphate by fructokinase.

The PTS is the main transport system responsible for the uptake of exogenous fructose, forming fructose 1-phosphate, which is phosphorylated to fructose 1,6-bisphosphate. In *L. lactis*, part of the fructose 1,6-bisphosphate is hydrolyzed to fructose 6-phosphate, which is used for the formation of biomass precursors (UDP-glucose) [227]. However, fructose-bisphosphatase appears to be absent in the genome of *L. acidophilus*. Moreover, there was not found any enzyme able to convert fructose 1-phosphate to fructose 6-phosphate. Hence, fructose required for biomass may enter the cell through a mechanism other than the PTS. In *in silico* simulations, fructose required for biomass production enters the cell through an ABC system and is then phosphorylated to fructose 6-phosphate, which is converted to glucose 6-phosphate. Phosphoglucomutase converts glucose 6-phosphate to glucose 1-phosphate. Then, an UTP-glucose-1-phosphate uridylyltransferase allows the formation of UDP-glucose from glucose 1-phosphate and UTP.

The uptake of both lactose and galactose is made through symport with H^+. Lactose is degraded into glucose and galactose by a β-galactosidase. Glucose enters in the EMB pathway, while galactose follows the Leloir pathway. In this pathway, galactose is phosphorylated to galactose 1-phosphate, which reacts with UDP-glucose forming UDP-galactose and glucose 1-phosphate. UDP-galactose is then converted to UDP-glucose, and glucose 1-phosphate follows to glycolysis. The tagatose pathway, found in other LAB [228], is absent in *L. acidophilus* [161].

4.6.2 *Pyruvate metabolism*

As mentioned before, *L. acidophilus* ferments hexoses through the EMB pathway. Although the main fate of pyruvate is the production of lactic acid, heterofermentative LAB can also produce ethanol and acetate. In addition, several LAB can convert pyruvate to L-alanine or acetolactate (two compounds associated with flavor), in the presence of alanine dehydrogenase and acetolactate synthase [229], respectively. Figure 14 shows a representation of the pyruvate metabolism in *L. acidophilus*.

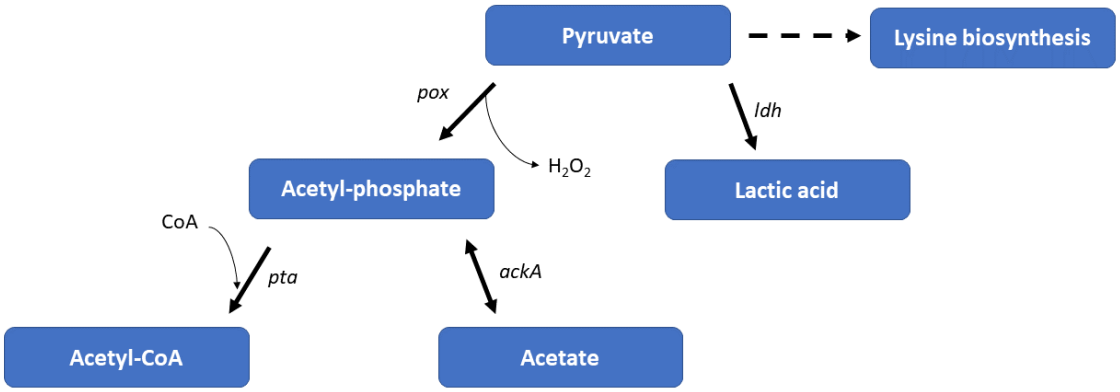


Figure 14.: Pyruvate metabolism in *L. acidophilus*. Pyruvate can be converted to lactic acid by lactate dehydrogenase (*ldh*), acetyl-phosphate by pyruvate oxidase (*pox*) or be used for the synthesis of lysine. Phosphate acetyltransferase (*pta*) converts acetyl-phosphate to acetyl-CoA. Acetyl-phosphate can also be converted to acetate by acetate kinase (*ack*).

If lysine is not available in the growth medium, pyruvate enters the "Lysine biosynthesis" pathway, in an intermediary reaction catalyzed by 4-hydroxy-tetrahydrodipicolinate synthase (EC 4.3.3.7).

Genes encoding pyruvate dehydrogenase and pyruvate formate lyase are absent in the genomes of species belonging to the *L. acidophilus* group [194]. Hence, these species need an alternative pathway to obtain acetyl-CoA, essential for the production of biomass precursors, such as N-acetyl-glucosamine.

In anaerobic conditions, *L. acidophilus* requires acetate for growth [200], which was also observed in *L. johnsonii* and *L. lactis* [194, 230]. In these species, the activity of acetate kinase allows the phosphorylation of acetate, generating acetyl-phosphate, which can be converted to acetyl-CoA by phosphate acetyltransferase. In *L. johnsonii*, the requirement for acetate can be replaced by oxygen, due to the activity of pyruvate oxidase. However, in such conditions hydrogen peroxide is produced, arresting growth by causing oxidative damage in the cell [194], which is not possible to simulate with *GSM* models.

4.6.3 *Fatty acid and lipid metabolism*

L. acidophilus La-14 does not have the necessary genes for the biosynthesis of fatty acids, as only two of the 13 required genes are present in the genome of this strain, which is in accordance with KEGG's reference annotation for *L. acidophilus* NCFM. However, regarding *L. acidophilus* 30SC and *L. acidophilus* FSI4, the genes required for the synthesis of fatty

acids are present in the genomes of both strains, thus this auxotrophy might be strain dependent. The requirement for fatty acids is described for *Lactobacillus johnsonii* [231, 232], which uses an exogenous source of fatty acids, usually polysorbate 80 (also known as Tween 80). This substance is predominantly composed of oleate (octadecenoic acid), although other fatty acids are also present [233]. Tween 80 has been proven to be essential for the growth of *L. acidophilus*, and cannot be replaced by free oleate [200]. The cleavage of esther bonds present in Tween 80 by lipases, provides free fatty acids ready to be incorporated in the cell [234]. Figure 15 represents the incorporation pathway of exogenous fatty acids by *L. acidophilus*.

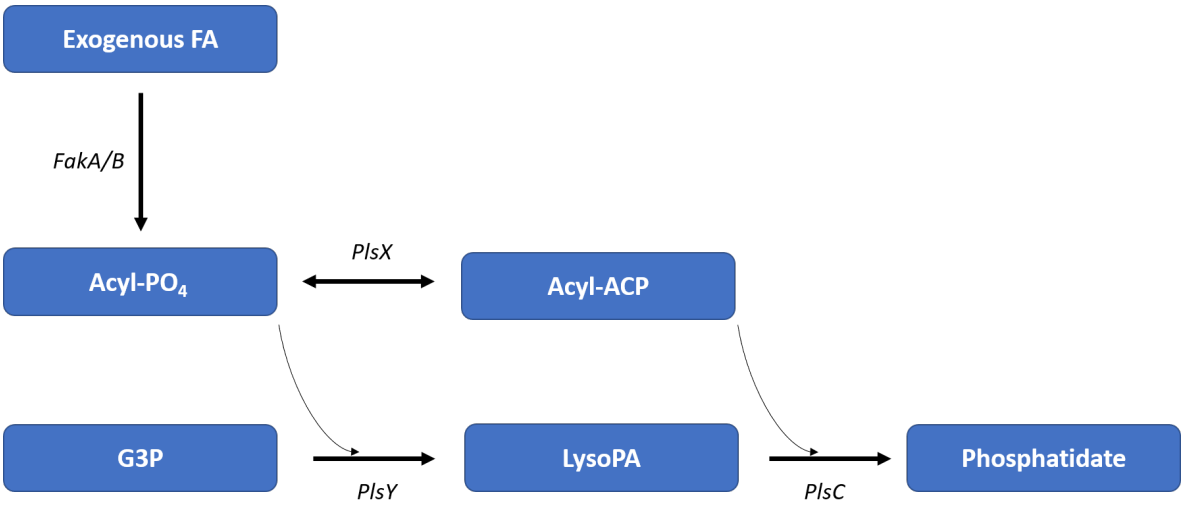


Figure 15.: Pathway for the incorporation of exogenous fatty acids, and phosphatidate synthesis. FA: Fatty acid; Acyl-PO_4: Acyl phosphate; G3P: glycerol-3-phosphate; LysoPA: Lysophosphatidate; Acyl-ACP: acyl-acyl-carrier-protein; FakA/B: Fatty acid kinase (2.7.-.-); PlsX: acyl-ACP:phosphate transacylase (EC 2.3.1.274); PlsY: glycerol-3-phosphate acyltransferase (EC 2.3.1.275); PlsC: 1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51)

In gram-positive bacteria, fatty acids cross the membrane by diffusional flipping, without requiring a transport protein [235, 236]. After entering the cell, fatty acids bound to the fatty acid-binding protein, FakB, and are then phosphorylated by fatty acid kinase, FakA, forming acyl phosphate. FakB delivers acyl phosphate to the membrane, where it is used by glycerol-3-phosphate acyltransferase to acylate glycerol-3-phosphate, forming lysophosphatidate. Additionally, acyl phosphate can be converted to acyl-ACP, by acyl-ACP:phosphate transacylase. In the final step, acyl-glycerol-3-phosphate acyltransferase converts lysophosphatidate and acyl-ACP to phosphatidate, which is then used in the phospholipid biosynthesis (figure 16).

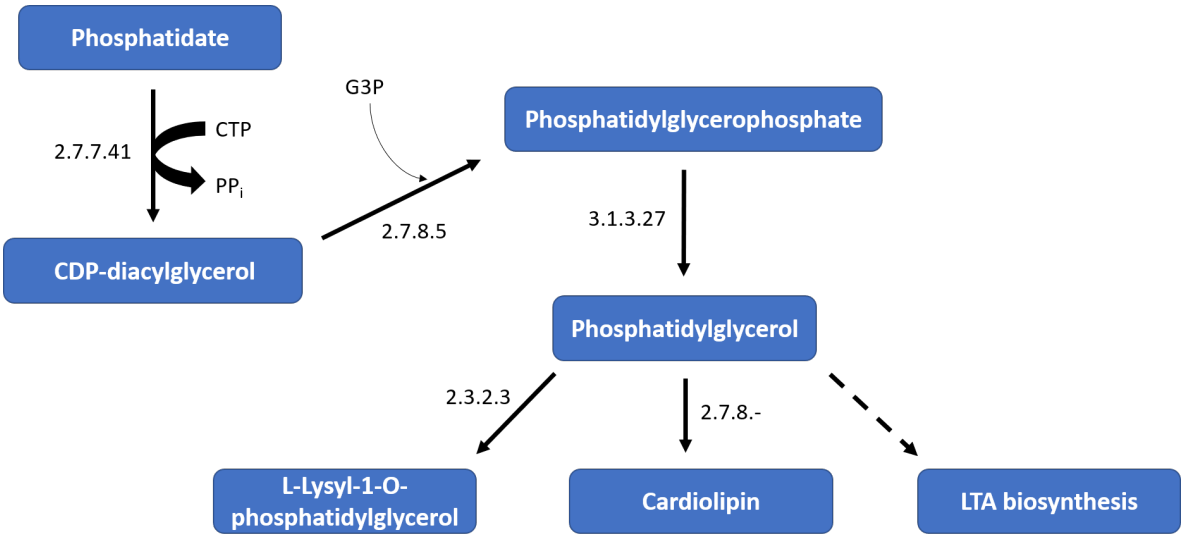


Figure 16.: Phospholipid metabolism in *L. acidophilus*. The present EC numbers have the following correspondence: 2.3.2.3 – lysyltransferase 2.7.7.41 – phosphatidate cytidylyltransferase; 2.7.8.- – Major cardiolipin synthase ClsA; 2.7.8.5 – CDP-diacylglycerol-glycerol-3-phosphate 1-phosphatidyltransferase; 3.1.3.27 – phosphatidylglycerophosphatase.

Phosphatidate cytidylyltransferase converts phosphatidate to CDP-diacylglycerol, which is used to obtain phosphatidyl glycerophosphate. The hydrolysis of this compound originates phosphatidylglycerol, which follows different pathways. It can be used to form cardiolipin, by cardiolipin synthase ClsA, and L-Lysyl-1-O-phosphatidylglycerol, by a lysyltransferase. These two phospholipids, together with phosphatidylglycerol, are the precursors of the lipid biomass component. Additionally, phosphatidylglycerol is used in the synthesis of the glycerol-phosphate polymer found in LTAs. In this polymerization reaction, diacylglycerol is released, which can be used to obtain phosphatidate over again, by diacylglycerol kinase.

4.6.4 *Purine and pyrimidine metabolism*

The requirements of *L. acidophilus* for nucleotides is not consensual: *Lovtrup et al* reported a need for deoxyribonucleotides; *Soska et al* [200] and *Morishita et al* [123] suggested that this species requires at least a purine (guanine or deoxyguanosine); *Lv et al* [206] verified that both purines and pyrimidines are necessary. These differences might be explained by the utilization of different strains in each study. No information regarding the nucleotide requirements was found for the La-14 strain.

The synthesis of both purine and pyrimidine nucleotides is dependent on the availability of HCO_3^-. This compound is used to obtain the pyrimidine precursor carbamoyl phosphate. In the "Purine metabolism" pathway, the intermediary reaction R07404 requires

HCO_3 as substrate. In several bacteria, carbonic anhydrase is responsible for the fast conversion of CO_2 to HCO_3^-. This enzyme is divided into five different families ($\alpha, \beta, \gamma, \delta$, and ζ) that do not share sequence similarities, as a result of convergent evolution [237]. However, no genomic evidence for the presence of any type of carbonic anhydrase in the *L. acidophilus* La-14 genome was found. The absence of this enzyme was also reported in *L. lactis* [238] and *L. johnsonii* [239]. Hence, these species are not capable of influencing the HCO_3^-/CO_2 equilibrium, depending on the non-enzymatic dissolution of CO_2 on water.

The complete pathway for the *de novo* synthesis of purines from ribose-5-phosphate and L-glutamine could be reconstructed based on genomic information. The presence of all essential genes regarding this biosynthetic capability was also reported in *L. acidophilus* NCFM [118].

However, guanine is often used as a source of purine bases in *L. acidophilus*. It can be converted to GMP by three phosphoribosyltransferases (EC 5.4.2.7, 5.4.2.8, 5.4.2.22), followed by phosphorylation to GDP by guanylate kinase. GTP can be formed from GDP either by pyruvate kinase (with phosphoenolpyruvate as phosphor donor) or by adenylate kinase (with ATP as phosphor donor). GTP is used for RNA synthesis, and to produce dGTP by ribonucleoside-triphosphate reductase, using thioredoxin or formate. In addition, deoxyguanosine and AMP (or other adenine nucleotide) are frequently included in CDM for this species.

The pathway for pyrimidine synthesis from L-glutamine and carbamoyl phosphate to UMP could be also completely reconstructed. After two phosphorylation reactions, UMP can be converted to UTP. CTP synthase converts UTP to CTP, which is used by ribonucleoside-triphosphate reductase to produce dCTP. Uridine phosphorylase can also catalyze the formation of deoxyuridine from uracil. Deoxyuridine is phosphorylated to dUMP, which is converted to dTMP. After two phosphorylation reactions catalyzed by dTMP kinase and adenylate kinase, dTTP is produced and directed to the DNA synthesis.

4.6.5 *Amino acid metabolism*

As mentioned before, *L. acidophilus* is auxotrophic for most amino acids, with the exception of alanine, asparagine, glutamine, glycine, and lysine. The metabolic model is able to produce threonine and cysteine, as demonstrated in figure 17.

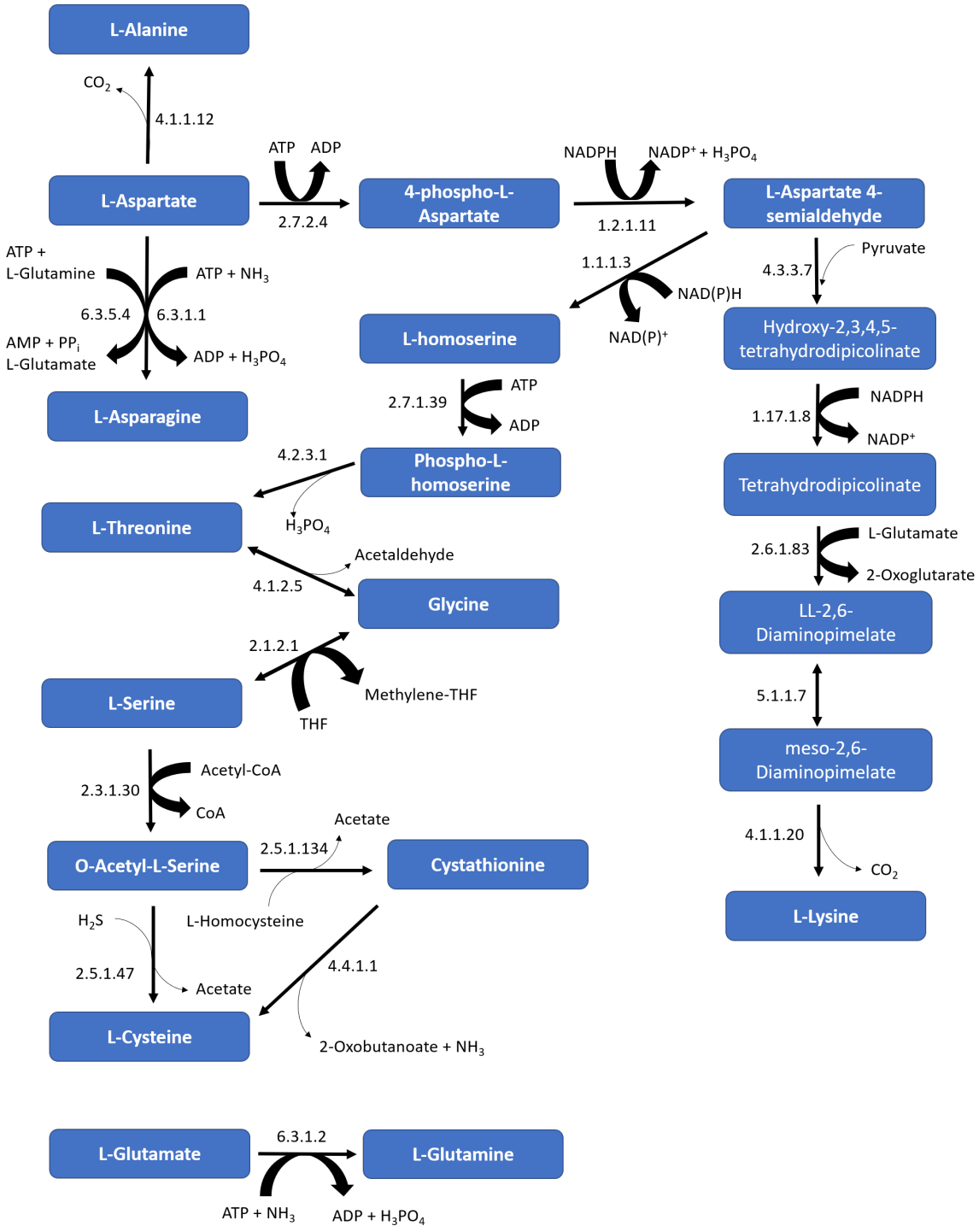


Figure 17.: Pathway for the synthesis of the non-essential amino acids alanine, asparagine, glycine, glutamine, and lysine in *L. acidophilus* La-14, from aspartate, serine, threonine, and glutamate. The shown EC numbers have the following correspondence: 1.1.1.3 – homoserine dehydrogenase; 1.2.1.11 – aspartate-semialdehyde dehydrogenase; 1.17.1.8 – 4-hydroxy-tetrahydrodipicolinate reductase; 2.1.2.1 – glycine hydroxymethyltransferase; 2.3.1.30 – serine O-acetyltransferase; 2.5.1.47 – cysteine synthase; 2.5.1.134 – cystathionine β-synthase; 2.6.1.83– LL-diaminopimelate aminotransferase; 2.7.1.39 – homoserine kinase; 2.7.2.4 – aspartate kinase; 4.4.1.1 – cystathionine γ-lyase; 4.1.1.12 – aspartate 4-decarboxylase; 4.1.1.20 – diaminopimelate decarboxylase; 4.1.2.5 – threonine aldolase; 4.2.3.1 – threonine synthase; 4.3.3.7 – 4-hydroxy-tetrahydrodipicolinate synthase; 5.1.1.7 – diaminopimelate epimerase; 6.3.1.1 – aspartate-ammonia ligase; 6.3.1.2 – glutamine synthetase; 6.3.5.4 – asparagine synthase.

Alanine can be produced directly from aspartate through a decarboxylation reaction catalyzed by aspartate 4-decarboxylase. Aspartate can also be used for the production of asparagine, either by asparagine synthase or aspartate-ammonia ligase. The phosphorylation of aspartate is the starting point for the lysine biosynthesis. This pathway includes seven reactions, and requires pyruvate and glutamate, besides ATP and NADPH. The intermediate aspartate 4-semialdehyde can be rerouted for the threonine biosynthesis. The bifunctional glycine hydroxymethyltransferase/threonine aldolase allows the interconversion among threonine, glycine, and serine.

Serine is the precursor for the synthesis of cysteine. After the conversion of serine to O-acetyl-serine by serine O-acetyltransferase, two routes can be followed: if hydrogen sulfide is available, cysteine can be produced directly from O-acetyl-serine; alternatively, L-homocysteine can be used as a sulfur source. This compound is produced from methionine, as observed in subsection 4.3.2.

Glutamine is obtained from glutamate by glutamine synthetase, using ammonia as a nitrogen source.

As expected, the metabolic model presents a limited amino acid biosynthesis capability. Non-essential amino acids are mainly produced directly from other amino acids. This limitation is a result of the adaptation of *L. acidophilus* to the gastrointestinal tract, where amino acids (and other nutrients) are available in abundance [3].

4.6.6 *Vitamin and cofactor metabolism*

L. acidophilus is auxotrophic for several vitamins and cofactors: pantothenate, folate, riboflavin, pyridoxal, and spermidine.

Pantothenate is present in the growth medium and is used in the e-Cofactor reaction, and for CoA synthesis. Pantothenate kinase phosphorylates pantothenate, forming 4-phosphopantothenate, which is converted to 4-phosphopantothenoyl-L-cysteine. The decarboxylation of this compound by phosphopantothenoylcysteine decarboxylase allows the obtention of 4-phosphopantetheine, followed by the conversion to dephospho-CoA. In the final step, dephospho-CoA is converted to CoA.

The activity of riboflavin kinase allows the direct conversion of riboflavin to FMN, which is used to produce FAD by FAD synthetase. These three cofactors are present in the e-Cofactor reaction.

Nicotinamide is the precursor of NAD^+ and $NADP^+$. Nicotinamide is hydrolyzed, forming nicotinate, which is directly converted to nicotinate D-ribonucleotide by nicotinate phosphoribosyltransferase. Then, nicotinate-nucleotide adenylyltransferase transfers an adenyl group from ATP, forming deamino-NAD^+. NAD^+ synthase catalyzes the final step, producing NAD^+, which can be also converted to $NADP^+$ by NAD^+ kinase.

Pyridoxal phosphate can be obtained directly from pyridoxal. Alternatively, pyridoxine can be used in the growth medium but requires oxygen to be converted to pyridoxal. S-adenosyl-L-methionine is directly obtained from L-methionine by methionine adenosyltransferase, with ATP consumption. Folate is the precursor of tetrahydrofolate, in a two-step reduction reaction, involving the oxidation of two NADPH molecules.

4.6.7 *Response to oxidative stress*

LAB inhabiting the human gut lack efficient mechanisms for protection against reactive oxygen species, such as catalase [204]. Although most *L. acidophilus* strains are aerotolerant, lower growth rates and biomass yields are achieved in aerobic conditions. The oxygen availability allows the activity of a few oxygen-dependent enzymes: pyruvate oxidase, pyridoxal 5-phosphate synthase, glycolate oxidase, and NADH oxidase. All these enzymes produce hydrogen peroxide, which can generate hydroxyl radicals, damaging the cell and thus inhibiting growth. The metabolic model presents two mechanisms to convert H_2O_2 to nontoxic compounds: NADH peroxidase, and tryparedoxin peroxidase. The presence of these mechanisms has been reported in *L. acidophilus* [215, 203, 216, 240].

4.7 PRODUCTION OF INDUSTRIALLY RELEVANT COMPOUNDS

Lactic acid bacteria produce several flavor and texture contributing compounds, like diacetyl, acetaldehyde, L-alanine, and EPS. Additionally, other compounds like lactic acid can be used as a preservation contributing agent. Strategies for more efficient production of these compounds have been focused on rerouting the pyruvate metabolism. However, as seen in subsection 4.6.2, the pyruvate flux in *L. acidophilus* is forwarded to lactic acid and acetyl-phosphate (in aerobic environments). Hence, redirecting the pyruvate metabolism will probably have to be associated with the introduction of heterologous genes.

4.7.1 *L-Lactic acid*

L. acidophilus produces lactic acid through the glycolytic pathway, in a racemic moisture of L and D-lactic acid. This fermentation product can be used as an acidifier (preserving food), flavor-enhancing agent by the food industry. In addition, it is employed by the cosmetics and pharmaceutical industries [241], and as the starting material of the synthetic biopolymer polylactic acid [242].

The margin for optimization of the production of lactic acid in homofermentative organisms is limited since this is the unique (or almost) fermentative product. In *L. acidophilus*, 90% of the glucose is converted to lactic acid. However, metabolic engineering approaches

have been focused on the production of L-lactic acid, since the D-isomer is toxic to humans. An obvious strategy is the inactivation of the *ldh*D gene, and the construction of strains with an additional copy of the *ldh*L gene. This approach has been successfully used in homofermentative LAB, such as *L. lactis* [243], *L. helveticus* [244, 245], and *L. johnsonii* [246].

The knockout of the *ldh*D gene was simulated, using LMOMA as simulation method, producing only the L-isomer. The production of L-lactic acid remained in the proportion of 1.8 $mol_{lactate}/mol_{glucose}$.

4.7.2 *Acetaldehyde*

Acetaldehyde is an important aroma compound present in dairy products. In the metabolic model, acetaldehyde can be obtained from three precursors: acetyl-CoA (acetaldehyde dehydrogenase); threonine (threonine aldolase); and deoxyribose 5-phosphate (deoxyribose-phosphate aldolase).

The rerouting of pyruvate for acetaldehyde production is possible under aerobic conditions due to the activity of pyruvate oxidase. The obtained acetyl phosphate can be converted to acetyl-CoA, which then forms acetaldehyde by acetaldehyde dehydrogenase. However, growth under aerobic conditions induce the production and accumulation of H_2O_2, limiting growth. Another alternative is the expression of pyruvate decarboxylase, which was already reported in *L. lactis*, using the *pdc* gene from *Zygomonas mobilis* [247].

The activity of threonine aldolase allows the production of glycine and acetaldehyde from threonine. The overexpression of this enzyme in *S. thermophillus* allowed an $80-90\%$ overproduction of acetaldehyde, compared with the wild-type strain [198].

Since *L. acidophilus* has two alcohol dehydrogenases, acetaldehyde can be converted to ethanol, lowering the aroma contribution of this species. Hence, the inactivation of this enzyme may increase the accumulation and secretion of acetaldehyde, as suggested by *Marshall et al* [197].

4.7.3 *L-Alanine*

The production of L-alanine (used as a food sweetener) in *L. lactis* was achieved through the introduction of the heterologous gene *ala*D, encoding alanine dehydrogenase of *Bacillus sphaericus*. This enzyme allows the conversion of pyruvate to alanine in a single enzymatic step. The introduction of *ala*D together with the knockout of *ldh* and *alr* (alanine racemase) genes resulted in the total conversion of pyruvate to alanine, turning the metabolism of *L. lactis* from homolactic to homoalanine [248].

In the *GSM* model, L-alanine can only be produced from L-aspartate, through aspartate decarboxylase. Since L-aspartate is an essential amino acid for *L. acidophilus*, the

rerouting of pyruvate to produce L-alanine must follow the same approach described for *L. lactis*.

4.7.4 *EPS*

The modulation of the production of EPS by lactic acid bacteria is an attractive subject of metabolic engineering, since these compounds influence the mouthfeel and texture of dairy products, presenting also prebiotic properties. Metabolic engineering approaches have been mainly focused on the synthesis of its precursors, to modulate the composition and structure of EPS [229]. As observed before, EPS precursors of *L. acidophilus* are UDP-glucose, UDP-galactose, and UDP-N-acetylglucosamine. The control points of the EPS synthesis are the conversion of glucose-6-phosphate to glucose-1-phosphate by phosphoglucomutase, and the synthesis of UDP-glucose from glucose-1-phosphate, by UDP-glucose phosphorylase. The overexpression of these two enzymes in *S. thermophilus* led to a 2-fold increase in EPS production [249]. Nevertheless, the utilization of the metabolic model to modulate the EPS production should involve alterations in the model, as EPS are biomass precursors.

5

CONCLUSION AND APPLICATIONS

A *GSM* model (iEC527) of *L. acidophilus* La-14 was developed in this work, containing 527 genes, 952 reactions, and 608 unique metabolites. The reconstruction of the metabolic model was based on information retrieved from biological databases and available literature and was accelerated by the utilization of automatic tools and plug-ins available in *merlin* and Optflux. Nevertheless, manual curation was still mandatory to account for specific characteristics of *L. acidophilus*.

The draft network was based on a previously obtained semi-automatic genome annotation. The high number of genes associated with transport proteins resulted in a high number of transport reactions, confirming the robust transport system of *L. acidophilus*. In fact, the metabolic model presents several auxotrophies regarding amino acids, pyrimidines and purines, fatty acids, and vitamins, demonstrating the low biosynthetic capability of this species.

The biomass composition was mainly inferred from data available in *GSM* models of closely related species, namely of *L. plantarum* and *L. lactis*. Although not having a significant impact on the qualitative predicting ability of the metabolic model, the experimental determination of the biomass composition of *L. acidophilus* La-14 would allow improving the qualitative *in silico* predictions of the model.

The growth rate assessment involved adjusting the maintenance ATP requirements. Lactic acid is the main compound produced from carbohydrate fermentation, confirming the homofermentative metabolism of *L. acidophilus*. Although lactose is the main sugar present in milk, higher growth rates are achieved using glucose as carbon source. The *in silico* fermentation pattern and the amino acid requirements are generally in agreement with available data. Slight differences were observed between critical genes identified in aerobic and anaerobic conditions. Genes associated with the biosynthesis of nucleotides and non-essential amino acids were additionally identified as critical in minimal media. The availability of data regarding gene essentiality of this species would be valuable for a deeper model validation. Moreover, transcriptomics and gene regulation data could also be integrated into the model. The lack of applicable experimental data (a CDM instead of rich/complex media, and quantification of consumed and produced metabolites) has

limited the validation of the metabolic model. The availability of quantitative experimental data would improve model curation, and consequently the phenotype predicting capability.

An overview of the main metabolic pathways in *L. acidophilus* was also performed in this work. The pathway for the metabolism of most commonly used carbohydrates was reconstructed according to *in silico* simulations, and compared with available information. The pathways for fatty acid incorporation, amino acid, nucleotide biosynthesis, and response to oxidative stress were also described.

The iEC527 model can be used in strain design approaches for optimization of the production of commercially relevant compounds, such as acetaldehyde and EPS. The rerouting of the pyruvate to the production of desirable compounds would probably involve the introduction of heterologous genes due to the limited pyruvate metabolism of this species.

L. acidophilus is usually used with other probiotics in dairy products. It is used in different types of cheese in addition to *L. casei*, *L. paracasei* or *Bifidobacterium* spp. The combination of this metabolic model with similar ones for these species should be useful to study and predict the phenotypic behavior of different LAB in the same environment.

BIBLIOGRAPHY

[1] B. Teusink and D. Molenaar, "Systems biology of lactic acid bacteria: For food and thought," *Current Opinion in Systems Biology*, vol. 6, pp. 7–13, 2017.

[2] Y. Zhu, Y. Zhang, and Y. Li, "Understanding the industrial application potential of lactic acid bacteria through genomics.," *Applied microbiology and biotechnology*, vol. 83, no. 4, pp. 597–610, 2009.

[3] N. Anjum, S. Maqsood, T. Masud, A. Ahmad, A. Sohail, and A. Momin, "*Lactobacillus acidophilus*: characterization of the species and application in food production.," *Critical reviews in food science and nutrition*, vol. 54, no. 9, pp. 1241–51, 2014.

[4] B. Stahl and R. Barrangou, "Complete genome sequence of probiotic strain *Lactobacillus acidophilus* la-14," *Microbiology Resource Announcements*, vol. 1, no. 3, 2013.

[5] H. Han, A. M. Segal, J. L. Seifter, and J. T. Dwyer, "Nutritional Management of Kidney Stones (Nephrolithiasis).," *Clinical nutrition research*, vol. 4, no. 3, pp. 137–52, 2015.

[6] D. Paineau, D. Carcano, G. Leyer, S. Darquy, M. A. Alyanakian, G. Simoneau, J. F. Bergmann, D. Brassart, F. Bornet, and A. C. Ouwehand, "Effects of seven potential probiotic strains on specific immune responses in healthy adults: A double-blind, randomized, controlled trial," *FEMS Immunology and Medical Microbiology*, 2008.

[7] R. Russell, M. Paterson, and N. Lima, "Molecular Biology of Food and Water Borne Mycotoxigenic and Mycotic Fungi," *Food Microbiology*, 2015.

[8] I. Thiele and B. Ø. Palsson, "A protocol for generating a high-quality genome-scale metabolic reconstruction.," *Nature protocols*, vol. 5, no. 1, pp. 93–121, 2010.

[9] O. Dias, M. Rocha, E. C. Ferreira, and I. Rocha, "Reconstructing genome-scale metabolic models with merlin.," *Nucleic acids research*, vol. 43, no. 8, pp. 3899–910, 2015.

[10] O. Dias, D. Gomes, P. Vilaca, J. Cardoso, M. Rocha, E. C. Ferreira, and I. Rocha, "Genome-Wide Semi-Automated Annotation of Transporter Systems.," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 14, no. 2, pp. 443–456, 2017.

[11] W. Cannon, *Bodily Changes in Pain, Hunger, Fear and Rage: An Account of Recent Researches into the Function of Emotional Excitement*. New York: D. Appleton and Company, 1915.

[12] N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*. Cambrige, MA: MIT Press, 1948.

[13] L. V. Bertalanffy, "An outline of general system theory," *British Journal for the Philosophy of Science*, 1950.

[14] H. Kitano, "Systems biology: A brief overview," *Science*, vol. 295, no. 5560, pp. 1662–1664, 2002.

[15] E. Stalidzans, A. Seiman, K. Peebo, V. Komasilovs, and A. Pentjuss, "Model-based metabolism design: constraints for kinetic and stoichiometric models," *Biochemical Society Transactions*, 2018.

[16] O. Øyås and J. Stelling, "Genome-scale metabolic networks in time and space," *Current Opinion in Systems Biology*, 2017.

[17] E. T. Papoutsakis, "Equations and calculations for fermentations of butyric acid bacteria," *Biotechnology and Bioengineering*, vol. 26, pp. 174–187, 1984.

[18] E. T. Papoutsakis and C. L. Meyer, "Fermentation equations for propionic-acid bacteria and production of assorted oxychemicals from various sugars.," *Biotechnology and bioengineering*, vol. 27, no. 1, pp. 67–80, 1985.

[19] R. A. Majewski and M. M. Domach, "Simple constrained-optimization view of acetate overflow in *E. coli*.," *Biotechnology and bioengineering*, vol. 35, no. 7, pp. 732–8, 1990.

[20] A. M. Feist, M. J. Herrgård, I. Thiele, J. L. Reed, and B. Ø. Palsson, "Reconstruction of biochemical networks in microorganisms.," *Nature reviews. Microbiology*, vol. 7, no. 2, pp. 129–43, 2009.

[21] I. Rocha, J. Förster, and J. Nielsen, "Design and application of genome-scale reconstructed metabolic models.," *Methods in molecular biology*, vol. 416, pp. 409–31, 2008.

[22] M. A. Oberhardt, B. Palsson, and J. A. Papin, "Applications of genome-scale metabolic reconstructions," *Molecular Systems Biology*, vol. 5, no. 320, pp. 1–15, 2009.

[23] E. W. Sayers et al, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research*, vol. 37, no. SUPPL. 1, 2009.

[24] S. Mukherjee, D. Stamatis, J. Bertsch, G. Ovchinnikova, H. Y. Katta, A. Mojica, I.-M. A. Chen, N. C. Kyrpides, and T. Reddy, "Genomes online database (gold) v.7: updates and new features," *Nucleic Acids Research*, p. gky977, 2018.

[25] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes.," *Nucleic acids research*, vol. 28, no. 1, pp. 27–30, 2000.

[26] R. Caspi et al, "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases," *Nucleic Acids Research*, 2014.

[27] P. Artimo, M. Jonnalagedda, K. Arnold, D. Baratin, G. Csardi, E. de Castro, S. Duvaud, V. Flegel, A. Fortier, E. Gasteiger, A. Grosdidier, C. Hernandez, V. Ioannidis, D. Kuznetsov, R. Liechti, S. Moretti, K. Mostaguir, N. Redaschi, G. Rossier, I. Xenarios, and H. Stockinger, "ExPASy: SIB bioinformatics resource portal.," *Nucleic acids research*, vol. 40, no. Web Server issue, pp. W597–603, 2012.

[28] The UniProt Consortium, "Ongoing and future developments at the Universal Protein Resource.," *Nucleic acids research*, vol. 39, pp. D214–9, 2011.

[29] I. Schomburg, A. Chang, O. Hofmann, C. Ebeling, F. Ehrentreich, and D. Schomburg, "BRENDA: a resource for enzyme data and metabolic information.," *Trends in biochemical sciences*, vol. 27, no. 1, pp. 54–6, 2002.

[30] Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. O. Palsson, and N. E. Lewis, "BiGG Models: A platform for integrating, standardizing and sharing genome-scale models.," *Nucleic acids research*, vol. 44, no. D1, pp. D515–22, 2016.

[31] M. H. Saier, "TCDB: the Transporter Classification Database for membrane transport protein analyses and information," *Nucleic Acids Research*, vol. 34, no. 90001, pp. D181–D186, 2006.

[32] Q. Ren, K. Chen, and I. T. Paulsen, "TransportDB: A comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels," *Nucleic Acids Research*, 2007.

[33] J. Hastings, G. Owen, A. Dekker, M. Ennis, N. Kale, V. Muthukrishnan, S. Turner, N. Swainston, P. Mendes, and C. Steinbeck, "ChEBI in 2016: Improved services and an expanding collection of metabolites.," *Nucleic acids research*, vol. 44, no. D1, pp. D1214–9, 2016.

[34] N. Y. Yu, J. R. Wagner, M. R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S. C. Sahinalp, M. Ester, L. J. Foster, and F. S. L. Brinkman, "PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes.," *Bioinformatics (Oxford, England)*, vol. 26, no. 13, pp. 1608–15, 2010.

[35] T. Goldberg, M. Hecht, T. Hamp, T. Karl, G. Yachdav, N. Ahmed, U. Altermann, P. Angerer, S. Ansorge, K. Balasz, M. Bernhofer, A. Betz, L. Cizmadija, K. T. Do, J. Gerke, R. Greil, V. Joerdens, M. Hastreiter, K. Hembach, M. Herzog, M. Kalemanov, M. Kluge,

A. Meier, H. Nasir, U. Neumaier, V. Prade, J. Reeb, A. Sorokoumov, I. Troshani, S. Vorberg, S. Waldraff, J. Zierer, H. Nielsen, and B. Rost, "LocTree3 prediction of localization.," *Nucleic acids research*, vol. 42, no. Web Server issue, pp. W350–5, 2014.

[36] O. Emanuelsson, H. Nielsen, S. Brunak, and G. von Heijne, "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.," *Journal of molecular biology*, vol. 300, no. 4, pp. 1005–16, 2000.

[37] A. Bairoch, "Serendipity in bioinformatics, the tribulations of a Swiss bioinformatician through exciting times!," *Bioinformatics (Oxford, England)*, vol. 16, no. 1, pp. 48–64, 2000.

[38] Y. M. Park, S. Squizzato, N. Buso, T. Gur, and R. Lopez, "The EBI search engine: EBI search as a servicemaking biological data accessible for all," *Nucleic Acids Research*, vol. 45, no. W1, pp. W545–W549, 2017.

[39] W. C. Barker, J. S. Garavelli, H. Huang, P. B. McGarvey, B. C. Orcutt, G. Y. Srinivasarao, C. Xiao, L. S. Yeh, R. S. Ledley, J. F. Janda, F. Pfeiffer, H. W. Mewes, A. Tsugita, and C. Wu, "The protein information resource (PIR).," *Nucleic acids research*, vol. 28, no. 1, pp. 41–4, 2000.

[40] I. U. of Biochemistry, M. B. N. Committee, and E. C. E. C. Webb, *Enzyme nomenclature 1992 : recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*. San Diego ; Sydney : Published for the International Union of Biochemistry and Molecular Biology by Academic Press, rev. ed ed., 1992. "A revision of the Recommendations (1984) of the Nomenclature Committee of IUB.".

[41] M. H. Saier, "A functional-phylogenetic classification system for transmembrane solute transporters.," *Microbiology and molecular biology reviews : MMBR*, vol. 64, no. 2, pp. 354–411, 2000.

[42] O. Dias, A. K. Gombert, E. C. Ferreira, and I. Rocha, "Genome-wide metabolic (re-) annotation of Kluyveromyces lactis.," *BMC genomics*, vol. 13, p. 517, 2012.

[43] S. L. Salzberg, A. L. Delcher, S. Kasif, and O. White, "Microbial gene identification using interpolated Markov models.," *Nucleic acids research*, vol. 26, no. 2, pp. 544–8, 1998.

[44] S. L. Salzberg, M. Pertea, A. L. Delcher, M. J. Gardner, and H. Tettelin, "Interpolated Markov models for eukaryotic gene finding.," *Genomics*, vol. 59, no. 1, pp. 24–31, 1999.

[45] C. Burge and S. Karlin, "Prediction of complete gene structures in human genomic DNA," *Journal of Molecular Biology*, 1997.

[46] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool.," *Journal of molecular biology*, vol. 215, no. 3, pp. 403–10, 1990.

[47] S. R. Eddy, "Profile hidden Markov models.," *Bioinformatics (Oxford, England)*, vol. 14, no. 9, pp. 755–763, 1998.

[48] P. M. Bowers, M. Pellegrini, M. J. Thompson, J. Fierro, T. O. Yeates, and D. Eisenberg, "Prolinks: a database of protein functional linkages derived from coevolution.," *Genome biology*, vol. 5, no. 5, p. R35, 2004.

[49] R. Overbeek, M. Fonstein, M. D'Souza, G. D. Pusch, and N. Maltsev, "The use of gene clusters to infer functional coupling.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 6, pp. 2896–901, 1999.

[50] A. J. Enright, I. Illopoulos, N. C. Kyrpides, and C. A. Ouzounis, "Protein interaction maps for complete genomes based on gene fusion events," *Nature*, 1999.

[51] J. L. Reed, I. Famili, I. Thiele, and B. O. Palsson, "Towards multidimensional genome annotation.," *Nature reviews. Genetics*, vol. 7, no. 2, pp. 130–41, 2006.

[52] N. Swainston, K. Smallbone, P. Mendes, D. Kell, and N. Paton, "The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks.," *Journal of integrative bioinformatics*, vol. 8, no. 2, p. 186, 2011.

[53] D. Machado, M. J. Herrgård, and I. Rocha, "Stoichiometric Representation of Gene-Protein-Reaction Associations Leverages Constraint-Based Analysis from Reaction to Gene-Level Phenotype Prediction.," *PLoS computational biology*, vol. 12, no. 10, p. e1005140, 2016.

[54] G. J. E. Baart and D. E. Martens, "Genome-Scale Metabolic Models: Reconstruction and Analysis," in *Methods in Molecular Biology*, pp. 107–126, Humana Press, 2012.

[55] J. Izard and R. J. Limberger, "Rapid screening method for quantitation of bacterial cell lipids from whole cells.," *Journal of microbiological methods*, vol. 55, no. 2, pp. 411–8, 2003.

[56] A. Varma and B. O. Palsson, "Metabolic Capabilities of *Escherichia coli* II. Optimal Growth Patterns," *Journal of Theoretical Biology*, vol. 165, no. 4, pp. 503–522, 1993.

[57] J. M. Comeron and M. Aguadé, "An evaluation of measures of synonymous codon usage bias.," *Journal of molecular evolution*, vol. 47, no. 3, pp. 268–74, 1998.

[58] S. Santos and I. Rocha, "Estimation of biomass composition from genomic and transcriptomic information.," *Journal of integrative bioinformatics*, vol. 13, no. 2, p. 285, 2016.

[59] J. C. Xavier, K. R. Patil, and I. Rocha, "Integration of Biomass Formulations of Genome-Scale Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in Prokaryotes.," *Metabolic engineering*, vol. 39, pp. 200–208, 2017.

[60] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. Ø. Palsson, "A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.," *Molecular systems biology*, vol. 3, p. 121, 2007.

[61] L. Liu, R. Agren, S. Bordel, and J. Nielsen, "Use of genome-scale metabolic models for understanding microbial physiology.," *FEBS letters*, vol. 584, no. 12, pp. 2556–64, 2010.

[62] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, and SBML Forum, "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.," *Bioinformatics (Oxford, England)*, vol. 19, no. 4, pp. 524–31, 2003.

[63] I. Rocha, P. Maia, P. Evangelista, P. Vilaça, S. Soares, J. P. Pinto, J. Nielsen, K. R. Patil, E. C. Ferreira, and M. Rocha, "OptFlux: an open-source software platform for in silico metabolic engineering," *BMC Systems Biology*, vol. 4, no. 1, p. 45, 2010.

[64] J. Schellenberger, R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. Ø. Palsson, "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0," *Nature Protocols*, vol. 6, no. 9, pp. 1290–1307, 2011.

[65] G. Stephanopoulos, A. Aristidou, and J. Nielsen, "Metabolic engineering: principles and methodologies," *San Diego: Academic Google Scholar*, 1998.

[66] I. Famili, J. Forster, J. Nielsen, and B. 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*, vol. 100, no. 23, pp. 13134–9, 2003.

[67] J. D. Orth, I. Thiele, and B. Ø. Palsson, "What is flux balance analysis?," *Nature Biotechnology*, vol. 28, no. 3, pp. 245–248, 2010.

[68] R. Mahadevan and C. H. Schilling, "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models.," *Metabolic engineering*, vol. 5, no. 4, pp. 264–76, 2003.

[69] D. Segre, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 23, pp. 15112–15117, 2002.

[70] T. Shlomi, O. Berkman, and E. 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*, vol. 102, no. 21, pp. 7695–700, 2005.

[71] K. R. Patil, I. Rocha, J. Förster, and J. Nielsen, "Evolutionary programming as a platform for in silico metabolic engineering.," *BMC bioinformatics*, vol. 6, p. 308, 2005.

[72] M. Rocha, P. Maia, R. Mendes, J. P. Pinto, E. C. Ferreira, J. Nielsen, K. R. Patil, and I. Rocha, "Natural computation meta-heuristics for the in silico optimization of microbial strains.," *BMC bioinformatics*, vol. 9, p. 499, 2008.

[73] O. R. Gonzalez, C. Küper, K. Jung, P. C. Naval, and E. Mendoza, "Parameter estimation using Simulated Annealing for S-system models of biochemical networks.," *Bioinformatics (Oxford, England)*, vol. 23, no. 4, pp. 480–6, 2007.

[74] S. Pabinger, R. Snajder, T. Hardiman, M. Willi, A. Dander, and Z. Trajanoski, "MEMOSys 2.0: an update of the bioinformatics database for genome-scale models and genomic data," *Database*, vol. 2014, p. bau004, 2014.

[75] E. Pitkänen, P. Jouhten, J. Hou, M. F. Syed, P. Blomberg, J. Kludas, M. Oja, L. Holm, M. Penttilä, J. Rousu, and M. Arvas, "Comparative genome-scale reconstruction of gapless metabolic networks for present and ancestral species.," *PLoS computational biology*, vol. 10, no. 2, p. e1003465, 2014.

[76] R. Overbeek, R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A. R. Wattam, F. Xia, and R. Stevens, "The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).," *Nucleic acids research*, vol. 42, no. Database issue, pp. D206–14, 2014.

[77] R. Agren, L. Liu, S. Shoaie, W. Vongsangnak, I. Nookaew, and J. Nielsen, "The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum.," *PLoS computational biology*, vol. 9, no. 3, p. e1002980, 2013.

[78] D. Lagoa, F. Liu, E. C. Ferreira, J. Faria, C. Henry, and O. Dias, "Towards a genome-wide transport systems encoding genes tracker.," in *COBRA 2018 - 5th Conference on Constraint-Based Reconstruction and Analysis.*, Seattle WA, USA, 10 2018.

[79] H. Ma and A.-P. Zeng, "Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms.," *Bioinformatics (Oxford, England)*, vol. 19, pp. 270–7, jan 2003.

[80] S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson, and M. J. Herrgard, "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox.," *Nature protocols*, vol. 2, no. 3, pp. 727–38, 2007.

[81] A. Funahashi, M. Morohashi, H. Kitano, and N. Tanimura, "CellDesigner: a process diagram editor for gene-regulatory and biochemical networks," *BIOSILICO*, vol. 1, pp. 159–162, nov 2003.

[82] A. Funahashi, Y. Matsuoka, A. Jouraku, M. Morohashi, N. Kikuchi, and H. Kitano, "CellDesigner 3.5: A versatile modeling tool for biochemical networks," *Proceedings of the IEEE*, 2008.

[83] A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke, "COBRApy: COnstraints-Based Reconstruction and Analysis for Python.," *BMC systems biology*, vol. 7, p. 74, 2013.

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

[85] N. E. Lewis, H. Nagarajan, and B. O. Palsson, "Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods.," *Nature reviews. Microbiology*, vol. 10, no. 4, pp. 291–305, 2012.

[86] K. R. Patil, M. Akesson, and J. Nielsen, "Use of genome-scale microbial models for metabolic engineering.," *Current opinion in biotechnology*, vol. 15, no. 1, pp. 64–9, 2004.

[87] W. J. Kim, H. U. Kim, and S. Y. Lee, "Current state and applications of microbial genome-scale metabolic models," *Current Opinion in Systems Biology*, vol. 2, pp. 10–18, 2017.

[88] O. Folger, L. Jerby, C. Frezza, E. Gottlieb, E. Ruppin, and T. Shlomi, "Predicting selective drug targets in cancer through metabolic networks.," *Molecular systems biology*, vol. 7, p. 501, 2011.

[89] B. Teusink and E. J. Smid, "Modelling strategies for the industrial exploitation of lactic acid bacteria.," *Nature reviews. Microbiology*, vol. 4, no. 1, pp. 46–56, 2006.

[90] B. Teusink, H. Bachmann, and D. Molenaar, "Systems biology of lactic acid bacteria: a critical review.," *Microbial cell factories*, vol. 10 Suppl 1, p. S11, 2011.

[91] F. Branco dos Santos, W. M. de Vos, and B. Teusink, "Towards metagenome-scale models for industrial applicationsthe case of Lactic Acid Bacteria," *Current Opinion in Biotechnology*, vol. 24, no. 2, pp. 200–206, 2013.

[92] M. H. Rau and A. A. Zeidan, "Constraint-based modeling in microbial food biotechnology.," *Biochemical Society transactions*, vol. 46, no. 2, pp. 249–260, 2018.

[93] N. A. L. Flahaut, A. Wiersma, B. van de Bunt, D. E. Martens, P. J. Schaap, L. Sijtsma, V. A. M. Dos Santos, and W. M. de Vos, "Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation.," *Applied microbiology and biotechnology*, vol. 97, no. 19, pp. 8729–39, 2013.

[94] A. P. Oliveira, J. Nielsen, and J. Förster, "Modeling *Lactococcus lactis* using a genome-scale flux model.," *BMC microbiology*, vol. 5, p. 39, 2005.

[95] M. I. Pastink, B. Teusink, P. Hols, S. Visser, W. M. de Vos, and J. Hugenholtz, "Genome-scale model of *Streptococcus thermophilus* LMG18311 for metabolic comparison of lactic acid bacteria.," *Applied and environmental microbiology*, vol. 75, no. 11, pp. 3627–33, 2009.

[96] E. Vinay-Lara, J. J. Hamilton, B. Stahl, J. R. Broadbent, J. L. Reed, and J. L. Steele, "Genome -scale reconstruction of metabolic networks of *Lactobacillus casei* ATCC 334 and 12A," *PLoS ONE*, 2014.

[97] N. Xu, J. Liu, L. Ai, and L. Liu, "Reconstruction and analysis of the genome-scale metabolic model of *Lactobacillus casei* LC2W," *Gene*, vol. 554, no. 2, pp. 140–147, 2015.

[98] S. N. Mendoza, P. M. Cañón, Á. Contreras, M. Ribbeck, and E. Agosín, "Genome-Scale Reconstruction of the Metabolic Network in Oenococcus oeni to Assess Wine Malolactic Fermentation.," *Frontiers in microbiology*, vol. 8, p. 534, 2017.

[99] L. Koduru, Y. Kim, J. Bang, M. Lakshmanan, N. S. Han, and D.-Y. Lee, "Genome-scale modeling and transcriptome analysis of Leuconostoc mesenteroides unravel the redox governed metabolic states in obligate heterofermentative lactic acid bacteria.," *Scientific reports*, vol. 7, no. 1, p. 15721, 2017.

[100] J. Hugenholtz, "The lactic acid bacterium as a cell factory for food ingredient production," *International Dairy Journal*, vol. 18, no. 5, pp. 466 – 475, 2008. Netherlands Association for the Advancement of Dairy Science 1908-2008.

[101] D. M. Waters, A. Mauch, A. Coffey, E. K. Arendt, and E. Zannini, "Lactic Acid Bacteria as a Cell Factory for the Delivery of Functional Biomolecules and Ingredients in Cereal-Based Beverages: A Review," *Critical Reviews in Food Science and Nutrition*, 2015.

[102] L. Brown, E. Vera Pingitore, F. Mozzi, L. Saavedra, J. M. Villegas, and E. M. Hebert, "Lactic acid bacteria as cell factories for the generation of bioactive peptides," *Protein & Peptide Letters*, 2017.

[103] S. S. Fong and B. Palsson, "Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes," *Nature Genetics*, 2004.

[104] M. B. Biggs, G. L. Medlock, G. L. Kolling, and J. A. Papin, "Metabolic network modeling of microbial communities.," *Wiley interdisciplinary reviews. Systems biology and medicine*, vol. 7, no. 5, pp. 317–34, 2015.

[105] C. S. Henry, H. C. Bernstein, P. Weisenhorn, R. C. Taylor, J.-Y. Lee, J. Zucker, and H.-S. Song, "Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction.," *Journal of cellular physiology*, vol. 231, no. 11, pp. 2339–45, 2016.

[106] E. Bosi, G. Bacci, A. Mengoni, and M. Fondi, "Perspectives and challenges in microbial communities metabolic modeling," *Frontiers in Genetics*, vol. 8, p. 88, 2017.

[107] F. Mozzi, R. R. Raya, and G. M. Vignolo, *Biotechnology of lactic acid bacteria: Novel applications: Second edition*. Wiley-Blackwell, 2015.

[108] P. G. Heineman, "Orla-Jensen's Classification of Lactic Acid Bacteria," *Journal of Dairy Science*, vol. 3, no. 2, pp. 143–155, 1920.

[109] S. Lahtinen, A. Ouwehand, S. Salminen, and A. von Wright, *Lactic Acid Bacteria: Microbiological and Functional Aspects, Fourth Edition*. Taylor & Francis, 2011.

[110] E. J. Quinto, P. Jiménez, I. Caro, J. Tejero, J. Mateo, and T. Girbés, "Probiotic Lactic Acid Bacteria: A Review," *Food and Nutrition Sciences*, vol. 05, no. 18, pp. 1765–1775, 2014.

[111] E. Stefanovic, G. Fitzgerald, and O. McAuliffe, "Advances in the genomics and metabolomics of dairy lactobacilli: A review.," *Food microbiology*, vol. 61, pp. 33–49, 2017.

[112] O. O'Sullivan, J. O'Callaghan, A. Sangrador-Vegas, O. McAuliffe, L. Slattery, P. Kaleta, M. Callanan, G. F. Fitzgerald, R. P. Ross, and T. Beresford, "Comparative genomics of lactic acid bacteria reveals a niche-specific gene set," *BMC Microbiology*, vol. 9, no. 1, p. 50, 2009.

[113] A. A. Zeidan, V. K. Poulsen, T. Janzen, P. Buldo, P. M. F. Derkx, G. Øregaard, and A. R. Neves, "Polysaccharide production by lactic acid bacteria: from genes to industrial applications.," *FEMS microbiology reviews*, vol. 41, no. Supp_1, pp. S168–S200, 2017.

[114] M. J. Bull, K. A. Jolley, J. E. Bray, M. Aerts, P. Vandamme, M. C. J. Maiden, J. R. Marchesi, and E. Mahenthiralingam, "The domestication of the probiotic bacterium *Lactobacillus acidophilus*," *Scientific Reports*, vol. 4, no. 1, p. 7202, 2015.

[115] J. L. JOHNSON, C. F. PHELPS, C. S. CUMMINS, J. LONDON, and F. GASSER, "Taxonomy of the *Lactobacillus acidophilus* Group," *International Journal of Systematic Bacteriology*, vol. 30, no. 1, pp. 53–68, 1980.

[116] E. Lauer, C. Helming, and O. Kandler, "Heterogeneity of the Species *Lactobacillus acidophilus* (Moro) Hansen and Moquot as Revealed by Biochemical Characteristics and DNA-DNA hybridisation," *Zentralblatt für Bakteriologie: I. Abt. Originale C: Allgemeine, angewandte und ökologische Mikrobiologie*, vol. 1, no. 2, pp. 150–168, 1980.

[117] K. Selle, T. Klaenhammer, and W. Russell, "*Lactobacillus — Lactobacillus acidophilus*," in *Encyclopedia of Food Microbiology*, pp. 412–417, Elsevier, 2014.

[118] E. Altermann, W. M. Russell, M. A. Azcarate-Peril, R. Barrangou, B. L. Buck, O. McAuliffe, N. Souther, A. Dobson, T. Duong, M. Callanan, S. Lick, A. Hamrick, R. Cano, and T. R. Klaenhammer, "Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM," *Proceedings of the National Academy of Sciences*, vol. 102, no. 11, pp. 3906–3912, 2005.

[119] O. Iartchouk, S. Kozyavkin, V. Karamychev, and A. Slesarev, "Complete Genome Sequence of *Lactobacillus acidophilus* FSI4, Isolated from Yogurt," *Genome Announcements*, vol. 3, no. 2, 2015.

[120] W.-H. Chung, J. Kang, M. Y. Lim, T.-J. Lim, S. Lim, S. W. Roh, and Y.-D. Nam, "Complete genome sequence and genomic characterization of *Lactobacillus acidophilus* la1 (11869bp)," *Frontiers in Pharmacology*, vol. 9, p. 83, Aug 2018.

[121] S. Oh, H. Roh, H.-J. Ko, S. Kim, K. H. Kim, S. E. Lee, I. S. Chang, S. Kim, and I.-G. Choi, "Complete Genome Sequencing of *Lactobacillus acidophilus* 30SC, Isolated from Swine Intestine," *Journal of Bacteriology*, vol. 193, no. 11, pp. 2882–2883, 2011.

[122] P. K. Gopal, "Lactic Acid Bacteria: *Lactobacillus* spp.: *Lactobacillus acidophilus*," in *Encyclopedia of Dairy Sciences: Second Edition*, pp. 91–95, Academic Press, 2011.

[123] T. Morishita, Y. Deguchi, M. Yajima, T. Sakurai, and T. Yura, "Multiple nutritional requirements of lactobacilli: genetic lesions affecting amino acid biosynthetic pathways.," *Journal of bacteriology*, vol. 148, no. 1, pp. 64–71, 1981.

[124] R. Temmerman, B. Pot, G. Huys, and J. Swings, "Identification and antibiotic susceptibility of bacterial isolates from probiotic products," *International Journal of Food Microbiology*, vol. 81, no. 1, pp. 1–10, 2003.

[125] D. Olukoya, S. Ebigwei, O. Adebawo, and F. Osiyemi, "Plasmid profiles and antibiotic susceptibility patterns of *Lactobacillus* isolated from fermented foods in Nigeria," *Food Microbiology*, vol. 10, no. 4, pp. 279–285, 1993.

[126] W. P. CHARTERIS, P. M. KELLY, L. MORELLI, and J. K. COLLINS, "Antibiotic Susceptibility of Potentially Probiotic *Lactobacillus* Species," *Journal of Food Protection*, vol. 61, no. 12, pp. 1636–1643, 1998.

[127] S. D. Todorov, D. N. Furtado, S. M. I. Saad, and B. D. Gombossy de Melo Franco, "Bacteriocin production and resistance to drugs are advantageous features for *Lactobacillus acidophilus* La-14, a potential probiotic strain.," *The new microbiologica*, vol. 34, no. 4, pp. 357–70, 2011.

[128] J. Delcour, T. Ferain, M. Deghorain, E. Palumbo, and P. Hols, "The biosynthesis and functionality of the cell-wall of lactic acid bacteria.," *Antonie van Leeuwenhoek*, vol. 76, no. 1-4, pp. 159–84, 1999.

[129] Z. Wu, D. Pan, X. Zeng, Y. Sun, and J. Cao, "Phosphorylation of peptidoglycan from *Lactobacillus acidophilus* and its immunoregulatory function," *International Journal of Food Science & Technology*, vol. 51, no. 3, pp. 664–671, 2016.

[130] Parvaneh Jafarei, "*Lactobacillus acidophilus* cell structure and application," *African Journal of Microbiology Research*, vol. 5, no. 24, pp. 4033–4042, 2011.

[131] A. Bera, R. Biswas, S. Herbert, E. Kulauzovic, C. Weidenmaier, A. Peschel, and F. Gotz, "Influence of Wall Teichoic Acid on Lysozyme Resistance in Staphylococcus aureus," *Journal of Bacteriology*, vol. 189, no. 1, pp. 280–283, 2007.

[132] A. Peschel, C. Vuong, M. Otto, and F. Gotz, "The D-Alanine Residues of Staphylococcus aureus Teichoic Acids Alter the Susceptibility to Vancomycin and the Activity of Autolytic Enzymes," *Antimicrobial Agents and Chemotherapy*, vol. 44, no. 10, pp. 2845–2847, 2000.

[133] H. Yamamoto, Y. Miyake, M. Hisaoka, S.-i. Kurosawa, and J. Sekiguchi, "The major and minor wall teichoic acids prevent the sidewall localization of vegetative dl-endopeptidase LytF in *Bacillus subtilis*," *Molecular Microbiology*, vol. 70, no. 2, pp. 297–310, 2008.

[134] S. HEPTINSTALL, A. R. ARCHIBALD, and J. BADDILEY, "Teichoic Acids and Membrane Function in Bacteria," *Nature*, vol. 225, no. 5232, pp. 519–521, 1970.

[135] M.-P. Chapot-Chartier, "Interactions of the cell-wall glycopolymers of lactic acid bacteria with their bacteriophages," *Frontiers in Microbiology*, vol. 5, p. 236, 2014.

[136] H. J. Boot, C. P. Kolen, J. M. van Noort, and P. H. Pouwels, "S-layer protein of *Lactobacillus acidophilus* ATCC 4356: purification, expression in *Escherichia coli*, and nucleotide sequence of the corresponding gene.," *Journal of Bacteriology*, vol. 175, no. 19, pp. 6089–6096, 1993.

[137] H. J. Boot, C. P. Kolen, F. J. Andreadaki, R. J. Leer, and P. H. Pouwels, "The *Lactobacillus acidophilus* S-layer protein gene expression site comprises two consensus promoter sequences, one of which directs transcription of stable mRNA.," *Journal of Bacteriology*, vol. 178, no. 18, pp. 5388–5394, 1996.

[138] E. K. Lam, L. Yu, H. P. Wong, W. K. Wu, V. Y. Shin, E. K. Tai, W. H. So, P. C. Woo, and C. Cho, "Probiotic *Lactobacillus rhamnosus* GG enhances gastric ulcer healing in rats," *European Journal of Pharmacology*, vol. 565, no. 1-3, pp. 171–179, 2007.

[139] L. De Vuyst and B. Degeest, "Heteropolysaccharides from lactic acid bacteria," *FEMS Microbiology Reviews*, vol. 23, no. 2, pp. 153–177, 1999.

[140] S. Riaz, S. K. Nawaz, and S. Hasnain, "Bacteriocins produced by L. fermentum and *L .acidophilus* can inhibit cephalosporin resistant *E .coli*.," *Brazilian Journal of Microbiology*, vol. 41, no. 3, pp. 643–648, 2010.

[141] S. Amdekar, P. Roy, V. Singh, A. Kumar, R. Singh, and P. Sharma, "Anti-Inflammatory Activity of *Lactobacillus* on Carrageenan-Induced Paw Edema in Male Wistar Rats," *International Journal of Inflammation*, vol. 2012, pp. 1–6, 2012.

[142] V. Rosenfeldt, K. F. Michalsen, M. Jakobsen, C. N. Larsen, P. L. MØller, P. Pedersen, M. Tvede, H. Weyrehter, N. H. Valerius, and A. PÆrregaard, "Effect of probiotic *Lactobacillus* strains in young children hospitalized with acute diarrhea," *The Pediatric Infectious Disease Journal*, vol. 21, no. 5, pp. 411–416, 2002.

[143] Y.-J. Yang, C.-C. Chuang, H.-B. Yang, C.-C. Lu, and B.-S. Sheu, "*Lactobacillus acidophilus* ameliorates *H. pylori*-induced gastric inflammation by inactivating the Smad7 and NFκB pathways," *BMC Microbiology*, vol. 12, no. 1, p. 38, 2012.

[144] R. Nagpal, A. Kumar, M. Kumar, P. V. Behare, S. Jain, and H. Yadav, "Probiotics, their health benefits and applications for developing healthier foods: a review," *FEMS Microbiology Letters*, vol. 334, no. 1, pp. 1–15, 2012.

[145] H. Maroof, Z. M. Hassan, A. M. Mobarez, and M. A. Mohamadabadi, "*Lactobacillus acidophilus* Could Modulate the Immune Response Against Breast Cancer in Murine Model," *Journal of Clinical Immunology*, vol. 32, no. 6, pp. 1353–1359, 2012.

[146] H. S. Ejtahed, J. Mohtadi-Nia, A. Homayouni-Rad, M. Niafar, M. Asghari-Jafarabadi, and V. Mofid, "Probiotic yogurt improves antioxidant status in type 2 diabetic patients," *Nutrition*, vol. 28, no. 5, pp. 539–543, 2012.

[147] S. F. Vilela, J. O. Barbosa, R. D. Rossoni, J. D. Santos, M. C. Prata, A. L. Anbinder, A. O. Jorge, and J. C. Junqueira, "*Lactobacillus acidophilus* ATCC 4356 inhibits biofilm formation by C. albicans and attenuates the experimental candidiasis in Galleria mellonella," *Virulence*, vol. 6, no. 1, pp. 29–39, 2015.

[148] O. Cortés-Zavaleta, A. López-Malo, A. Hernández-Mendoza, and H. García, "Antifungal activity of *Lactobacilli* and its relationship with 3-phenyllactic acid production," *International Journal of Food Microbiology*, vol. 173, pp. 30–35, 2014.

[149] S.-E. Jang, J.-J. Jeong, S.-Y. Choi, H. Kim, M. Han, and D.-H. Kim, "*Lactobacillus rhamnosus* HN001 and *Lactobacillus acidophilus* La-14 Attenuate Gardnerella vaginalis-Infected Bacterial Vaginosis in Mice," *Nutrients*, vol. 9, no. 6, p. 531, 2017.

[150] S. Giardina, C. Scilironi, A. Michelotti, A. Samuele, F. Borella, M. Daglia, and F. Marzatic, "Re: In Vitro Anti-Inflammatory Activity of Selected Oxalate-Degrading Probiotic Bacteria: Potential Applications in the Prevention and Treatment of Hyperoxaluria," *The Journal of Urology*, vol. 79, no. 3, pp. 384–391, 2014.

[151] F. Cruz, D. Lagoa, J. Mendes, I. Rocha, E. C. Ferreira, M. Rocha, and O. Dias, "SamPler a novel method for selecting parameters for gene functional annotation routines," *BMC Bioinformatics*, 2019.

[152] Y.-K. Oh, B. O. Palsson, S. M. Park, C. H. Schilling, and R. Mahadevan, "Genome-scale Reconstruction of Metabolic Network in *Bacillus subtilis* Based on High-throughput Phenotyping and Gene Essentiality Data," *Journal of Biological Chemistry*, vol. 282, pp. 28791–28799, sep 2007.

[153] S. T. Santos, "Development Of Computational Methods For The Determination Of Biomass Composition And Evaluation Of Its Impact In Genome-Scale Models Predictions," *Masters thesis, Universidade do Minho, Braga*, 2013.

[154] S. Magnúsdóttir, A. Heinken, L. Kutt, D. A. Ravcheev, E. Bauer, A. Noronha, K. Greenhalgh, C. Jäger, J. Baginska, P. Wilmes, R. M. T. Fleming, and I. Thiele, "Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota," *Nature Biotechnology*, vol. 35, pp. 81–89, jan 2017.

[155] S. Brinster, G. Lamberet, B. Staels, P. Trieu-Cuot, A. Gruss, and C. Poyart, "Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens," *Nature*, vol. 458, pp. 83–86, mar 2009.

[156] S. Turroni, B. Vitali, C. Bendazzoli, M. Candela, R. Gotti, F. Federici, F. Pirovano, and P. Brigidi, "Oxalate consumption by lactobacilli: evaluation of oxalyl-CoA decarboxylase and formyl-CoA transferase activity in *Lactobacillus acidophilus*," *Journal of Applied Microbiology*, vol. 103, pp. 1600–1609, nov 2007.

[157] M. A. Azcarate-Peril, J. M. Bruno-Barcena, H. M. Hassan, and T. R. Klaenhammer, "Transcriptional and Functional Analysis of Oxalyl-Coenzyme A (CoA) Decarboxylase and Formyl-CoA Transferase Genes from *Lactobacillus acidophilus*," *Applied and Environmental Microbiology*, vol. 72, pp. 1891–1899, mar 2006.

[158] D. G. Assimos, "In Vitro Anti-Inflammatory Activity of Selected Oxalate-Degrading Probiotic Bacteria: Potential Applications in the Prevention and Treatment of Hyperoxaluria," *Journal of Urology*, vol. 192, pp. 1130–1130, oct 2014.

[159] C. Bendazzoli, S. Turroni, R. Gotti, S. Olmo, P. Brigidi, and V. Cavrini, "Determination of oxalyl-coenzyme A decarboxylase activity in Oxalobacter formigenes and *Lactobacillus acidophilus* by capillary electrophoresis," *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 2007.

[160] M. J. Kullen and T. R. Klaenhammer, "Identification of the pH-inducible, proton-translocating F1Fo-ATPase (atpBEFHAGDC) operon of *Lactobacillus acidophilus* by differential display: gene structure, cloning and characterization," *Molecular Microbiology*, vol. 33, pp. 1152–1161, mar 2002.

[161] R. Barrangou, M. A. Azcarate-Peril, T. Duong, S. B. Conners, R. M. Kelly, and T. R. Klaenhammer, "Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays," *Proceedings of the National Academy of Sciences*, vol. 103, pp. 3816–3821, mar 2006.

[162] M. A. Azcarate-Peril, E. Altermann, R. L. Hoover-Fitzula, R. J. Cano, and T. R. Klaenhammer, "Identification and Inactivation of Genetic Loci Involved with *Lactobacillus acidophilus* Acid Tolerance," *Applied and Environmental Microbiology*, vol. 70, pp. 5315–5322, sep 2004.

[163] M. Azcarate-Peril, R. Tallon, and T. Klaenhammer, "Temporal gene expression and probiotic attributes of *Lactobacillus acidophilus* during growth in milk," *Journal of Dairy Science*, vol. 92, pp. 870–886, mar 2009.

[164] K. Abe, H. Hayashi, and P. C. Maloney, "Exchange of Aspartate and Alanine," *Journal of Biological Chemistry*, vol. 271, pp. 3079–3084, feb 1996.

[165] M. J. Allison, K. A. Dawson, W. R. Mayberry, and J. G. Foss, "Oxalobacter formigenes gen. nov., sp. nov.: oxalate-degrading anaerobes that inhabit the gastrointestinal tract," *Archives of Microbiology*, vol. 141, pp. 1–7, feb 1985.

[166] A. Taku, K. G. Gunetileke, and R. A. Anwar, "Biosynthesis of uridine diphospho-N-acetylmuramic acid. 3. Purification and properties of uridine diphospho-N-acetylenolpyruvyl-glucosamine reductase.," *Journal of Biological Chemistry*, vol. 245, no. 19, pp. 5012–5016, 1970.

[167] R. Blakley and McDougall, "Dihydrofolic reductase from Streptococcus faecalis R.," *Journal of Biological Chemistry*, vol. 236, pp. 1163–1167, 1961.

[168] N. C. Major and A. T. Bull, "The physiology of lactate production byLactobacillus delbreuckii in a chemostat with cell recycle," *Biotechnology and Bioengineering*, vol. 34, pp. 592–599, aug 1989.

[169] D. Bergmaier, C. Champagne, and C. Lacroix, "Growth and exopolysaccharide production during free and immobilized cell chemostat culture of *Lactobacillus rhamnosus* RW-9595M," *Journal of Applied Microbiology*, vol. 98, pp. 272–284, feb 2005.

[170] J. H. Veerkamp, "Fatty acid composition of *Bifidobacterium* and *Lactobacillus* strains.," *Journal of Bacteriology*, vol. 108, no. 2, pp. 861–867, 1971.

[171] F. A. Exterkate, B. J. Otten, H. W. Wassenberg, and J. H. Veerkamp, "Comparison of the phospholipid composition of *Bifidobacterium* and *Lactobacillus* strains.," *Journal of Bacteriology*, vol. 106, no. 3, pp. 824–829, 1971.

[172] W. Fischer, H. U. Koch, P. Rösel, and F. Fiedler, "Alanine ester-containing native lipoteichoic acids do not act as lipoteichoic acid carrier. Isolation, structural and functional characterization.," *Journal of Biological Chemistry*, vol. 155, no. 10, pp. 4557–4562, 1980.

[173] O. V. Ledesma, A. P. De Ruiz Holgado, G. Oliver, G. S. D. Giori, P. Raibaud, and J. V. Galpin, "A Synthetic Medium for Comparative Nutritional Studies of Lactobacilli," *Journal of Applied Bacteriology*, vol. 42, pp. 123–133, feb 1977.

[174] S. E. Jones, K. Whitehead, D. Saulnier, C. M. Thomas, J. Versalovic, and R. A. Britton, "Cyclopropane fatty acid synthase mutants of probiotic human-derived *Lactobacillus reuteri* are defective in TNF inhibition," *Gut Microbes*, vol. 2, pp. 69–79, mar 2011.

[175] G. Robijn, "Structural characterization of the exopolysaccharide produced by *Lactobacillus acidophilus* LMG9433," *Carbohydrate Research*, vol. 288, pp. 203–218, jul 1996.

[176] A. P. Laws, M. J. Chadha, M. Chacon-Romero, V. M. Marshall, and M. Maqsood, "Determination of the structure and molecular weights of the exopolysaccharide produced by *Lactobacillus acidophilus* 5e2 when grown on different carbon feeds," *Carbohydrate Research*, vol. 343, pp. 301–307, feb 2008.

[177] Z. Yang, M. Staaf, E. Huttunen, and G. Widmalm, "Structure of a viscous exopolysaccharide produced by *Lactobacillus helveticus* K16," *Carbohydrate Research*, vol. 329, pp. 465–469, nov 2000.

[178] Y. Yamamoto, S. Murosaki, R. Yamauchi, K. Kato, and Y. Sone, "Structural study on an exocellular polysaccharide produced by *Lactobacillus helveticus* TY1-2," *Carbohydrate Research*, vol. 261, pp. 67–78, aug 1994.

[179] Y. Yamamoto, T. Nunome, R. Yamauchi, K. Kato, and Y. Sone, "Structure of an exocellular polysaccharide of *Lactobacillus helveticus* TN-4, a spontaneous mutant strain of *Lactobacillus helveticus* TY1-2," *Carbohydrate Research*, vol. 275, pp. 319–332, oct 1995.

[180] M. Staaf, Z. Yang, E. Huttunen, and G. Widmalm, "Structural elucidation of the viscous exopolysaccharide produced by *Lactobacillus helveticus* Lb161," *Carbohydrate Research*, vol. 326, pp. 113–119, jun 2000.

[181] M. Staaf, G. Widmalm, Z. Yang, and E. Huttunen, "Structural elucidation of an extracellular polysaccharide produced by *Lactobacillus helveticus*," *Carbohydrate Research*, vol. 291, pp. 155–164, sep 1996.

[182] G. W. Robijn, J. R. Thomas, H. Haas, D. J. van den Berg, J. P. Kamerling, and J. F. Vliegenthart, "The structure of the exopolysaccharide produced by *Lactobacillus helveticus* 766," *Carbohydrate Research*, vol. 276, pp. 137–154, oct 1995.

[183] J. Hugenholtz, E. Looijesteijn, M. Starrenburg, and C. Dijkema, "Analysis of sugar metabolism in an eps producing *Lactococcus lactis* by 31p nmr," *Journal of Biotechnology*, vol. 77, no. 1, pp. 17 – 23, 2000. NMR in Biotechnology.

[184] T. Shiraishi, S. Yokota, S. Fukiya, and A. Yokota, "Structural diversity and biological significance of lipoteichoic acid in Gram-positive bacteria: focusing on beneficial probiotic lactic acid bacteria," *Bioscience of Microbiota, Food and Health*, vol. 35, no. 4, pp. 147–161, 2016.

[185] K. Hamana, T. Akiba, F. Uchino, and S. Matsuzaki, "Distribution of spermine in *bacilli* and lactic acid bacteria," *Canadian Journal of Microbiology*, vol. 35, pp. 450–455, apr 1989.

[186] M. Rogosa, "Characters used in the classification of lactobacilli," *International Journal of Systematic Bacteriology*, vol. 20, pp. 519–533, oct 1970.

[187] F. Ozogul and I. Hamed, "Lactic acid bacteria: *Lactobacillus* spp.: *Lactobacillus acidophilus*," in *Reference Module in Food Science*, Elsevier, 2016.

[188] M. Kilstrup, K. Hammer, P. Ruhdaljensen, and J. Martinussen, "Nucleotide metabolism and its control in lactic acid bacteria," *FEMS Microbiology Reviews*, vol. 29, pp. 555–590, aug 2005.

[189] V. Crow and G. Pritchard, "Pyruvate kinase from streptococcus lactis," in *Carbohydrate Metabolism - Part E* (W. A. Wood, ed.), vol. 90 of *Methods in Enzymology*, pp. 165 – 170, Academic Press, 1982.

[190] J.-W. Baick, J.-H. Yoon, S. Namgoong, D. Soll, S.-I. Kim, S. Eom, and K.-W. Hong, "Growth inhibition of *Escherichia coli* during heterologous expression of *Bacillus subtilis* glutamyl-trna synthetase that catalyzes the formation of mischarged glutamyl-trna1 gln," *Journal of microbiology (Seoul, Korea)*, vol. 42, pp. 111–6, 07 2004.

[191] M. Frederix and J. A. Downie, "Chapter 2 - quorum sensing: Regulating the regulators," vol. 58 of *Advances in Microbial Physiology*, pp. 23 – 80, Academic Press, 2011.

[192] B. Buck, M. Azcarate-Peril, and T. Klaenhammer, "Role of autoinducer-2 on the adhesion ability of *Lactobacillus acidophilus*," *Journal of Applied Microbiology*, vol. 107, pp. 269–279, jul 2009.

[193] M. W Hickey, A. Hillier, and G. Jago, "Metabolism of Pyruvate and Citrate in *Lactobacilli*," *Australian Journal of Biological Sciences*, vol. 36, no. 6, p. 487, 1983.

[194] R. Y. Hertzberger, R. D. Pridmore, C. Gysler, M. Kleerebezem, and M. J. Teixeira de Mattos, "Oxygen Relieves the CO2 and Acetate Dependency of *Lactobacillus johnsonii* NCC 533," *PLoS ONE*, vol. 8, p. e57235, feb 2013.

[195] R. Hertzberger, J. Arents, H. L. Dekker, R. D. Pridmore, C. Gysler, M. Kleerebezem, and M. J. T. de Mattos, "H_2O_2 Production in Species of the *Lactobacillus acidophilus* Group: a Central Role for a Novel NADH-Dependent Flavin Reductase," *Applied and Environmental Microbiology*, vol. 80, pp. 2229–2239, apr 2014.

[196] H. M. stlie, M. H. Helland, and J. A. Narvhus, "Growth and metabolism of selected strains of probiotic bacteria in milk," *International Journal of Food Microbiology*, vol. 87, no. 1, pp. 17 – 27, 2003.

[197] V. M. Marshall and W. M. Cole, "Threonine aldolase and alcohol dehydrogenase activities in *Lactobacillus bulgaricus* and *Lactobacillus acidophilus* and their contribution to flavour production in fermented milks," *Journal of Dairy Research*, vol. 50, pp. 375–379, aug 1983.

[198] A. C. S. D. Chaves, M. Fernandez, A. L. S. Lerayer, I. Mierau, M. Kleerebezem, and J. Hugenholtz, "Metabolic Engineering of Acetaldehyde Production by *Streptococcus thermophilus*," *Applied and Environmental Microbiology*, vol. 68, pp. 5656–5662, nov 2002.

[199] Y. Miyamoto, T. Masaki, and S. Chohnan, "Characterization of n-deoxyribosyltransferase from *Lactococcus lactis* subsp. lactis," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1774, no. 10, pp. 1323 – 1330, 2007.

[200] J. Soska, "Growth of *Lactobacillus acidophilus* in the absence of folic acid.," *Journal of Bacteriology*, vol. 91, p. 18401847, May 1966.

[201] E. Collins and K. Aramaki, "Production of Hydrogen Peroxide by *Lactobacillus acidophilus*," *Journal of Dairy Science*, vol. 63, pp. 353–357, mar 1980.

[202] E. Kot, S. Furmanov, and A. Bezkorovainy, "Ferrous iron oxidation by *Lactobacillus acidophilus* and its metabolic products," *Journal of Agricultural and Food Chemistry*, vol. 43, pp. 1276–1282, may 1995.

[203] A. Talwalkar, K. Kailasapathy, J. Hourigan, P. Peiris, and R. Arumugaswamy, "An improved method for the determination of NADH oxidase in the presence of NADH peroxidase in lactic acid bacteria," *Journal of Microbiological Methods*, 2003.

[204] A. Talwalkar and K. Kailasapathy, "The role of oxygen in the viability of probiotic bacteria with reference to L. acidophilus and Bifidobacterium spp," 2004.

[205] R. D. Pridmore, A. C. Pittet, F. Praplan, and C. Cavadini, "Hydrogen peroxide production by *Lactobacillus johnsonii* NCC 533 and its role in anti-Salmonella activity," *FEMS Microbiology Letters*, 2008.

[206] X. Lv, G. Liu, X. Sun, H. Chen, J. Sun, and Z. Feng, "Short communication: Nutrient consumption patterns of *Lactobacillus acidophilus* KLDS 1.0738 in controlled pH batch fermentations," *Journal of Dairy Science*, vol. 100, pp. 5188–5194, jul 2017.

[207] A. M. Gomes and F. Malcata, "Bifidobacterium spp. and *Lactobacillus acidophilus*: biological, biochemical, technological and therapeutical properties relevant for use as probiotics," *Trends in Food Science & Technology*, vol. 10, pp. 139–157, apr 1999.

[208] P. G. Bruinenberg, G. De Roo, and G. K. Limsowtin, "urification and characterization of cystathionine γ-lyase from *Lactococcus lactis* subsp. cremoris SK11: Possible role in flavor compound formation during cheese maturation," *Applied and Environmental Microbiology*, vol. 63, no. 2, p. 561566, 1997.

[209] D. J. Manning, "Chemical production of essential Cheddar flavour compounds," *Journal of Dairy Research*, vol. 46, pp. 531–537, jul 1979.

[210] R. Sreekumar, Z. AlAttabi, H. Deeth, and M. Turner, "Volatile sulfur compounds produced by probiotic bacteria in the presence of cysteine or methionine," *Letters in Applied Microbiology*, vol. 48, pp. 777–782, 6 2009.

[211] W.-J. Lee, D. S. Banavara, J. E. Hughes, J. K. Christiansen, J. L. Steele, J. R. Broadbent, and S. A. Rankin, "Role of Cystathionine -Lyase in Catabolism of Amino Acids to Sulfur Volatiles by Genetic Variants of *Lactobacillus helveticus* CNRZ 32," *Applied and Environmental Microbiology*, vol. 73, pp. 3034–3039, may 2007.

[212] S. Irmler, H. SchÃfer, B. Beisert, D. Rauhut, and H. Berthoud, "Identification and characterization of a strain-dependent cystathionine Î/Î-lyase in *Lactobacillus casei* potentially involved in cysteine biosynthesis," *FEMS Microbiology Letters*, vol. 295, pp. 67–76, jun 2009.

[213] B. Aslm, "Poly-β-hydroxybutyrate production by lactic acid bacteria," *FEMS Microbiology Letters*, vol. 159, pp. 293–297, feb 1998.

[214] A. Hamieh, Z. Olama, and H. Holail, "Microbial production of polyhydroxybutyrate , a biodegradable plastic using agro-industrial waste," *Global Advanced Research Journal of Microbiology*, vol. 2, no. 3, 2013.

[215] A. Talwalkar, K. Kailasapathy, P. Peiris, and R. Arumugaswamy, "Application of rbgra simple way for screening of oxygen tolerance in probiotic bacteria," *International Journal of Food Microbiology*, vol. 71, no. 2, pp. 245 – 248, 2001.

[216] M. Sakamoto and K. Komagata, "Aerobic growth of and activities of NADH oxidase and NADH peroxidase in lactic acid bacteria," *Journal of Fermentation and Bioengineering*, 1996.

[217] D. M. Wheater, "The Characteristics of *Lactobacillus acidophilus* and *Lactobacillus bulgaricus*," *Journal of General Microbiology*, vol. 12, pp. 123–132, feb 1955.

[218] L. Axelsson and S. Lindgren, "Characterization and DNA homology of *Lactobacillus* strains isolated from pig intestine," *Journal of Applied Bacteriology*, vol. 62, pp. 433–440, may 1987.

[219] N. Stern, F. Konishi, C. Hesseltine, and H. Wang, "*Lactobacillus acidophilus* Utilization of Sugars and Production of a Fermented Soybean Product," *Canadian Institute of Food Science and Technology Journal*, vol. 10, pp. 197–200, jul 1977.

[220] D. Srinivas, B. Mital, and S. Garg, "Utilization of sugars by *Lactobacillus acidophilus* strains," *International Journal of Food Microbiology*, vol. 10, pp. 51–57, jan 1990.

[221] D. Charalampopoulos, S. Pandiella, and C. Webb, "Growth studies of potentially probiotic lactic acid bacteria in cereal-based substrates," *Journal of Applied Microbiology*, vol. 92, pp. 851–859, may 2002.

[222] J. B. Smart, C. J. Pillidge, and J. H. Garman, "Growth of lactic acid bacteria and bifidobacteria on lactose and lactose-related mono-, di- and trisaccharides and correlation with distribution of β galactosidase and phospho- β galactosidase," *Journal of Dairy Research*, vol. 60, pp. 557–568, nov 1993.

[223] J. M. Andersen, R. Barrangou, M. Abou Hachem, S. Lahtinen, Y. J. Goh, B. Svensson, and T. R. Klaenhammer, "Transcriptional and functional analysis of galactooligosaccharide uptake by lacS in *Lactobacillus acidophilus*," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 17785–17790, oct 2011.

[224] J. M. Andersen, R. Barrangou, M. A. Hachem, S. J. Lahtinen, Y.-J. Goh, B. Svensson, and T. R. Klaenhammer, "Transcriptional Analysis of Prebiotic Uptake and Catabolism by *Lactobacillus acidophilus* NCFM," *PLoS ONE*, vol. 7, p. e44409, sep 2012.

[225] T. Duong, R. Barrangou, W. M. Russell, and T. R. Klaenhammer, "Characterization of the tre Locus and Analysis of Trehalose Cryoprotection in *Lactobacillus acidophilus* NCFM," *Applied and Environmental Microbiology*, vol. 72, pp. 1218–1225, feb 2006.

[226] J. Thompson, A. Pikis, S. B. Ruvinov, B. Henrissat, H. Yamamoto, and J. Sekiguchi, "The gene *glvA* of *Bacillus subtilis* 168 encodes a metal-requiring, nad(h)-dependent 6-phospho--glucosidase: Assignment to family 4 of the glycosylhydrolase superfamily," *Journal of Biological Chemistry*, vol. 273, no. 42, pp. 27347–27356, 1998.

[227] S. Benthin, J. Nielsen, and J. Villadsen, "Two uptake systems for fructose in *Lactococcus lactis* subsp. cremoris FD1 produce glycolytic and gluconeogenic fructose phosphates and induce oscillations in growth and lactic acid formation," *Applied and Environmental Microbiology*, vol. 59, no. 10, pp. 3206–3211, 1993.

[228] O. Kandler, "Carbohydrate metabolism in lactic acid bacteria," *Antonie van Leeuwenhoek*, vol. 49, no. 3, pp. 209–224, 1983.

[229] M. Papagianni, "METABOLIC ENGINEERING OF LACTIC ACID BACTERIA FOR THE PRODUCTION OF INDUSTRIALLY IMPORTANT COMPOUNDS," *Computational and Structural Biotechnology Journal*, vol. 3, p. e201210003, oct 2012.

[230] E. B. Collins and J. C. Bruhn, "Roles of acetate and pyruvate in the metabolism of Streptococcus diacetilactis.," *Journal of Bacteriology*, vol. 103, no. 3, pp. 541–546, 1970.

[231] J. Boekhorst, "The complete genomes of *Lactobacillus plantarum* and *Lactobacillus johnsonii* reveal extensive differences in chromosome organization and gene content," *Microbiology*, vol. 150, pp. 3601–3611, nov 2004.

[232] J. A. Muller, R. P. Ross, W. F. H. Sybesma, G. F. Fitzgerald, and C. Stanton, "Modification of the Technical Properties of *Lactobacillus johnsonii* NCC 533 by Supplementing

the Growth Medium with Unsaturated Fatty Acids," *Applied and Environmental Microbiology*, vol. 77, pp. 6889–6898, oct 2011.

[233] D. Ilko, A. Braun, O. Germershaus, L. Meinel, and U. Holzgrabe, "Fatty acid composition analysis in polysorbate 80 with high performance liquid chromatography coupled to charged aerosol detection," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 94, pp. 569–574, aug 2015.

[234] C. M. Toutain-Kidd, S. C. Kadivar, C. T. Bramante, S. A. Bobin, and M. E. Zegans, "Polysorbate 80 Inhibition of Pseudomonas aeruginosa Biofilm Formation and Its Cleavage by the Secreted Lipase LipA," *Antimicrobial Agents and Chemotherapy*, vol. 53, pp. 136–145, jan 2009.

[235] J. B. Parsons, T. C. Broussard, J. L. Bose, J. W. Rosch, P. Jackson, C. Subramanian, and C. O. Rock, "Identification of a two-component fatty acid kinase responsible for host fatty acid incorporation by Staphylococcus aureus," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 10532–10537, jul 2014.

[236] J. E. Cronan, "A new pathway of exogenous fatty acid incorporation proceeds by a classical phosphoryl transfer reaction," *Molecular Microbiology*, vol. 92, pp. 217–221, apr 2014.

[237] G. Dubey, B. Kollah, U. Ahirwar, and S. Mohanty, "Carbon dioxide metabolism and ecological significance of enzyme complex systems in terrestrial ecosystem," *Life Science Clusters*, vol. 1, pp. 35–45, 04 2018.

[238] F. Arsène-Ploetze and F. Bringel, "Role of inorganic carbon in lactic acid bacteria metabolism," in *Lait*, 2004.

[239] R. Y. Hertzberger, "Encounters with oxygen: Aerobic physiology and ho production of *Lactobacillus johnsonii*.," 2014.

[240] Q. Guo, S. Li, Y. Xie, Q. Zhang, M. Liu, Z. Xu, H. Sun, and Y. Yang, "The nad+-dependent deacetylase, bifidobacterium longum sir2 in response to oxidative stress by deacetylating sigh (h) and foxo3a in bifidobacterium longum and hek293t cell respectively," *Free Radical Biology and Medicine*, vol. 108, pp. 929 – 939, 2017.

[241] A. J. Maris, W. N. Konings, J. P. Dijken, and J. T. Pronk, "Microbial export of lactic and 3-hydroxypropanoic acid: implications for industrial fermentation processes," *Metabolic Engineering*, vol. 6, pp. 245–255, oct 2004.

[242] D.-M. Bai, X.-M. Zhao, X.-G. Li, and S.-M. Xu, "Strain improvement and metabolic flux analysis in the wild-type and a mutant *Lactobacillus lactis* strain for L(+)-lactic acid production," *Biotechnology and Bioengineering*, vol. 88, pp. 681–689, dec 2004.

[243] H. W. Andersen, M. B. Pedersen, K. Hammer, and P. R. Jensen, "Lactate dehydrogenase has no control on lactate production but has a strong negative control on formate production in *Lactococcus lactis*," *European Journal of Biochemistry*, vol. 268, pp. 6379–6389, dec 2001.

[244] T. Bhowmik and J. L. Steele, "Cloning, characterization and insertional inactivation of the*Lactobacillus helveticus* D() lactate dehydrogenase gene," *Applied Microbiology and Biotechnology*, vol. 41, pp. 432–439, jun 1994.

[245] K. Kyla-Nikkila, M. Hujanen, M. Leisola, and A. Palva, "Metabolic Engineering of *Lactobacillus helveticus* CNRZ32 for Production of Pure L-(+)-Lactic Acid," *Applied and Environmental Microbiology*, vol. 66, pp. 3835–3841, sep 2000.

[246] L. Lapierre, J. E. Germond, A. Ott, M. Delley, and B. Mollet, "D-Lactate dehydrogenase gene (ldhD) inactivation and resulting metabolic effects in the *Lactobacillus johnsonii* strains La1 and N312," *Applied and Environmental Microbiology*, vol. 65, no. 9, pp. 4002–4007, 1999.

[247] R. S. Bongers, M. H. N. Hoefnagel, and M. Kleerebezem, "High-Level Acetaldehyde Production in *Lactococcus lactis* by Metabolic Engineering," *Applied and Environmental Microbiology*, vol. 71, pp. 1109–1113, feb 2005.

[248] P. Hols, M. Kleerebezem, A. N. Schanck, T. Ferain, J. Hugenholtz, J. Delcour, and W. M. de Vos, "Conversion of *Lactococcus lactis* from homolactic to homoalanine fermentation through metabolic engineering," *Nature Biotechnology*, vol. 17, pp. 588–592, jun 1999.

[249] F. Levander, M. Svensson, and P. Rådström, "Enhanced exopolysaccharide production by metabolic engineering of *Streptococcus thermophilus*," *Applied and Environmental Microbiology*, 2002.

A

SUPPORT MATERIAL

A.1 GENOME ANNOTATION WORKFLOW

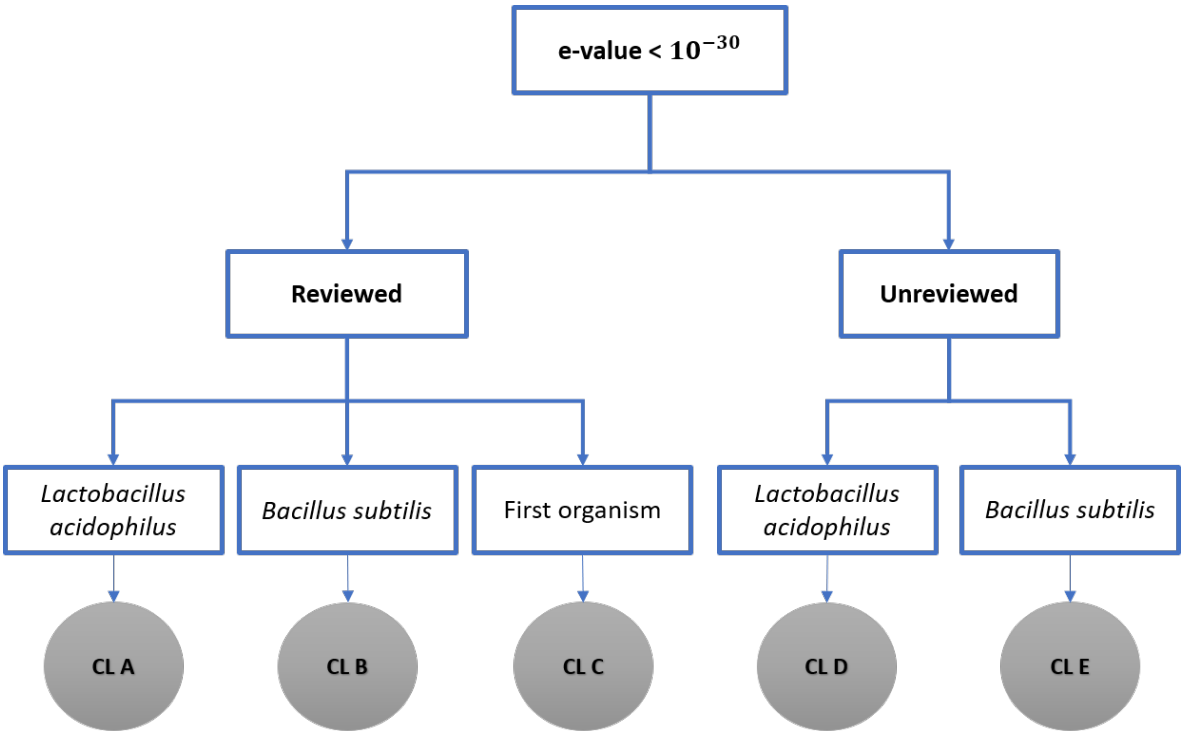


Figure S1.: Enzymes annotation workflow. CL: confidence level classification. Each manual enzymes annotation has a confidence level associated, from CL A (maximum confidence level) to CL E (minimum confidence level).

A.2 REMOVED PATHWAYS

Table S1.: Pathways removed from the model before the pathway-by-pathway analysis.

Pathway	**Pathway**
2-Oxocarboxylic acid metabolism	Inositol phosphate metabolism
alpha-Linolenic acid metabolism	Lysine degradation
Aminobenzoate degradation	Metabolic pathways
Ascorbate and aldarate metabolism	Metabolism of xenobiotics by cytochrome P450
Benzoate degradation	Microbial metabolism in diverse environments
Betalain biosynthesis	Monoterpenoid biosynthesis
Biosynthesis of amino acids	Naphthalene degradation
Biosynthesis of ansamycins	Nitrogen metabolism
Biosynthesis of antibiotics	Nitrotoluene degradation
Biosynthesis of secondary metabolites	Novobiocin biosynthesis
Biosynthesis of type II polyketide products	Penicillin and cephalosporin biosynthesis
Biosynthesis of unsaturated fatty acids	Pentose and glucuronate interconversions
Bisphenol degradation	Phenazine biosynthesis
C5-Branched dibasic acid metabolism	Phenylalanine metabolism
Caprolactam degradation	Phenylalanine, tyrosine and tryptophan biosynthesis
Carbon fixation in photosynthetic organisms	Phenylpropanoid biosynthesis
Carbon metabolism	Prodigiosin biosynthesis
Chlorocyclohexane and chlorobenzene degradation	Retinol metabolism
D-Alanine metabolism	Sphingolipid metabolism
Degradation of aromatic compounds	Steroid degradation
Dioxin degradation	Streptomycin biosynthesis
Drug metabolism - cytochrome P450	Synthesis and degradation of ketone bodies
Fatty acid degradation	Taurine and hypotaurine metabolism
Fatty acid elongation	Tropane, piperidine and pyridine alkaloid biosynthesis
Fatty acid metabolism	Tryptophan metabolism

Continued on next page

Table S1 – continued from previous page

Pathway	Pathway
Glucosinolate biosynthesis	Tyrosine metabolism
Glycosaminoglycan degradation	Ubiquinone and other terpenoid-quinone biosynthesis
Glycosphingolipid biosynthesis - ganglio series	Xylene degradation
Glycosphingolipid biosynthesis - globo and isoglobo series	Zeatin biosynthesis
Indole alkaloid biosynthesis	

A.3 GENERIC REACTIONS

Table S2.: Generic and glycan-associated reactions removed from the model.

Reaction ID	Reaction
R00135	Peptide + H2O→ L-Proline + Peptide
R00162	ATP + Protein→ ADP + Phosphoprotein
R00164	Phosphoprotein + H2O→ Protein + Orthophosphate
R00281	Acceptor + NADH + H+→ Reduced acceptor + NAD+
R00375	dATP + DNA→ Diphosphate + DNA
R00376	dGTP + DNA→ Diphosphate + DNA
R00377	dCTP + DNA→ Diphosphate + DNA
R00378	dTTP + DNA→ Diphosphate + DNA
R00379	Deoxynucleoside triphosphate + DNA(n)→ Diphosphate + DNA(n+1)
R00382	NAD+ + DNA(n) + 5′-Phospho-DNA(m) ⇌ AMP + Nicotinamide D-ribonucleotide + DNA(n+m)
R00435	ATP + RNA→ Diphosphate + RNA
R00437	RNA + Orthophosphate ⇌ RNA + ADP
R00441	GTP + RNA→ Diphosphate + RNA
R00442	CTP + RNA→ Diphosphate + RNA
R00443	UTP + RNA→ Diphosphate + RNA
R00444	Nucleoside triphosphate + RNA ⇌ Diphosphate + RNA
R00539	Acyl phosphate + H2O→ Carboxylate + Orthophosphate

Continued on next page

Table S2 – continued from previous page

Reaction ID	Reaction
R00600	S-Adenosyl-L-methionine + tRNA guanine ⇌ S-Adenosyl-L-homocysteine + tRNA containing N7-methylguanine
R00623	Primary alcohol + NAD+ ⇌ Aldehyde + NADH + H+
R00624	Secondary alcohol + NAD+ ⇌ Ketone + NADH + H+
R00804	Sugar phosphate + H2O→ Sugar + Orthophosphate
R00857	Glycerophosphodiester + H2O ⇌ Alcohol + sn-Glycerol 3-phosphate
R01122	Dimethylallyl diphosphate + tRNA ⇌ Diphosphate + tRNA containing 6-isopentenyladenosine
R01341	(S)-2-Hydroxyacid + Oxygen ⇌ 2-Oxo acid + Hydrogen peroxide
R01369	Triacylglycerol + H2O→ Diacylglycerol + Carboxylate
R01516	2,3-Bisphospho-D-glycerate + Protein histidine→ 3-Phospho-D-glycerate + Protein N(tau)-phospho-L-histidine
R02164	Quinone + Succinate ⇌ Hydroquinone + Fumarate
R02320	Nucleoside triphosphate + Pyruvate→ NDP + Phosphoenolpyruvate
R02584	ATP + Protein tyrosine→ ADP + Protein tyrosine phosphate
R02585	Protein tyrosine phosphate + H2O→ Protein tyrosine + Orthophosphate
R02628	Phosphoenolpyruvate + Protein histidine→ Pyruvate + Protein N(pi)-phospho-L-histidine
R02780	alpha,alpha-Trehalose + Protein N(pi)-phospho-L-histidine→ alpha,alpha'-Trehalose 6-phosphate + Protein histidine
R02806	Thymidine + Base ⇌ Deoxynucleoside + Thymine
R02961	S-Adenosyl-L-methionine + DNA adenine ⇌ S-Adenosyl-L-homocysteine + DNA 6-methylaminopurine
R03060	3'-Ribonucleotide + H2O ⇌ Ribonucleoside + Orthophosphate
R03232	Protein N(pi)-phospho-L-histidine + D-Fructose→ Protein histidine + D-Fructose 1-phosphate
R03259	3',5'-Cyclic nucleotide + H2O ⇌ Nucleoside 5'-phosphate
R03319	alpha-D-Hexose 1-phosphate ⇌ alpha-D-Hexose 6-phosphate
R03422	2',3'-Cyclic nucleotide + H2O ⇌ Nucleoside 2'-phosphate
R03423	2',3'-Cyclic nucleotide + H2O ⇌ Nucleoside 3'-phosphate
R03704	5,10-Methylenetetrahydrofolate + tRNA containing uridine at position 54 + FADH2 ⇌ Tetrahydrofolate + tRNA containing ribothymidine at position 54 + FAD
R03743	beta-Lactam + H2O ⇌ Substituted beta-amino acid
R03815	Dihydrolipoylprotein + NAD+ ⇌ Lipoylprotein + NADH + H+

Continued on next page

Table S2 – continued from previous page

Reaction ID	Reaction
R03910	Acetyl-CoA + Alkane-alpha,omega-diamine ⇌ CoA + N-Acetyldiamine
R03923	L-Cysteine + 'Activated' tRNA ⇌ L-Serine + tRNA containing a thionucleotide
R04111	Protein N(pi)-phospho-L-histidine + Maltose→ Protein histidine + Maltose 6'-phosphate
R04168	Deoxynucleoside + Base ⇌ Deoxynucleoside + Base
R04238	N-Substituted aminoacyl-tRNA + H2O ⇌ N-Substituted amino acid + tRNA
R04268	N-Formyl-L-methionylaminoacyl-tRNA + H2O ⇌ Formate + L-Methionylaminoacyl-tRNA
R04273	Peptidylproline (omega=180) ⇌ Peptidylproline (omega=0)
R04294	2'-Deoxyribonucleoside diphosphate + Thioredoxin disulfide + H2O ⇌ Ribonucleoside diphosphate + Thioredoxin
R04314	DNA containing 6-O-methylguanine + [Protein]-L-cysteine→ Protein S-methyl-L-cysteine + DNA containing guanine
R04315	Deoxynucleoside triphosphate + Thioredoxin disulfide + H2O ⇌ Ribonucleoside triphosphate + Thioredoxin
R04511	Phosphatidylglycerol + Membrane-derived-oligosaccharide D-glucose ⇌ 1,2-Diacyl-sn-glycerol + Membrane-derived-oligosaccharide 6-(glycerophospho)-D-glucose
R05209	Diacylglycerol + H2O ⇌ 1-Acylglycerol + Carboxylate
R05994	GA1 + H2O ⇌ GA2 + D-Galactose
R06034	Sucrose + Orthophosphate ⇌ D-Fructose + D-Glucose 1-phosphate
R06040	Maltose + Orthophosphate ⇌ D-Glucose + beta-D-Glucose 1-phosphate
R06049	1,4-alpha-D-Glucan(n) + ADP-glucose ⇌ 1,4-alpha-D-Glucan(n+1) + ADP
R06050	1,4-alpha-D-Glucan(n) + Orthophosphate ⇌ 1,4-alpha-D-Glucan(n-1) + D-Glucose 1-phosphate
R06070	Raffinose + H2O ⇌ D-Galactose + Sucrose
R06080	Isomaltose + H2O ⇌ alpha-D-Glucose + D-Glucose
R06084	Maltose + H2O ⇌ 2 alpha-D-Glucose
R06087	Sucrose + H2O ⇌ D-Fructose + D-Glucose
R06088	Sucrose + H2O ⇌ beta-D-Fructose + alpha-D-Glucose
R06091	Melibiose + H2O ⇌ D-Galactose + D-Glucose

Continued on next page

Table S2 – continued from previous page

Reaction ID	Reaction
R06093	Galactinol + H2O ⇌ myo-Inositol + D-Galactose
R06098	Lactose + H2O ⇌ alpha-D-Glucose + D-Galactose
R06100	Raffinose + H2O ⇌ Melibiose + D-Fructose
R06101	Stachyose + H2O ⇌ Manninotriose + D-Fructose
R06112	G10518 + H2O ⇌ D-Glucose + D-Glucose 6-phosphate
R06113	H2O + Trehalose 6-phosphate ⇌ D-Glucose + D-Glucose 6-phosphate
R06114	Lactose + H2O ⇌ D-Glucose + D-Galactose
R06115	H2O + Maltose 6′-phosphate ⇌ D-Glucose + D-Glucose 6-phosphate
R06131	alpha-Amino acid→ 2-Oxo acid + Ammonia
R06134	Amide + H2O ⇌ Carboxylate + Ammonia
R06142	Epimelibiose + H2O ⇌ D-Mannose + D-Galactose
R06152	Stachyose + H2O ⇌ Raffinose + D-Galactose
R06176	ATP + Cellobiose ⇌ ADP + G10518
R06185	Starch + Orthophosphate ⇌ 1,4-alpha-D-Glucan + D-Glucose 1-phosphate
R06186	1,4-alpha-D-Glucan ⇌ Starch
R06199	Starch(n+1) + H2O ⇌ alpha-D-Glucose + Starch(n)
R06202	G10534(n+1) + H2O ⇌ D-Galactose + G10534(n)
R06229	Trehalose + Protein N(pi)-phospho-L-histidine→ Trehalose 6-phosphate + Protein histidine
R06236	Protein N(pi)-phospho-L-histidine + Maltose→ Protein histidine + Maltose 6′-phosphate
R07180	ROOH + 2 Thiol-containing reductant ⇌ Alcohol + H2O + Oxidized thiol-containing reductant
R07297	5′-Ribonucleotide + H2O ⇌ Ribonucleoside + Orthophosphate
R07461	[Enzyme]-S-sulfanylcysteine + Adenylyl-[sulfur-carrier protein] + Reduced acceptor ⇌ AMP + Thiocarboxy-[sulfur-carrier protein] + [Enzyme]-cysteine + Acceptor
R07618	Enzyme N6-(dihydrolipoyl)lysine + NAD+ ⇌ Enzyme N6-(lipoyl)lysine + NADH + H+
R07807	G01977 + H2O ⇌ G13073 + D-Galactose
R08363	2′-Deoxyribonucleoside diphosphate + Tryparedoxin disulfide + H2O ⇌ Ribonucleoside diphosphate + Tryparedoxin
R08364	2′-Deoxyribonucleoside diphosphate + Trypanothione disulfide + H2O ⇌ Ribonucleoside diphosphate + Trypanothione
R08372	1-Hydroxyalkyl-sn-glycerol ⇌ Aldehyde + Glycerol

Continued on next page

Table S2 – continued from previous page

Reaction ID	Reaction
R08384	9-cis-Retinal + Oxygen + H2O ⇌ 9-cis-Retinoic acid + Hydrogen peroxide
R08550	Protein N6-(dihydrolipoyl)lysine + NAD+ ⇌ Protein N6-(lipoyl)lysine + NADH + H+
R08763	2 6-Iminocyclohexa-2,4-dienone + Acceptor ⇌ 2-Aminophenoxazin-3-one + Reduced acceptor
R09382	tRNA precursor + 2 CTP + ATP ⇌ tRNA with a 3′ CCA end + 3 Diphosphate
R09383	tRNA precursor + CTP ⇌ tRNA with a 3′ cytidine + Diphosphate
R09384	tRNA with a 3′ cytidine + CTP ⇌ tRNA with a 3′ CC end + Diphosphate
R09386	tRNA with a 3′ CC end + ATP ⇌ tRNA with a 3′ CCA end + Diphosphate
R09597	[tRNA(Ile2)]-cytidine34 + L-Lysine + ATP ⇌ [tRNA(Ile2)]-lysidine34 + AMP + Diphosphate + H2O
R09662	Reduced flavin + NAD+ ⇌ Flavin + NADH + H+
R10223	tRNA adenine + H2O ⇌ tRNA hypoxanthine + Ammonia
R10613	FAD + [Protein]-L-threonine ⇌ [Protein]-FMN-L-Threonine + AMP
R10648	L-Threonylcarbamoyladenylate + tRNA adenine ⇌ AMP + N6-L-Threonylcarbamoyladenine in tRNA
R10803	Hydrogen sulfide + 3 Coenzyme F420 + 3 H2O ⇌ Sulfite + 3 Reduced coenzyme F420
R10806	Protein glutamine + S-Adenosyl-L-methionine ⇌ Protein N5-methyl-L-glutamine + S-Adenosyl-L-homocysteine
R11029	2′-Deoxy-5-hydroxymethylcytidine-5′-triphosphate ⇌ DNA 5-hydroxymethylcytosine
R11528	L-Cysteine + [Protein]-L-cysteine ⇌ L-Alanine + [Protein]-S-sulfanyl-L-cysteine
R11529	[Enzyme]-S-sulfanylcysteine + [Protein]-L-cysteine ⇌ [Enzyme]-cysteine + [Protein]-S-sulfanyl-L-cysteine
R11592	Ribonucleoside triphosphate + Formate ⇌ Deoxynucleoside triphosphate + CO2 + H2O
R11865	2-(Hydroxysulfanyl)hercynine + Reduced acceptor ⇌ Ergothioneine + Acceptor + H2O

A.4 CHEMICALLY DEFINED MEDIUM

Table S3.: *In silico* chemically defined medium for *L. acidophilus*, defined according to [123]. The lower bound of the respective exchange reaction was left unconstrained, except for glucose, whose lower bound was settled as 16.4 $mmol\, h^{-1}\, gDW^{-1}$

Metabolite	**Metabolite**	**Metabolite**	**Metabolite**
Glucose	L-Isoleucine	L-Valine	Pantothenate
Acetate	L-Leucine	Guanine[a]	Spermidine
Orthophosphate	L-Lysine[a]	Uracil[a]	Folate
D/L-Alanine[a]	L-Methionine	AMP[a]	Tetradecanoic acid
L-Arginine	L-Phenylalanine	CMP[a]	Hexadecanoic acid
L-Aspartate	L-Proline	Deoxyguanosine[a]	Hexadecenoic acid
L-Cysteine[b]	L-Serine	Thymine[a]	Octadecanoic acid
L-Glutamate	L-Threonine[b]	Riboflavin	Octadecenoic acid
Glycine[a]	L-Tryptophan	Pyridoxal	
L-Histidine	L-Tyrosine	Nicotinamide	

[a] These metabolites are not essential, but are included in rich media to stimulate growth;
[b] Cysteine and threonine were considered essential according to [123], but are not essential in *in silico* simulations.

A.5 FERMENTATION PATTERN

Table S4.: Comparison between experimental fermentation growth from different carbohydrates of *L. acidophilus* and *in silico* simulations.

Carbohydrate	*Wheater et al* [217]	*Axelsson et al* [218]	Model (This work)	AGORA model
Glucose	+	ND	+	+
Fructose	+	+	+	+
Galactose	+	+-	+	+
Lactose	+	+	+	+
Cellobiose	+	+	+	+
N-Acetyl-Glucosamine	ND	+	+	+
Amygdalin	+	+	-	-
Maltose	+-	ND	+	+
Stachyose	ND	ND	+	+
Mannose	+	+	+	+
Trehalose	+-	+-	+	+
Salicin	+	+	+	+
Sucrose	+	ND	+	+
Melibiose	+-	+-	+	-
Raffinose	+-	+	+	-
Melezitose	-	-	-	-
Mannitol	-	ND	-	-
Arabinose	-	-	-	-
Rhamnose	-	ND	-	-
Ribose	ND	-	+	+
Xylose	ND	-	-	-

+ Growth; - Non growth; +- Strain specific growth; ND, Not determined.