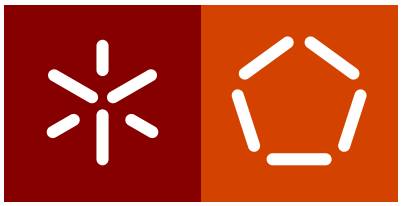
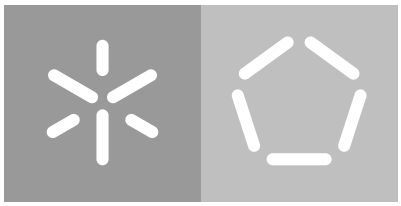


Universidade do Minho
Escola de Engenharia
Departamento de Informática

Catarina Moreira Ribeiro

Assessment of microbial community interactions using different tools

March 2020



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Catarina Moreira Ribeiro

Assessment of microbial community interactions using different tools

Master dissertation
Master Degree in Bioinformatics

Dissertation supervised by
Oscar Dias
Ahmad Zeidan

March 2020

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

ACKNOWLEDGEMENTS

First of all, I would like to begin by saying that during this work I faced several challenges, which allowed me to develop myself, both on a professional and personal level and I am very grateful for everyone who took part in this chapter of my life.

To begin with, I would like express my gratitude towards my supervisors, for giving me the opportunity to study such an emergent field in systems biology. Specifically, I would like to thank Doctor Oscar Dias for introducing me to the Systems Biology field, for the availability to answer all questions and for all the given help through this work. Also, I would like to thank Doctor Ahmad Zeidan, for all the help during my stay in Denmark, for always helping me to face the challenges this work presented and for making me feel very welcome at Chr. Hansen A/S.

I would also like to acknowledge everyone I met in Denmark. Thank you to everyone at the company for creating such a fantastic work environment which made my experience much more special, I cannot imagine a better place to work at. Thank you to my roomie, Helle, for receiving me at your house, the *higgiest* place in Copenhagen!

The biggest thank you to my parents, Ílidia and José, without whom I would certainly not be here. Thank you for all the encouragement through my years of studying. Thank you to my sister, Marta, who is one of my biggest supporters and always knows what needs to be said.

A huge thank you to all my friends, specially to Cata, Chica, Costa, Dani, Diana, Mesquita and Nini, who grew up with me and were there through all moments of my life, including this one.

Last but not least, I have to thank my boyfriend Miguel, who was always there when I needed him, at any time of any day! Thank you for all the love and support, for all the laughs and for cheering me up when I need to. Thank you for your calming words when I am stressed up. This accomplishment would not have been possible without you.

STATEMENT OF INTEGRITY

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

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

ABSTRACT

Microbial communities participate in many biological processes, directly affecting its surrounding environment. Thus, the study of a community's behaviour and interactions among its members can be very useful in the biotechnology, environmental and human health fields. Nevertheless, decoding the metabolic exchanges between microorganisms and community dynamics remains a challenge.

Computational modelling methods have gained interest as a way to unravel the interactions and behaviour. GSM models allow the prediction of an organism's response to changes in genetic and environmental conditions. Thus, the extension of such method to a community level can help decode a community's phenotype.

In this work, different GSM models and current bioinformatics tools were used to model the metabolism of different microbial communities. The different tools' performances were compared to assess which is currently the best method to perform an analysis on a community level. Distinct case studies regarding microbial communities for which its interactions were already known, were selected. To assess the tools' performances, each tools output was compared to what was expected in theory.

COBRA Toolbox's methods proved to be useful to build a community structure from individual GSM models, while pFBA and SteadyCom's simulation methods can predict exchange between the organisms and the environment. Additionally, *Dynamic Flux Balance Analysis (dFBA)* approaches, such as DFBAlab and DyMMM, can successfully simulate metabolite and biomass variation over time. Nevertheless, these methods are more limited as they require specific organism information, which is not always available.

Several GSM models are available for use. Nonetheless, their quality control has to gain attention as the simulations' results are directly affected by the individual models accuracy to represent an organism's metabolism. Thus, community model builders should carefully chose a GSM model, or combination of models before performing simulations.

Keywords: Microbial Community; Systems Biology; Genome-scale Metabolic Model; Community Modelling

RESUMO

Comunidades microbianas participam em inúmeros processos biológicos, afetando diretamente o ambiente que as engloba. Assim, o estudo do comportamento de uma comunidade e interações entre os seus membros pode ser muito útil nas áreas da biotecnologia, ambiente e saúde. No entanto, descodificar as trocas entre microganismos e a dinâmica de comunidades continua um desafio.

Métodos de modelação computacional têm ganho interesse como forma de desvendar tais interações e comportamento de comunidades. Modelos metabólicos à escala genómica permitem prever a resposta de um certo organismo a mudanças genéticas e ambientais. Assim, a extensão de tal método ao nível de comunidade pode ajudar a prever o fenótipo de uma certa comunidade.

No presente trabalho, diferentes modelos metabólicos à escala genómica e ferramentas bioinformáticas foram utilizados para modelar o metabolismo de diferentes comunidades microbianas, comparando o desempenho destas ferramentas para avaliar qual o melhor método para análise ao nível da comunidade. Casos de estudo distintos, relativos a comunidades para as quais se conhecem as interações, foram selecionados. Por fim, para aferir o desempenho das ferramentas, os respetivos resultados foram comparados ao teoricamente esperado.

Os métodos da ferramenta COBRA Toolbox provaram ser úteis para construir a estrutura da comunidade, usando modelos metabólicos à escala genómica dos organismos individuais. Quanto a métodos de simulação, pFBA e SteadyCom são úteis para prever trocas entre os organismos e o ambiente que os envolve. Para além disso, abordagens dFBA, como DFBAlab e DyMMM, podem simular a variação da concentração de metabolitos e biomassa ao longo do tempo. No entanto, estes métodos apresentam limitações por requererem informação específica ao organismo, que nem sempre se encontra disponível.

Vários modelos metabólicos à escala genómica estão disponibilizados. No entanto, o controlo na qualidade destes tem que ganhar atenção, visto que os resultados das simulações são diretamente afetados pela sua precisão na representação do metabolismo de um organismo e consequentemente, da comunidade. Assim, para construir um modelo de comunidades, é necessária uma seleção cuidadosa dos modelos individuais a usar, antes de serem feitas simulações.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

B

BIGG Biochemical, Genetic and Genomic.

C

CFBA Community Flux Balance Analysis.

COA Coenzyme A.

COBRA Constraint-Based Reconstruction and Analysis.

D

DFBA Dynamic Flux Balance Analysis.

DFBALAB Dynamic Flux Balance Analysis laboratory.

DYMMM Dynamic Multispecies Metabolic Modeling.

E

EC Enzyme Commission.

F

FBA Flux Balance Analysis.

FVA Flux Variability Analysis.

G

GSM Genome-scale Metabolic.

I

IBM Individual-based Model.

ID Identifier.

K

KEGG Kyoto Encyclopedia of Genes and Genomes.

L

LDH Lactate Dehydrogenase.

LP Linear Programming.

M

MIP Metabolic Interaction Potential.

MRO Metabolic Resource Overlap.

MTHFC Methenyltetrahydrofolate Cyclohydrolase.

O

ODE Ordinary Differential Equation.

P

PFBA parsimonious Flux Balance Analysis.

PFL Pyruvate-formate Lyase.

PLM Population-level Model.

S

SBML Systems Biology Markup Language.

SMETANA Species Metabolic Interaction Analysis.

1

INTRODUCTION

This dissertation was developed in the context of the Masters in Bioinformatics, assured by Centre of Biological Engineering and Department of Informatics at University of Minho, and Chr. Hansen A/S.

The context and motivation, goals and the structure of the current work are described in the present chapter.

1.1 CONTEXT AND MOTIVATION

Microbial environments are commonly described by microbial communities and their interactions (99). Understanding such interactions can be useful for several practical applications (59). For instance, food fermentation processes are not usually carried out by a single strain, but a complex mixture of lactic acid bacteria (89). Thus, characteristics such as flavor or texture of the final food product are directly affected by culture functions and the metabolic interactions of the community members (71), making the study of such aspects of extreme economical value.

The emergence of high-throughput technologies led to a fast increase of fully sequenced genomes and systems biology assumes nowadays a leading role in biological sciences, as there is a clear need to decode all the information available (11; 78).

GSM network reconstructions can be obtained by combining genome sequence information with biochemical knowledge, resulting in a model containing detailed descriptions of all biochemical reactions, metabolites and genes for a specific organism, acting as a tool to predict an organism's response to changes in genetic and environmental conditions (78). This approach is gaining attention as a promising way to understand interactions in microbial communities, whether competitive, mutualistic, commensal, or others.

Microbial interactions depend on the potential of each species and the nutrients available (8; 57; 99). Thus, the usage of systems biology approaches allows the understanding of how different species interact and affect the environment surrounding them.

Different tools and methods have been developed to aid the study of such microbial communities. Thus, frameworks such as SteadyCom (8), for predicting a stable commu-

nity composition; OptCom (99), a comprehensive *Flux Balance Analysis (FBA)* framework for microbial communities; modelling approaches such as SMETANA (95); and dynamic modelling frameworks such as DFBAlab (26) and DyMMM (97) can be helpful for understanding the interaction between modeled species and the dynamics of a community.

In this project, we will use GSM models and apply community modeling techniques to simulate and study interspecies interactions.

1.2 GOALS

The main goal of this work is analyzing different tools' performances in regards to the prediction of interactions between different species, using genome-scale metabolic models.

Thus, the process to achieve such goal is described by the following steps:

- Selection of different case studies regarding known interspecies interactions;
- Creation of the community models using different tools;
- Simulation of interactions using different methods;
- Analysis of the obtained results;
- Performance assessment of the different tools.

1.3 STRUCTURE OF THE DOCUMENT

This document has the following structure:

- **Chapter 2:** State-of-the-art
  - Description of the GSM model reconstruction process
  - Overview of methodologies used in microbial communities analysis
  - Single organism and community based computational tools
- **Chapter 3:** Materials and Methods
  - Case studies and respective GSM models;
  - Software and tools;
  - Definition of case studies' environmental conditions;
  - Analysis using different tools;
  - Comparison of the different tools performances.
- **Chapter 4:** Results and Discussion

  - Results obtained regarding the different community models' structures, for each case study;
  - The different tools' performance, for each case study.

- **Chapter 5:** Conclusion
  - The best method to build a community model;
  - The best method to run simulations for community models;
  - Future work;
  - Final remarks regarding the state of the field of study.

2

STATE OF THE ART

The following chapter provides state-of-the-art methods on microbial community analyses. Specifically, the process to reconstruct an individual GSM model and a microbial community model and individual and community-based computational tools are described.

2.1 BACKGROUND

Systems biology encompasses the quantification of cell components and analyses the interactions between them. Thus, it aims at predicting the whole-cell behavior through computer simulations of biochemical models (11).

Whole-genome high-throughput sequencing techniques allowed the whole-genome sequencing of several organisms. From the genome sequence it is possible to identify gene products involved in biological processes. Thus, genomics arised as a field in which gene functions are determined (11).

As a genome does not change significantly over time, other *omics* emerged in the effort to identify and characterize additional cellular components, which regulate the expression levels of genes. These include proteomics, transcriptomics and metabolomics, which provide information regarding the physiological state of the system (11).

Genome-scale metabolic models have become an important tool in systems biology, as these allow decoding information of the increasingly higher number of fully sequenced genomes. *Haemophilus influenzae* was the first organism to have its complete genome sequenced, and one of the first genome-scale reconstructions in systems biology as well (18).

Over time, progress has been achieved and modelling has transitioned from highly characterized organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae*, to species with less information and characterization, which are of extreme importance to specific applications. As an example, study of pathogens allows the development of new strategies in metabolic engineering and facilitates patient treatment (47).

Efforts in automating the model reconstruction process have been made, as it is faster to obtain a new genome sequence than reconstructing its GSM model. These advancements,

in bioinformatics tools and algorithms, allowed the reconstruction of GSM models of less-characterized organisms (47).

Recently, the Assembly of Gut Organisms through Reconstruction and analysis (AGORA) resource presented 773 genome-scale metabolic reconstructions for human gut microbes (57). Approaches such as these, are a starting point of high-quality reconstructions.

It is well known that bacteria and microorganisms cohabit in dense, surface-associated communities with several species and strains (62; 94). Thus, members of the consortium affect each other by establishing different metabolic interactions (84). Such interactions play an important role in biotechnology, environment, human health and food industry as they have applications with great commercial value (27).

An effort to extend individual modeling techniques to community-level modeling has been made (94). In general, a model of metabolic interactions of microbes in co-culture accounts the metabolites that can be exchanged (91). Thus, constraint-based stoichiometric modelling in microbial communities is more complex than single organism modelling as both the exchange of metabolites and the biomass abundance of each organism have to be considered. Additionally, there are some challenges such as isolating individual members from the consortium and having poorly characterized species (37; 46).

There are few studies combining experimental and microbial communities metabolic models data (30). Stolyar and coworkers, performed the first study regarding a multispecies metabolic model and compared the results to experimental data on growth of a co-culture of methanogenic bacteria (88). To the date, purely computational studies have been conducted to determine the potential of a community's interactions, calculate biomass ratios and fluxes under balanced growth conditions of microbial communities (46) or designed medium compositions that enforce metabolic interactions (49).

In food fermentation, community-level modelling provides phenotype predictions of individual-strain and the community as a whole. Furthermore, community models' ability to identify metabolic interactions between the community members is its most significant benefit (94). As an example, flavor compounds production can be optimized.

Computational studies can be used to investigate the potential interactions in a community, such as competition, cross-feeding, syntrophy and mutualism (37; 88). Originally, the primary focus of such studies was mostly the prediction of a community phenotype. Currently, these are used to infer metabolic exchange fluxes between microorganisms from experimental data (30). For instance, the characterization of interactions between gut microbiota and their interactions with diet and their host is important application of community-level models (94).

Community interactions are quantified by combining experimental data, regarding metabolite concentrations and specific growth rates, with stoichiometric metabolic models (30).

Although the medium composition determines the growth potential of an organism, on a community level, other members' metabolic capabilities will affect the metabolism and growth of another particular species (27). There are three categories of effects that organisms cause on each other: positive, negative and neutral. These can then be classified as different interactions: mutualism, parasitism, competition, commensalism, amensalism and neutralism (83), as shown in Table 1 .

Table 1: The six main effector and target microorganism interactions.

Effect on target	**Effect on effector**		
	Beneficial	**Detrimental**	**Neutral**
Beneficial	Mutualism		Commensalism
Detrimental	Parasitism	Competition	Amensalism
Neutral			Neutralism

Mutualism is a common interaction in microbial communities where two or more species provide benefit to one another, creating a dependency (83; 88). It is important to note that, although similar, it is different from commensalism, in which one organism benefits from the interaction but the other is not particularly affected. On the opposite side, parasitism arises from one species benefiting at the expense of another (84).

Amensalism is when an organism negatively affects the other without being affected itself. Such can happen when products from primary metabolism (e.g. alcohols) inhibit the growth of the other consortium individuals (84).

When both organisms suffer a negative effect from the interaction, there is competition. In fermentation, it is common to find microbes competing for energy sources and nutrients (e.g. competing for free amino acids in milk fermentation) (84).

Finally, when each organism can sustain growth without cross-feeding, the interaction is called neutralism. This kind of interaction can potentially lead to competition since the same resources could be consumed by both species (49; 61).

When studying microbial communities, it is important to consider the phylogenetic distance between the interacting species (62; 95). For instance, considering a clonal group of cells, cooperative behavior is expected. Natural selection is expected to favor phenotypes that maximize the overall survival and reproduction of the genotype. However, cooperation is not guaranteed as loss-of-function mutations can occur at the loci driving the cooperative trait (62).

In Zeidan et al. (2010), an altruistic cooperation interaction was proposed. This is characterized by an individual species passing on its genes to the next generation indirectly, by aiding a closely related species to reproduce. Altruistic traits are visible in cases in which a specific cell is harmed to promote reproduction of other group members (62). However,

the precise mechanism for this interaction and the nature of the involved molecules, are yet not fully understood (62; 94).

Among different genotypes, the ecology of a system should also be taken into consideration. For instance, the overlap in metabolite intake in the niche indicates the degree of competition (95). Thus, if the number of metabolites that both organisms consume simultaneously is high, it creates an overlap that implies strong competition for nutritional resources. Additionally, when a certain species is harmful to another, by consuming its nutritional resources, the second might increase competitiveness in return (61).

2.2 GENOME-SCALE METABOLIC MODELS

The development of the high-throughput technologies generated bulk loads of information, which facilitates the study of cells metabolism (63). With the availability of whole-genome sequences and information regarding biochemical reactions in several biological databases, GSM networks can be generated (11). These networks represent a set of biological reactions retrieved from enzymes encoded in the organisms genome. Thus, GSM networks describe reactions and the relations between them.

Even though GSM networks can represent some of the cell's physiological and biochemical properties, GSM models allow predicting the metabolic capabilities of the biological system. These models include energetic needs and biomass composition specificities, in addition to the network data (11). However, it is important to note that these networks do not consider enzyme kinetic information (46; 65). Kinetic modelling is used on a pathway scale rather than a genome scale. It models reaction fluxes and metabolite concentration as a function of time, thus requiring more information (85). As there is a lack of such information for the majority of the sequenced organisms, the usage of these models at a genome-scale has been held back (74).

GSM models are widely used in biotechnology and medicine (46; 56) as these contain most metabolic reactions of an organism, associated to genes encoding enzymes that catalyze such reactions, thus proving to be very powerful tools. As a result, these models are currently used, *in silico*, to identify possible drug targets and predict the microorganisms response to gene knockouts or nutritional changes (11).

An increase in the number of available GSM models is visible, as a consequence of the availability of tools that automate such process. Nevertheless, the lack of quality in such models can compromise its usage. Thus, the quality and accurate predictions should be priorities, and efforts in this topic are being performed (47).

2.2.1 *Genome-scale metabolic models reconstruction*

Several works, representing the bottom-up and top-down approaches to reconstruct a metabolic model, have been published (11), the first being the most common (56).

The bottom-up reconstruction approach begins by associating metabolic functions to the respective genes through means of genome annotation. From there on it is necessary to collect a set of biochemical reactions from a reaction database, such as *Kyoto Encyclopedia of Genes and Genomes (KEGG)* (44). As a result, a draft metabolic network is assembled and subsequently improved by manual curation procedures (56).

As previously stated, there is the alternative top-down approach. It begins by reconstructing a universal model, which is manually curated and includes import/export reactions and a universal biomass equation. Then, the universal model is adapted to an organism, by removing reactions and metabolites that are not predicted to be in the organism (56).

The process of reconstructing a model (Figure 1) is a laborious task. Four main stages constitute the process herein described, specifically genome annotation, metabolic network assembly, conversion of the network to a stoichiometric model and the validation of the metabolic model (11).

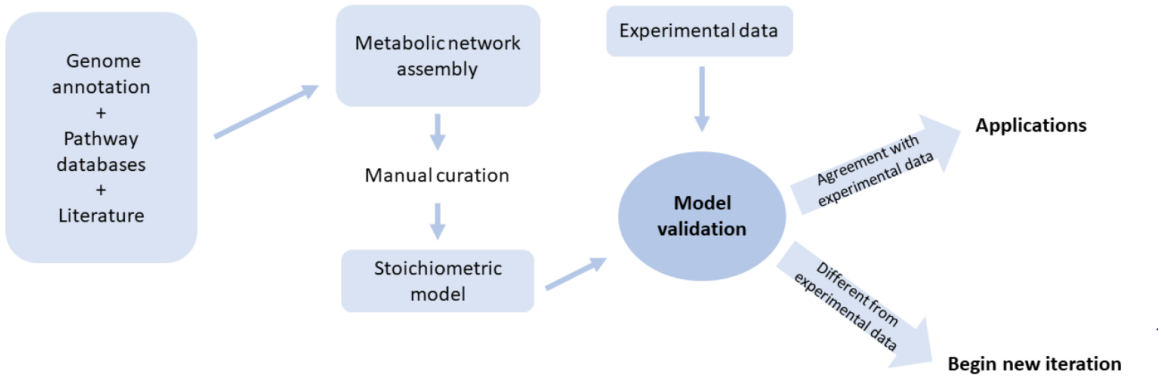


Figure 1: The reconstruction of a metabolic network is an iterative process. It begins by assessing information about the organism metabolism. Then, the reaction set is constructed manually curated and a stoichiometric model arises. If its *in silico* predictions are in agreement with experimental values, the model is then applied to several ends. If *in silico* and experimental results are in disagreement, another iteration begins, where the process is repeated until results are satisfactory. Adapted from *Rocha I., Förster J. and Nielsen* (74)

2.2.1.1 *Genome Annotation*

The start of a GSM model reconstruction consists in the genome annotation, which is the assignment of functions to genes (90). Whole genome annotations can be downloaded

from online databases such as NCBI (77) or KEGG (44). However, such annotations must be reliable and updated. When annotations are not available, sequence alignment algorithms such as Basic Local Alignement Search Tool (BLAST) (1) or HMMER (17) should be used.

Notice that genes encoding enzymes or transport systems are labeled as metabolic genes and subunits of protein complexes should be identified, as several genes may be required to encode an enzyme. Enzymes' information can be retrieved from curated databases such as BRENDA (42), where *Enzyme Commission (EC)* numbers and other metabolic items can be retrieved. Gene ontology (GO) information can also be used to find enzymes (90). Finally, biochemical reaction databases, such as KEGG (44), can be used to combine EC numbers with metabolic reactions.

After collecting the candidate metabolic genes and their potential reactions, the draft reconstruction can be assembled (90).

2.2.1.2 *Assembling the Metabolic Network*

This is the longest part of the reconstruction, due to extensive manual curation. Genes and reactions should be individually analysed and compared with organism-specific data, as organism-unspecific reactions can affect the predictive capacity of the model. If such information is not available in literature, phylogenetic close organisms can be considered.

Firstly, the network assembly starts by collecting information regarding transport systems and reactions catalyzed by enzymes encoded in the annotated genome. Genes-protein-reactions (GPR) associations are performed by analyzing literature and biological databases. Additionally, spontaneous and non-enzymatic reactions, available in KEGG or published literature, should be included in the network. After collecting the reactions, stoichiometry should be reexamined.

Subsequently, network reactions should be compartmentalized, assigning reactions to the organelles where these take place. Based on amino acid sequences and physiological characteristics of the organism, tools such as PSORTb (93), TargetP (20), LocTree3 (25), among others, can be used to predict subcellular localization.

Note that organism complexity affects results, as prokaryotic organisms are usually characterized by extracellular space, cytoplasm and periplasmatic space, whereas eukaryotic cells are complex and have several organelles. In higher eukaryotes, different tissues may also be considered.

Finally, as automatic methods are prone to missing information and to generate errors, manual curation should be performed. Reactions should be revised to correctly represent organisms specificities in literature and organism-specific databases. Likewise, reactions reversibility should also be assessed, and can be estimated from the Gibbs free energy of formation and of reaction (24; 90). Manual curation also addresses the following issues:

- Misrepresentation of substrate/cofactor usage;

- Adding new organism-specific reactions;
- Assessing the reactions assigned to ambiguous identifiers, such as incomplete EC numbers (28).

Comparative genomics can be used to accelerate the process of reconstructing a GSM model. The draft network is compared with curated models from closely related organisms, known biological pathways or a combination of both. Moreover, gap finding and gap-filling of missing reactions in the network has to be performed. These gaps refer to missing reactions which would result in the accumulation of produced products and in the arrest of the pathway's flux (11).

Additionally, organisms may differ in cofactors usage and directionality of certain reactions. Hence, organism-specific biochemical studies and organism-specific databases such as EcoCyc (45) can be used to find such information.

2.2.1.3 *Conversion from Metabolic Network to Stoichiometric Model*

This stage starts with the conversion of the network into a stoichiometric matrix. The coefficients of each reaction are represented in a matrix and constraints representing reaction limits are added to the model (11; 65).

A biomass formation equation should be included in the GSM model's reactions set. This equation, which represents all macromolecules and building blocks that establish the biomass, is necessary to simulate the growth rate (h^{-1}) of the desired organism and can be represented as below:

$$\sum_{k=1}^{p} c_k.X_k \rightarrow biomass \tag{1}$$

Where c_k represents the coefficient of the metabolite X_k.

Such reaction is important for *in silico* simulations and its accuracy depends on retrieving information specific to the organism. Thus, the biomass composition should be determined experimentally (4; 38; 41), or estimated using organism-specific literature. When such information is unavailable, phylogenetically close organisms can be used (11).

The model is not complete without including energy requirements, both growth or non-growth associated (90). The first covers several cellular processes, for instance amino acid, protein and nucleotides polymerization. The latter, refers to non-growth ATP requirements of the cell to survive (90).

Chemical engineering principles are used to represent the behavior of metabolites concentration. Thus, each metabolite has a respective *Ordinary Differential Equation (ODE)* within the metabolic network, in which its stoichiometry along the reactions set is considered (11; 74).

The following equation represents the variation of a metabolite concentration over time:

$$\frac{dX_i}{dt} = \sum_{j=0}^{N} S_{ij}.v_j + \mu X, i = 1, ..., M \quad (2)$$

Where X_i is the metabolite i concentration, v_j is the metabolic flux of reaction j, S_{ij} is the stoichiometric coefficient of metabolite i in reaction j and μX is the system's growth rate.

As there is an impossibility to collect and estimate kinetic rates and parameters from an organism's genome, a steady-state approach is used. Accordingly, mass balance constraints are applied, in which the total amount of any compound being produced equals the consumption, resulting in the following equation:

$$S.v = 0 \quad (3)$$

where $\boldsymbol{S}$ is the sparse matrix where each row represents an unique metabolite and each column a reaction, and $\boldsymbol{v}$ is a vector that represents the flux through all the reactions.

As most metabolic networks are undetermined systems and the number of fluxes is greater than the number of mass balance constraints, an infinite number of solutions may satisfy the mass balance constraints. Therefore, the establishment of these constraints can reduce the null space of S to a series of solutions, the flux cone of solutions.

Furthermore, reversibility and directionality define the stoichiometric matrix's constraints. Hence, upper and lower bounds are set as model constraints for specific reactions as they impose a maximum or minimum desirable reaction fluxes (65; 70). The following inequality represents these constraints:

$$\beta_j \geq v_j \geq \alpha_j, j = 1, ...N \quad (4)$$

Where

- b_j is the upper bound
- a_j is the lower bound
- v_j is the flux vector

Finally, the resulting representation of the mathematical model should be exported in a standard universal format such as the SBML (40).

2.2.1.4 *Metabolic Model Validation*

The last stage of the reconstruction relies on comparison between experimental data and the *in silico* prediction. Thus, the assessment of information regarding physiology, biochemistry and genetics of the target organism is necessary to improve the model's predictive capabilities.

This is an iterative process and when predictions and experimental values are not in agreement with experimental results, the stoichiometric model has to be reevaluated (74). As different environmental conditions are tested, the need for more iterations arises.

Constraint-based analyses of metabolic networks have become popular for simulations regarding cellular metabolism. FBA (65) is the dominant approach for flux estimation (70; 34). It uses linear optimization by maximizing or minimizing an objective function to determine the steady-state reaction flux distribution in the metabolic network, providing one optimal solution (70). Maximization of biomass formation is the most common objective function. However, different simulations can be performed (e.g. maximizing/minimizing a target compound production).

The following expressions illustrate the FBA formulation:

$$maximize/minimize \rightarrow Z$$

$$subject\ to \rightarrow S.v = 0$$

$$\beta_j \geq v_j \geq \alpha_j, \text{ j = 1, ... N}$$

Where

- Z is the linear objective function to be maximized or minimized
- S is the sparse stoichiometric matrix
- v is the flux vector
- β_j and α_j are the upper and lower bounds, respectively.

This method can be used in several reconstruction steps, such as model refinement, gap filling and energy requirements fine-tuning.

After a model is validated, one can determine how robust a metabolic model is in different simulation conditions, using *Flux Variability Analysis (FVA)* (29). Minimum and maximum fluxes for reactions in the network, while maintaining a specific state, can be determined using this method.

Among the common applications of FVA is the studying of flux distributions under suboptimal growth, optimization of process feed formulation for production of desired products, understanding networks flexibility and optimal strain design procedures.

2.2.2 *Microbial community models*

Mathematical models provide information impossible to acquire with experimental data alone (35). Models consolidate experimental knowledge, underlying assumptions, and expose gaps and inconsistencies in the knowledge. Once a model is finished, it can be used

to predict a system's response to conditions that have not yet been tested experimentally or predict system properties that are not directly observed.

Identically to single organisms, communities might be modeled by using top-down or bottom-up approaches (6). Thereby, definitions such as *Population-level Model (PLM)* and *Individual-based Model (IBM)* arised (6; 35).

2.2.2.1 *Population-level Models*

The traditional approach to modelling a microbial consortium is based on PLMs. These represent a top-down approach and are simple models that provide general explanations. The biological variability among individual cells is not considered and all cells are grouped together and represent one single macroscopic variable (86).

For example, these can be used to understand predator-prey interactions and generate oscillatory population dynamics. When using differential equations, a PLM is applied to spatially homogeneous environments and when using partial differential equations these describe spatially structured environments. However, there are limitations regarding these models. As previously mentioned, the lack of information describing the state and behavior of individual microorganisms can lead to incorrect predictions (35; 86).

2.2.2.2 *Individual-based Models*

Alternatively, IBM uses individual cells as modelling unit (86). It models the properties, activities and interactions of individuals in the population. Important examples of such are:

- Biomass of an individual species
- Uptake of substrates from the environment
- Competitive, synergistic or parasitic interactions between individuals

IBMs use single cell data resulting in individual-level information. Consequently, the collective information results in population-level data (35). Thus, IBMs incorporate physical, chemical and biological information at the individual bacterium level to produce a population-level output (86). Since IBMs account for individual organism and its interactions with the environment, the natural variability is simulated (23; 86).

On a final note, IBMs exhibit high computational demands (86) and alternative approaches are also used (e.g. combining individuals in a population in a super-individual (92)).

2.2.2.3 *Microbial Communities Design*

FBA simulations represent an important part in reconstructing a high-quality GSM model (27; 65; 74). Furthermore, FBA has applications in GSM modeling in microbial communities. Three different approaches in the usage of this method have been proposed (31). They vary in complexity and the way individual species are handled (31). There is a steady-state compartmentalized approach (5; 88), a supra-organism approach (46; 76), and a dynamic compartmentalized approach, based on dFBA (37). Alternatively, these frameworks are also referred to as species compartmentalization, mixed-bag modeling and multi-species dynamic modeling, respectively (36; 71).

The compartmentalized approach is the most frequently used when studying microbial interactions (5). It assigns a compartment to each organism, and creates a shared compartment for metabolite exchanges between them. With this new common compartment, the original external metabolites become internal ones in the community. Thus, they must be balanced following the mass balance equation (Eq. 3). The individual species' exchange metabolites are combined so that they only exist once in the new community model.

In the individual organisms' models, the specific growth rates' unit was referred as h^{-1}. Nevertheless, on a community level, the unit was changed to $gDW_i/gDW_c/h$, where i and c represent an individual and the community, as these are biomass synthesis reactions which produce the species' into the shared compartment. On the same note, a community growth reaction is introduced, resulting from the different synthesized biomass to the community medium. The rate of this reaction represents the community growth rate (h^{-1}).

The resulting community's structure facilitates the comprehension of species interactions and provides a detailed insight on pathogen or mutualistic interactions, for example (46). The shared byproducts and exchange reactions flow through the shared compartment. For instance, Stolyar and coworkers, modeled a microbial community as a multi-compartment metabolic network and used FBA to characterize community-level fluxes. Optimal growth rate and metabolite fluxes were estimated and the biomass functions for each species were combined and weighted (88).

Rodríguez and coworkers (76) used the supra-organism approach for the first time. All metabolic reactions from the community's individuals were merged, resulting in a sole, larger, reaction set. Thus, species are not segregated and its boundaries are not considered. Afterwards, it was used to study the metabolic capabilities in terms of product and substrate variation of the community (5). Recently, biomass concentrations were successfully integrated in the compartmentalization approach (46).

Finally, the dynamic compartmentalized approach (37) implements dFBA through the usage of substrate uptake kinetics and cross-feeding between organisms. Instead of using specific substrate uptake values, dFBA calculates the time-varying consumption and pro-

duction of metabolites, changes in biomass, and effects of the metabolism in the community environment (5).

2.3 COMPUTATIONAL TOOLS

Along with the development of high-throughput sequencing, techniques such as metagenomics and meta-transcriptomics facilitate the access to genomic data (99). Despite the increasing amount of information, there is still a need to understand individual microorganisms contributions to the whole community, resulting in the need to develop modelling frameworks and approaches to elucidate such aspects (8; 99).

2.3.1 *Single Organism Based Tools*

2.3.1.1 *merlin*

Metabolic Models Reconstruction Using Genome-Scale Information (*merlin*) (12) is an in-house developed tool. It is user-friendly and is divided in two modules: annotation module and models reconstruction module. It aids in all GSM model reconstruction stages for a single-species model.

2.3.1.2 *COBRA Toolbox and COBRApy*

COBRA methods have been commonly used in field of metabolic pathway engineering (21; 22; 64).

COBRA Toolbox was released as a MATLAB package (78). It contains several *in silico* methods for quantitative prediction of cellular biochemical networks, with constraint-based modelling. In addition, it reads and writes models in the universal format SBML.

The latter version, *COBRA Toolbox* v3.0 (69), includes methods and tools implemented in several categories. Among these are:

- Gap filling
- Network Reconstruction
- Reconstruction refinement
- FBA
- Network visualization
- FVA

After *COBRA Toolbox* creation, an extension was designed: COBRApy (16). A Python package that works in an object-oriented way and facilitates the representation of metabolism and gene expression.

2.3.1.3 *OptFlux*

OptFlux is an open-source and modular software used in metabolic engineering. It provides easy access to a number of tools, due to its user-friendly interface (73)

Among the different algorithms and tools OptFlux offers, are:

- Metabolic Flux analysis (MFA) (87)
- Evolutionary Algorithms (EA) (67), the *OptKnock* algorithm (7) or simulated annealing metaheuristics for strain optimization through identification of metabolic engineering targets(75).
- FBA, Minimization of Metabolic Adjustment (MOMA) (80) and Regulatory on/off minimization of metabolic flux changes (ROOM) (82).
- Model visualization module

2.3.1.4 *CarveMe*

CarveMe (56) is an open-source and user-friendly tool for reconstruction of both species and community models. Its novelty arises from the fact that a top-down reconstruction approach (Figure 2) is used.

Thus, a manually curated universal model is used for the automated reconstruction. It includes a universal biomass equation, import/export reactions and does not contain blocked or unbalanced reactions.

To create an organism-specific model, reactions and metabolites not predicted to be present in the desired organism are removed, and the manual curation and structural properties of the original model are maintained. Additionally, for the creation of a microbial community model, the desired single-species models are merged, resulting in a community network.

2.3.1.5 *KBase*

The DOE Systems Biology Knowledgebase (KBase) (3) is an open-source software and data platform. It aids comparative genomics of plants, prediction of microbiome interactions and metabolic modeling of microbes, by providing a set of applications for genome assembly, annotation, metabolic model reconstruction, FBA, expression analysis and comparative genomics. Thus, it can be used in single organisms and community modeling.

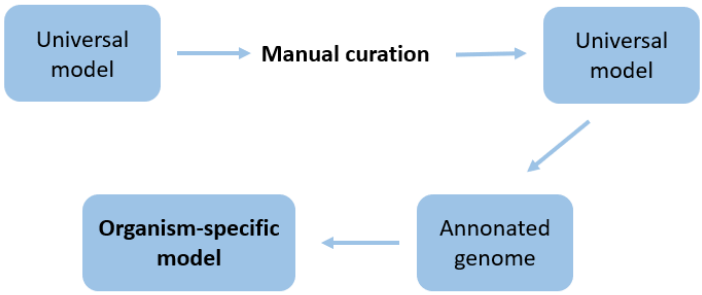


Figure 2: The top-down approach of model reconstruction. A universal model is generated and manually curated. This is then used as a template for organism-specific model generation. Adapted from Machado *et al.* (56)

2.3.2 *Community Based Tools*

2.3.2.1 *Community Flux Balance Analysis*

Community Flux Balance Analysis (cFBA) (46) is a computational method whose primary focus is elucidating metabolic capabilities of a community and understanding metabolic interactions. It uses genome-scale stoichiometric models of metabolism, metabolic interactions between species in the community, abiotic processes and considers constraints deriving from reaction stoichiometry, reaction thermodynamics, and the ecosystem.

Along with the abundance of all species present in the community, cFBA predicts fluxes distribution, growth rates and exchange fluxes between the microbes and the environment as well. It is a direct extension of FBA for single organisms to communities, by applying community-specific constraints.

Considering two organisms in a community, cFBA can be used after merging both species' GSM model reconstructions, resulting in one stoichiometric matrix of the consortium which contains all the internal and transport reactions, unique extracellular reactions and cross feeding reactions. The latter are determined by analyzing the common set of exchange metabolites. Biomass abundances of the individual organisms are also considered, in addition to the whole consortium simulations.

The problem definition of cFBA is non-linear if biomass fractions are considered a variable and the number of organisms in the community affects the number of linear programs to be solved. Even though this approach initially presents a non-linear programming problem if the biomass fraction is variable, by fixing the individual biomass values, a *Linear Programming (LP)* problem arises, thus providing a result which identifies the optimal spe-

cific flux values. This approach states that the entire community is at steady state and thereby all metabolites mass balances equal zero and reaction rates stay constant. It should also be highlighted that the community growth rate does not need to equal the maximal growth rate of any individual organism. However, when considering a mutualistic interaction between growing organisms, to maintain a steady-state of the cross-feeding metabolite, producing and consuming organisms need to grow at an equal pace.

2.3.2.2 *Dynamic Flux Balance Analysis*

Dynamic Flux Balance analysis (dFBA) (37; 58) is another direct extension of FBA, used to simulate, analyse and optimize synthetic microbial communities in diverse contexts. In comparison to other modelling approaches, it provides the incorporation of GSM models in a dynamic model. Thus, the result is a better prediction of time-varying species metabolism and interactions, as it accounts for dynamic changes in the concentration of metabolites. However, one could consider the fact that substrate uptake kinetics needs to be determined, a limitation.

A steady-state community flux balance model is created by merging the stoichiometric matrices of the individual organisms into one and solving an objective function for the whole community, consequently calculating the unknown fluxes by solving a LP problem.

To develop a dynamic flux balance model for a community, the steady-state flux balance model described above is combined with substrate and exchange metabolites uptake kinetics. This model also includes extracellular mass balances on substrates and products.

The following expression represents the LP problem representing the multispecies extension of single-species FBA.

$$\begin{aligned} max\mu = \mu_1 + \mu_2 + ... + \mu_n = w^T v \\ Sv = 0 \\ \beta \geq v \geq \alpha \end{aligned} \tag{5}$$

Where

- For a community of n organisms, μ_n is the respective individual growth rate
- μ is the community growth rate
- w is a vector of biomass weighing coefficients containing the coefficients of the individual species
- S is the stoichiometric matrix
- v is the fluxes vector

- β and α are the upper and lower bounds, respectively.

The extension of classical FBA is based on specifying substrate uptake fluxes of the individual species and solving equation (3) for the unknown fluxes. Although classical FBA is still used to compute the growth rate, intracellular fluxes v and the product secretion rates v_p, extracellular substrate concentrations S and product concentrations P are now used for calculation of time-varying substrate uptake rates v_s through expressions for the uptake kinetics, as described in Figure 3, a representation of the dFBA framework.

The substrate uptake rates describe the maximum rates possible caused by transport limitations and are incorporated as upper bounds on the uptake rates calculated in the FBA problem.

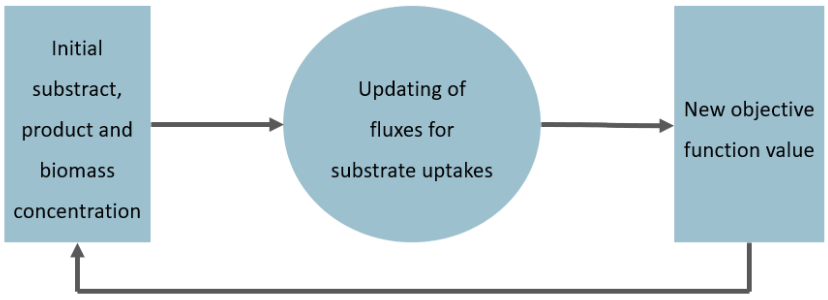


Figure 3: Representation of a dynamic flux balance analysis. Initial substrate, product and biomass concentrations are introduced. The subset of fluxes for substrate uptake and secretion rates changes as the concentration values are updated. For each change, a new objective function value is calculated. As a result, exchange values change continuously over time. Subsequently, available substrate and product concentration changes. Adapted from Henson *et al.* (37).

Ultimately, the dFBA approach provides certain advantages when studying communities due to its higher capability for capturing time-varying species metabolism and interactions. Nonetheless, its limitations are something to take into consideration. Substrate uptake kinetics determination is required, which has proved to be one of the main challenges of this approach.

2.3.2.3 *DyMMM*

Dynamic Multispecies Metabolic Modeling (DyMMM) (97) framework was implemented in MATLAB and is available as an add-on to COBRA toolbox. It is an extension of the aforementioned dFBA approach (58), which states that a steady state is reached as a consequence of changes in the extracellular environment. As FBA predicts rates, these are used to update the extracellular environment.

A community metabolic model must account for the metabolic exchanges between species and with the environment, as well as the changes in biomass of the modeled species. The DyMMM framework describes a community of N microbial species coexisting in an environment containing a specified number of metabolites. It integrates all the community's microbial species growth rates, as well as the production/consumption (exchange) rates of all metabolic species in the environment.

The growth rate (dX/dt) of every microbial species in the community is given by equation 6, whereas the consumption/production rate (dS/dt) of every metabolite in the environment is given by equation 7:

$$\frac{dX_j}{dt} = \mu_j X_j \tag{6}$$

$$\frac{dS_i}{dt} = \sum_{j=1}^{N} V_i^j X_j \tag{7}$$

Where j is a species in the community, X_j its biomass and μ_j its growth rate. As i is a metabolite, S_i represents its concentration in the environment and V_i^j its consumption/production rate by the organism j .

V_i^j and μ_j values are calculated for each microbial species in the community, using FBA, as represented in the equation set 8:

$$\begin{aligned} Maximize\ \mu_j &= c^T v_j \\ subject\ to\ S_j v_j &= 0 \\ {v_j}^{max} \geq v_j \geq {v_j}^{min} M &\geq i \geq 1 \\ N &\geq j \geq 1 \end{aligned} \tag{8}$$

Where c^T is the objective function, S_j is the stoichiometric matrix of the organism j, ${v_j}^{max}$ and ${v_j}^{min}$ are its flux capacity constraints, based on its respective genome-scale model. When considering external metabolites, ${v_j}^{max}$ and ${v_j}^{min}$ to the respective fluxes can be calculated environmental concentration of these metabolites, using either an On/Off method (if the uptake kinetics are not available) or the MichaelisMenten kinetics method (if kinetics information is available). M is the number of metabolites in the community and N the number of microbial species.

Given the flux constraints, a solution where the objective function (specific growth rate) is maximal is calculated by LP.

2.3.2.4 *DFBAlab*

Dynamic Flux Balance Analysis laboratory (DFBAlab) (26) is a MATLAB-based, dynamic simulation framework. Similarly to DyMMM, it is based on the dFBA approach.

Given a vector x_0 that contains the initial concentrations of metabolites and biomass in a culture, after specifying metabolites' exchange fluxes for each species (exchange fluxes), mass transfer rates, feed and discharge rates from the culture and other dynamic processes, a rate of change function f can be calculated for each of the elements of x_0. After integrating the function f, the concentration profiles with respect to time, x(t), are assessed.

Lexicographic optimization is used by DFBAlab. Thus, objectives are ordered by priority. In this case, the first objective is the maximization of biomass, followed by maximization or minimization of the specified metabolites exchange fluxes.

Additionally, as LPs have a single optimal objective function value, such value changes continuously with changes in the lower and upper bounds of a species, in a dynamic approach.

In sum, DFBAlab combines lexicographic optimization with a LP feasibility problem to generate an extended dynamic system for which the LP always has a solution.

2.3.2.5 *μbialSim*

μbialSim (68) is based on dFBA. Thus, dynamic shifts, as a consequence of the systems dynamics are simulated. Additionally, the time course in terms of composition and activity of a community, in batch or chemostat may be predicted.

As the compartmentalized approach (88) is used to generate a community model, separate genome-scale metabolic network models are used to model the activity and growth of individual species. Additionally, each organism has access to a common pool of metabolites, enabling metabolite exchange. Compound exchange and growth rates are computed through FBA and used to update the state variables of the model.

Discontinuities in intracellular fluxes over time can occur in dFBA simulations, as different flux distributions can lead to the same maximal growth rate. However, to avoid such discontinuities, μbialSim has two features, which can be used individually or one succeeding the other:

- pFBA, determining a FBA solution with a minimal sum of fluxes;
- A methodology similar to MOMA, where the optimal growth rate is found by flux distributions

All the simulated exchange fluxes are calculated by μbialSim. Thus, fluxes for both the exchange reactions that are coupled to the pool compounds and those that are not, can be obtained.

2.3.2.6 *OptCom*

OptCom (99) uses a multi-level optimization structure. It assesses the interactions between the species involved, either positive, negative or both.

It consists on two levels:

- An inner-level regarding individual fitness level of organisms
- An outer-level regarding community fitness.

The existence of two levels ensures that the combined biomasses (individual organisms and the community) is maximum. It is important to note that it maximizes community fitness, while not compromising the fitness of its individual organisms.

If more constraints are provided, OptCom is also able to compute cases where organisms do not present their maximum growth, exhibiting cooperative behavior instead.

Additionally, a new perspective of OptCom arised with the creation of d-OptCom (98), which takes dynamics of the microbial communities into consideration. It integrates dynamic mass balance equations referring to metabolite production, organism growth and substrate uptake. The result is the prediction of biomass and metabolite exchange between the community members.

Finally, a community-level objective function is obtained, subject to biomass maximization for each species.

2.3.2.7 *SteadyCom*

SteadyCom (8) is an optimization framework, integrated in COBRA Toolbox. It predicts the metabolic flux distribution, assuring that steady-state is imposed. In addition, it is compatible with flux balance analysis.

When a single organism is considered, the biomass flux is normalized by the organism's rates of consumption or production. However, when multiple organisms are growing, there is not a constant growth rate for all microbes. Therefore, the fastest growing organism can outgrow the rest of the community members. Thus, SteadyCom imposes a steady-state condition, where an organism cannot secrete metabolites if it is not growing, to avoid these situations.

Additionally, SteadyCom considers an aggregate flux, which is described as the total biomass of a population. It quantifies the metabolites that an organism can consume or produce in a microbial community. Furthermore, an organism can only have non-zero fluxes if both its biomass production is non-zero as well.

2.3.2.8 *micom*

"micom" (microbial communities) (13) is a Python software package, based on the COBRApy Python package (16).

As aforementioned, FBA assumes a steady state for all fluxes in a biological system and optimizes an organism-specific biomass, calculating the approximate fluxes of the organism.

Similarly, micom assumes that growth rates and relative abundances in the community are in steady state. Thus, its mathematical formulation is similar to the aforementioned mentioned OptCom and SteadyCom frameworks' formulation, where there are two different classifications for growth rates: individual growth rates, which estimate the growth rate of a single organism, and the community growth rates, which represent the growth of the entire community.

To represent the community, a particular abundance for each organism (in gDW) is considered and each organism is allocated to an external compartment which represents the community environment. Given a particular abundance of an organism, a sub-model in the whole community, micom scales the whole community's internal exchange fluxes to the respective abundance value.

Finally, both transport (organism-common compartment) and community exchange reactions (which can be used to constrain the model) are added, resulting in the final community model.

2.3.2.9 *RedCom*

RedCom (51) consists of an approach to build community models, which is based of the compartmentalized model approach (88). Nevertheless, it builds *reduced* community models.

Firstly, the individual models are reduced by eliminating reactions that are not relevant for the individual organism's growth. If the single organisms' exchange fluxes in the community models do not pass an imposed minimality criterion, these solutions are discarded and the respective reactions are removed. Subsequentely, the community model is created using the smaller models to avoid unrealistic solutions, where a species altruistically synthesizes large amounts of byproducts while not synthesising a corresponding proportion of biomass.

As the created models are smaller, smaller ranges of feasible community compositions and exchange fluxes are calculated.

2.3.2.10 *OptDeg*

In Koch *et al.* (2016), given a created community model, a community composition hierarchical optimization approach was presented, consisting of two objective functions. Firstly, the maximization of community growth rate is optimized, followed by the optimization of the maximization of biomass yield.

Regarding the quantification of the overall biomass yield (secondary objective), an optimality degree (OptDeg) was introduced.

OptDeg serves as a comparison between the biomass yields of each species population and the respective theoretically feasible maximum. Thus, for OptDeg = 1, all the community species grow with maximum specific growth rate *and* maximum biomass yields.

2.3.2.11 *SMETANA*

Species Metabolic Interaction Analysis (SMETANA) (95) is a mixed-integer LP method that begins by assembling a community model from the single-species' models. Then it allows predicting resource competition and metabolic cross-feeding the community.

The degree of competition in these communities is determined by *Metabolic Resource Overlap (MRO)*, which is an intrinsic property of any community and is defined by the maximum possible overlap between the minimal nutritional requirements of all member species.

This method uses three scores to identify possible inter-species interactions:

- Species Coupling Score (SCS)
- Metabolic Uptake Score (MUS)
- Metabolite Production Score (MPS)

After the simulation, a SMETANA score evaluates the strength of the community through enumeration of possible metabolites exchange and it can be defined as the sum of all inter-species dependencies (SCS, MUS and MPS).

Metabolic Interaction Potential (MIP) and phylogenetic relatedness are negatively correlated, which leads to the assumption that related species metabolism requirements are similar.

MIP represents the tendency of community members to exchange metabolites and is given by the maximum number of essential nutritional components that a community can provide for itself, through interspecies metabolic exchange. Thus, the higher its value, the higher the probability of community members benefiting from metabolite production from other members. The community's size and its MIP are positively correlated.

MIP is calculated as the difference between the minimal number of components required for the growth of all members in a non-interacting community and an interacting community.

MIP and MRO represent opposite situations. For instance, MIP represents the probability of community members benefiting from metabolite production from other members and MRO represents the overlap between the minimal nutritional requirements of all member species.

While the first metric represents the propensity of the community's organisms to depend on each other, not being able to grow on its own, MRO indicates the propensity to competition between the organisms, since they require both metabolites from a specific environment. Thus, SMETANA does not calculate both metrics simultaneously.

As MIP and MRO are intrinsic properties of any community, for its determination, only information regarding biosynthetic and metabolite transport capabilities of its member species is required.

3

MATERIALS AND METHODS

The present chapter describes the methods used to model different microbial communities and microbe-microbe interactions, subject to different environmental conditions, using different tools and software.

3.1 SOFTWARE

MATLAB v18.3.4 was used to run pFBA (55), SteadyCom (8), DyMMM (97) and DFBAlab's (26) simulations. Python 3.7 was used to run micom's methods (13) and SMETANA (95). Cplex v.12 (academic license) was used as a solver in all simulations.

Even though other tools were considered during this work's timeline, problems related to acquiring, installing or running these tools have emerged, which rendered their evaluation infeasible. The specific reasons preventing the evaluation of such tools are presented in Table 2.

Table 2: Tools and causes leading to the abdication of their usage during this works' timeline. Even though some tools were described in publications, they were not available for download. cFBA, which runs in Python 2, requires the usage of packages with compatability issues. μbialSim's usage was discarded as errors were arising during its usage.

Tool	**Cause**	**Publication**
cFBA	Incompatibility with Python packages	Khandelwal *et al.* (2013)
OptDeg	Not available	Koch *et al.* (2016)
RedCom	Not available	Koch *et al.* (2019)
OptCom	Not available	Zomorrodi *et al.* (2012)
d-OptCom	Not available	Zomorrodi *et al.* (2014)
μbialSim *	Errors during usage	Popp and Centler (2019)
CASINO	Not available	Shoaie *et al.* (2015)

* Errors arose during the the different GSM models import. However, this does not imply that the tool's methods are unavailable or unable to obtain community based solutions.

Different tools follow different simulation approaches. The general features of the software used in this work are discriminated in Table 3.

Table 3: Description of the used tools. Information regarding the algorithm a tool uses to perform simulations, the used approach to build the community; the used versions and what type of software the tool is based on. SMETANA does not perform simulations nor builds the community. Thus, information regarding such aspects is null.

Tool	Simulation type	Approach	Version	Type of software
SteadyCom	FBA	Compartmentalized	-	MATLAB package
SMETANA	-	-	1.0.0	command line interface
DFBAlab	dFBA	-	-	MATLAB scripts
DyMMM	dFBA	-	1.3	MATLAB scripts
micom	pFBA	Compartmentalized	0.9.3	Python package
COBRA Toolbox	pFBA	Compartmentalized	3.0.0	MATlab Toolbox

Regarding the compartmentalized model approach, implemented by Klitgord and Segre (2010), a new community compartment is created, bearing metabolite exchange between species (Figure 4). Thus, the resulting community model consists in three types of reactions: internal, membrane transport, and environment exchange reactions, commonly referred as drains. Membrane transport reactions present a connection between the individual organisms and the common community compartment, which serves as a mediator between these and the extracellular environment.

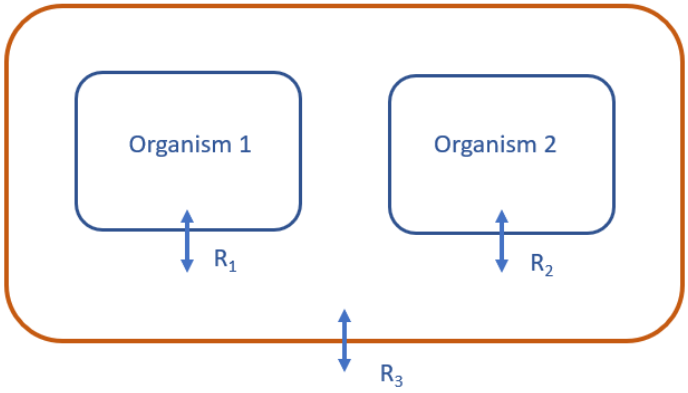


Figure 4: Schematic representation of the created community reactions. R_1 and R_2 represent the transport reaction connecting organism 1 and organism 2 to the common community compartment, respectively, and R_3 the exchange reactions between the common compartment and the environment conditions to which the community is subject to.

3.2 CASE STUDIES

To assess the similarities and differences between tools and the tools' performances, different case studies, with different practical applications, were selected and studied in the present work.

To select a case study, some aspects had to be considered. For a given community, the organisms' GSM models must have the same metabolite and reaction *Identifier (ID)*, to correctly create the community stoichiometric matrix. Additionally, information regarding growth conditions had to be present in literature, whether it was regarding *in silico* simulations or experimental data.

The selected case studies and the used GSM models were:

- Case A: A hypothetical community composed of four *Escherichia coli* mutants (8);
- Case B: A consortium where *Bacteroides caccae*, a fiber-degrading organism, commonly present in microbiomes, and *Lactobacillus rhamnosus*, a common human probiotic strain, grow together, using semi-automatically reconstructed models published in the AGORA collection (57);
- Case C: A *Lactobacillus bulgaricus* and *Stroptococcus thermophilus* consortium, usually used as a starter culture in the manufacturing of yoghurt (32);
- Case D: A community composed of *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*, two members of the gut microbiota (19);
- Case E: A consortium composed of *Desulfobrio vulgaris* and *Methanococcus marapaludis*. The organisms' cooperation was studied, due to the applications of such process for the biogas production, using stoichiometric models of the organisms' core metabolism (88; 51);
- Case F and G: A *Saccharomyces cerevisiae* and *Escherichia coli* co-culture, designed for studying efficient aerobic consumption of glucose/xylose mixtures in batch fermentation (33; 60; 26); The study of *S. cerevisiae* and *E. coli*'s co-culture behaviour, with only glucose as carbon source.

The used genome-scale models vary in number of metabolites, reactions, genes and compartments. As *S. cerevisiae* is an eukaryotic organism, its model's complexity is higher. Thus, the number of compartments is higher than the other GSM models' used in this work, which represent prokaryotic organisms and are composed of two compartments: cytosol and extracellular space, in exception to *D. vulgaris* and *M. maripaludis*, whose models are exclusively composed by one compartment which represents intracellular space.

3.2.1 *Case study A*

The studied consortium consists of four *E. coli* mutants, with cross-feeding of amino acids between them. Each mutant strain consumes two amino acids from the medium, for which it is auxotrophic. In addition, each mutant lacks one amino acid transporter, thus only having the ability to secrete one.

The *E. coli* K-12 MG1655 (iAF1260) GSM model (24) was used to design the four different strains, as performed by Chan and coworkers (8). Thus, four different copies of the original model were created, in order to simulate the presence of different organisms in the community.

To study the cross-feeding between the four *e. coli*'s, internal and amino acid uptake reaction bounds were limited in the individual models to obtain the following organisms:

- Strain one (Ec1): auxotrophic for lysine and methionine, unable to export phenylalanine;
- Strain two (Ec2): auxotrophic for arginine and phenylalanine, unable to export methionine;
- Strain three (Ec3): auxotrophic for arginine and phenylalanine, unable to export lysine;
- Strain four (Ec4): auxotrophic for lysine and methionine, unable to export arginine.

The expected production and respective consumption of the studied amino acids by the different strains, as described in (8), is represented in Figure 5.

Environmental constraints

The environmental conditions were constrained according to what Chan and co-wrokers (8). The study was performed under aerobic conditions. Thus, oxygen is provided in the community medium, in addition to glucose and ions, as represented in Table 4 (section 3.2.1). Additionally, the cross-feeding of amino acids between the organisms was limited by constraining the amino acid exchange reaction to a rate of 1 or -1 mmol/gDW/h, representing production and consumption of the metabolite, respectively.

As the purpose of the case study was to design mutants, incapable of transporting and producing specific amino acids, internal reactions (Table 5) were constrained by limiting its upper and lower bounds to 0. Thereby, these reactions' absence is simulated. Moreover, similar to what was performed by Chan and coworkers (8), a methionine transport reaction (METt3pp) was added to the model.

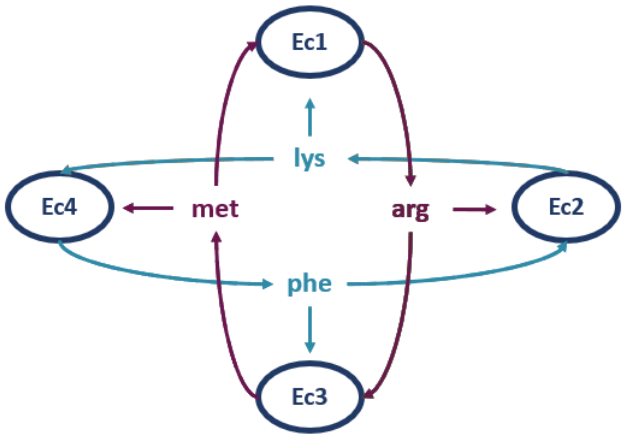


Figure 5: Amino acid production and consumption by the different strains. *arg*, *lys*, *met* and *phe* represent L-arginine, L-lysine, L-methionine and L-phenylalanine, respectively. Adapted from (8).

3.2.2 *Case study B*

The GSM models of *Lactobacillus rhamnosus* GG ATCC 53103 (LGG) and *Bacteroides caccae* ATCC 43185 (57) were used to study the consortium;

LGG has the potential to establish coaggregative interactions with *Bacteroides* strains, having potential for positive effects on the host's health. Specifically, probiotics are propense to inhibit the colonization of pathogens by competing with them for substrate (79). Thus, the possible interactions consisting of the cross-feeding of metabolites between the strains were studied (57).

The uptake and secretion rates of the major cross-fed metabolites in this community, according to the results in Magnúsdóttir et al., (2017), were analysed:

- Uptake of metabolites such as glucose and arabinogalactan by the community;
- Secretion of L-alanine by *B. caccae*, and its consequent consumption by LGG;
- Lactate production by LGG and its consumption by *B. caccae*;
- Acetate secretion by the community.

Environmental constraints

The community's lower bounds were constrained to values based on the DMEM 6429 defined culture medium composition supplemented with arabinogalactan (Table 6), as performed in (57). Therefore, glucose, galactose, amino acids and vitamins were provided in the medium.

Table 4: Community lower bounds' constraints (mmol/gDW/h) to simulate the desired environmental conditions. Ec1, Ec2, Ec3 and Ec4 represent the four *E. coli* strains. The limit of intake is defined by negative values and export limits are defined by positive values.

Metabolites	Community	Ec1	Ec2	Ec3	Ec4
Ca^{2+}	-1000	-1000	-1000	-1000	-1000
Cobalamin	-0.01	-0.01	-0.01	-0.01	-0.01
Cl^{+}	-1000	-1000	-1000	-1000	-1000
CO_2	-1000	-1000	-1000	-1000	-1000
Co^{2+}	-1000	-1000	-1000	-1000	-1000
Cu^{2+}	-1000	-1000	-1000	-1000	-1000
Fe^{2+}	-1000	-1000	-1000	-1000	-1000
Fe^{3+}	-1000	-1000	-1000	-1000	-1000
K^{+}	-1000	-1000	-1000	-1000	-1000
H_2O	-1000	-1000	-1000	-1000	-1000
D-Glucose	-8	-8	-8	-8	-8
H^{+}	-1000	-1000	-1000	-1000	-1000
Mg^{2+}	-1000	-1000	-1000	-1000	-1000
Mn^{2+}	-1000	-1000	-1000	-1000	-1000
Molybdate	-1000	-1000	-1000	-1000	-1000
Phosphate	-1000	-1000	-1000	-1000	-1000
Na^{+}	-1000	-1000	-1000	-1000	-1000
Ammonium	-1000	-1000	-1000	-1000	-1000
O_2	-18,5	-18,5	-18,5	-18,5	-18,5
SO_4	-1000	-1000	-1000	-1000	-1000
Tungsten	-1000	-1000	-1000	-1000	-1000
Zn^{2+}	-1000	-1000	-1000	-1000	-1000
L-lysine	0	-1	0	0	-1
L-methionine	0	-1	0	0	-1
L-phenylalanine	0	0	-1	-1	0
L-arginine	0	0	-1	-1	0

Table 5: Reactions which lower and upper-bounds values were constrained to 0 mmol/gDW/h, in order to simulate the desired mutations. DAPDC is essential for lysine biosynthesis, ARGSL for arginine's, HSST for methionine's and PPNDH for phenylalanine's. METt3pp is the transport reaction for methionine, PHEt2rpp for phenylalanine, LYSt3pp for lysine and ARGt3pp for arginine

Ec1	Ec2	Ec3	Ec4
DAPDC	ARGSL	ARGSL	DAPDC
HSST	PPNDH	PPNDH	HSST
PHEt2rpp	METt3pp	LYSt3pp	ARGt3pp

Table 6: Applied environmental conditions, by constraining the lower bounds (mmol/gDW/h) of the community's exchange reactions and internal transport reactions connecting *B. cacce* and LGG to the common compartment. The limit of intake is defined by negative values and export limits are defined by positive values.

Metabolite	Community	*B. caccae*	*LGG*
Arabinogalactan	-0.094	-0.094	-
Ca^{2+}	-1	-1	-1
Choline	-1	-1	-1
Cl^{-}	-1	-1	-1
Co^{2+}	-1	-1	-1
Cu^{2+}	-1	-1	-1
D-glucose	-4.5	-4.5	-4.5
Fe^{2+}	-1	-1	-1
Fe^{3+}	-1	-1	-1
Folate	-1	-1	-1
Glycine	-1	-1	-1
H_2O	-10	-5	-5
H_2S	-1	-1	-1
Inositol	-1	-1	-1
K^{+}	-1	-1	-1
Lactate	0	-20	0
L-alanine	0	0	-1
L-arginine	-1	-1	-
L-asparagine	-1	-1	-1
L-aspartate	-1	-	-1
L-cysteine	-1	-1	-1
L-cysteinylglycine	-1	-1	-1
L-glutamate	-1	-1	-1
L-glutamine	-1	-1	-1
L-histidine	0	0	-1
L-isoleucine	-1	-1	-1
L-lysine	-1	-1	-1
L-methionine	-1	-1	-1
L-phenylalanine	-1	-1	-1
L-proline	-1	-1	-1
L-serine	-1	-1	-1
L-threonine	-1	-1	-1
L-Tryptophan	-1	-1	-1
L-tyrosine	-1	-1	-1
L-valine	-1	-1	-1
Metaquinone 8	-1	-1	-1
Mg^{2+}	-1	-1	-1

Mn^{2+}	-1	-1	-1
NH_4	-1	-1	-1
Nicotinate	-1	-1	-1
Phosphate	-10	-5	-5
Pyridoxal	-1	-1	-1
R-Pantothenate	-1	-1	-1
Riboflavine	-1	-1	-1
Siroheme	-1	-1	-1
SO_4	-1	-1	-1
Spermidine	-1	-1	-1
Succinate	-1	-1	-1
Thiamine	-1	-1	-1
Ubiquinone-8	-1	-1	-1
Urea	-1	-1	-1
Zn^{2+}	-1	-1	-1

3.2.3 *Case study C*

The GSM models of *Lactobacillus delbruekii* ssp. *bulgaricus* ATCC BAA 365 (57) and *Streptococcus thermophilus* LMG18311 (66) were used for this case study.

To standardize the models, reaction and metabolite identifiers from the *S. thermophilus* GSM model were converted into *Biochemical, Genetic and Genomic (BIGG)* identifiers (48).

Concerning cross-feeding between the *L. bulgaricus* and *S. thermophilus*, the first is expected to have proteolytic activity. Therefore, reaction fluxes related to casein degradation by *L. bulgaricus*, its secretion of "milk peptide" and consequent uptake by *S. thermophilus* were throughly analysed. In addition to this first metabolite exchange, formate is expected to be secreted by *S. thermophilus* and consumed by *L. bulgaricus* (32). Hence, the following aspects were analysed:

- The community's lactose consumption;
- The community's lactate and galactose secretion;
- Formate cross-feeding between the organisms;
- Peptide degradation by both organisms;

Milk contains casein, a protein. In a consortium composed of *S. thermophilus* and *L. bulgaricus*, the latter can degrade casein into smaller peptides (32; 81). As *S. thermophilus* LMG18311 does not have proteolytic activity (10), it benefits from *L. bulgaricus* presence

(39; 81). The smaller peptides resulting of *L. bulgaricus*' activity, can be degraded by *S. thermophilus* to produce amino acids.

Due to not being present in *L. bulgaricus*' model, the following reactions were added to simulate casein and milk peptide degradation:

- A protease reaction regarding the degradation of casein into "milk peptide";
- The transport reaction of such peptide into *L. bulgaricus*' cell;
- A reaction regarding the degradation of the "milk peptide" into the different amino acids.

Additionally, reversible symport transport reactions for L-methionine, L-arginine and L-proline were added to *L. bulgaricus*' model. These amino acids are metabolized from milk peptide and can therefore be secreted by the organisms. Thus, the absence of the respective transport reactions could lead to the inability to reach a steady-state.

Environmental constraints

Environmental conditions (Table 7) were firstly designed by adding the required metabolites for the individual organisms' minimal growth. Additionally, free ions were added to the medium and lactose was considered the primary carbon source, similar to milk. As milk contains vitamins such as biotin and proteins such as casein (32; 54), both were added to the medium.

S. thermophilus LMG18311 is auxotrophic for L-histidine. Nevertheless, the free amino acid was not supplied to the medium, as the organism can metabolize it from the degradation of milk peptide, provided by *L. bulgaricus*.

Finally, to correctly simulate *S. thermophilus*' metabolite's consumption and production, the lower bound of the "Alanine dehydrogenase" (EC 1.4.1.1) reaction, was constrained to 0.

Table 7: Environmental conditions applied, by constraining the lower bounds (mmol/gDW/h) of the community's exchange reactions and internal transport reactions connecting *S. thermophilus* and *L. bulgaricus* to the common compartment. The limit of intake is defined by negative values and export limits are defined by positive values.

Metabolite	Community	S. thermophilus	L. bulgaricus
Acetate	0	0	0
Casein	-10	-	-10
Milk peptide	0	-5	-5
Biotin	-100	-	-100
Ca^{2+}	-100	-	-100

Cd^{2+}	-100	-	-100
Cl^{-}	-100	-	-100
CO_2	-100	-100	-100
Co^{2+}	-100	-	-100
Cu^{2+}	-100	-	-0.5
Fe^{2+}	-100	-	-100
Fe^{3+}	-100	-	-100
Formate	0	0	-1
Glycine	-1	-0.5	-0.5
H_2O	-100	-100	-100
K	-100	-	-100
Lactose	-10	-5	-5
L-asparagine	-1	-	-0.5
L-glutamine	-1	-0.5	-0.5
Mg^{2+}	-100	-	-100
Mn^{2+}	-100	-	-100
NH_4	-100	-	-100
Nicotinate	-1	-0.5	-0.5
Phosphate	-1	-0.5	-0.5
Pyridoxal	-1	-	-0.5
Riboflavin	-1	-0.5	-0.5
R-pantothenate	-1	-0.5	-0.5
SO_4	-0.5	-	-0.5
Thiamine	-1	-0.5	-0.5
Urea	-2	-1	-1
Zn^{2+}	-1	-	-0.5

3.2.4 *Case study D*

Bifidobacterium adolescentis L2-32 and *Faecalibacterium prausnitzii* A2-165's GSM models (19) were used to study this consortium.

As described in El-Semman *et al.* (2014), *B. adolescentis* is expected to generate acetate, which is consumed by *F. prausnitzii*. Subsequently, *F. prausnitzii* produces butyrate, which plays a critical role in colonic homeostasis and cancer prevention (19).

In vivo experiments showed the production of acetate, lactate, formate and ethanol by *Bifidobacterium*. However, in other *in silico* experiments, *Bifidobacterium* did not produce lactate under glucose limitation, as it maximized energy production by cleaving pyruvate into acetyl phosphate and formate (19).

In the present study, using El-Semman *et al.*'s GSM models, the community's behaviour regarding the following aspects was analysed:

- Production of acetate by *B. adolescentis* and its subsequent consumption by *F. prausnitzii*;
- Production of butyrate and formate by *F. prausnitzii*;
- Production of ethanol and formate by *B. adolescentis*

Environmental constraints

Community environmental conditions were created based on the original upper and lower bounds of the individual organisms, as described in El-Semman *et al.*'s (2014) work.

As both organisms compete for the available glucose in the medium as a carbon source, the following constraints were originally imposed:

- Community exchange reaction for glucose was constrained to a lower-bound value of -1 mmol/gDW/h;
- Individual organisms uptake of glucose from the common compartment was constrained to -1 mmol/gDW/h.

Once pFBA methods were performed in such environmental conditions, *F. prausnitzii* did not display any growth. Therefore, to ensure growth from both organisms and further study its interactions, glucose availability in the system was constrained as follows:

- Community exchange reaction for glucose was constrained to a lower-bound value of -2 mmol/gDW/h;
- Individual organisms uptake of glucose from the common compartment was constrained to -1 mmol/gDW/h.

In both models, directionality differed among the set of reactions. Thus, when constraining the flux bounds, such factor had to be considered.

3.2.5 *Case study E*

Stoichiometric models representing *Desulfobrio vulgaris* and *Methanococcus marapaludis*'s core metabolism, published in (50), were used to study a two-species community.

D. vulgaris, an acetogenic organism, consumes lactate and produces acetate, carbon dioxide, hydrogen or/and formate, while *M. maripaludis*, a methanogen, produces methane and

consumes the produced hydrogen, carbon dioxide, and formate, therefore benefiting from the presence of *D. vulgaris* in the environment.

These models' structure does not contain an extracellular compartment. Moreover, only the internal part of the cell is considered in the models as a compartment.

Only pFBA, available in MATlab, was used to perform simulations to study the community's interactions. SteadyCom, micom and SMETANA's methods were not used as:

- The models' structure does not contain an extracellular compartment, which is necessary for the creation of a community model in micom or to perform SteadyCom's simulations;
- SMETANA is not compatible with the models' structure as these lack an extracellular compartment and do not match any of the available SBML formats.

The community model was created using the COBRA Toolbox function *createMultipleSpeciesModel*. As the individual models are composed of one compartment only and lack an extracellular representation, the created community model does not have a new common compartment as it is not possible to follow the compartmentalized approach to generate the community model.

Nevertheless, the individual community members' models were merged, creating a larger stoichiometric matrix, as described in the supra-organism approach (76), resulting in a sole reaction set.

Environmental constraints

Regarding supply of metabolites in the environment, only lactate is required. The environmental conditions were simulated using the information in Table 8.

Table 8: Used lower bounds' constraints (mmol/gDW/h) to constrain *D. vulgaris* and *M. maripaludis* individual uptake reactions. The limit of intake is defined by negative values and export limits are defined by positive values.

Metabolite	***D. vulgaris***	***M. maripaludis***
Acetate	0	-5
CO_2	0	-5
Formate	0	-5
Lactate	-5	-5
Methionine	-	-5
H_2	0	-5

Finally, CO_2 secretion by *M. maripaludis* was constrained to 0 mmol/gDw/h and pFBA methods were used in MATlab, using COBRA Toolbox's functions.

3.2.6 *Case Study F and G*

Similar to what was performed by Hanly and coworkers (33), *Saccharomyces cerevisiae* S288C (iND750) (14) and *Escherichia coli* K-12 substr. MG1655's (iJR904) (72) GSM models were used to perform simulations and calculations on DFBAlab and DyMMM regarding two different co-cultures:

- The co-culture of *S. cerevisiae* and the engineered *E. coli* strain ZSC113;
- The co-culture of *S. cerevisiae* and a non-engineered *E. coli* strain.

E. coli ZSC113 only consumes xylose, differing from the original strain, which consumed glucose (33). The flux bounds for glucose kinase and glucose exchange were constrained to zero in *e. coli*'s model, to mathematically simulate the associated gene deletion. Thus, glucose is consumed as a carbon source by *S. cerevisiae* while *E. coli* ZSC113 consumes xylose. consequently, ethanol is produced by *S. cerevisiae*, having an inhibitory effect on*E. coli*'s growth.

When considering the non-engineered *E. coli* in co-culture with *S. cerevisiae*, it is expected that both organisms consume glucose from the medium. Thus, the respective simulations where the flux bounds for glucose kinase and glucose exchange are not constrained were performed to compare the different co-culture results.

3.3 ANALYSES

The usage of the different tools is described in this section. For instance, in MATlab, COBRA Toolbox's methods were used to create community models for the different case studies. Such models were used to run pFBA and SteadyCom's simulations, calculating community growth and predicting community interactions.

In Python, micom's methods were used to create the community models which were used to run pFBA simulations using COBRApy's methods. SMETANA was used to calculate metrics regarding interspecies interactions.

Finally, DFBAlab and DyMMM were exclusively used to analyse case studies F and G's dynamics.

3.3.1 *COBRA Toolbox*

A community model was created for each case study. Firstly, the individual GSM models were read in MATlab and assigned to variables in order to create community structures. The individual organisms' reaction, metabolite and compartment IDs need to match, to properly

create community reactions and metabolites. Otherwise, duplicate community metabolites, with different IDs, would be created. Subsequently, the presence of such metabolites would generate an unconnected metabolic network, as metabolites are assumed as different between the organisms.

To generate the community stoichiometric matrix, the COBRA Toolbox function *createMultipleSpeciesModel*, which follows the compartmentalized approach, was used. Thus, a new common compartment, where all metabolites are provided to the community, was created. For a given extracellular metabolite in the individual GSM models, a transport reaction is created, allowing its uptake from the common compartment.

pFBA

Before running pFBA, environmental conditions had to be defined. Thus, community exchange reactions and individual organisms' transport reactions from/to the common compartment were properly constrained as aforementioned in each case study's description.

The objective function of the community model was set as the maximization of the individual organisms' biomass reactions. Thus, the community's growth rate is given as the sum of the individual organisms' growth rate values.

Upon definition of environmental conditions, COBRA Toolbox's function *optimizeCbModel* was used to run pFBA methods, simulating community growth, cross-feeding of metabolites between the organisms and community uptake and secretion rates.

3.3.2 *SteadyCom*

To run SteadyCom's methods, a community model is required. Thus, the model created in section 3.3.1, with the environmental conditions already specified via reaction upper and lower bounds, is used. Furthermore, the addition of the following variables to the community structure is required to run SteadyCom:

- All the exchangeable metabolites, identified by the respective compartment identification;
- A structure containing community reaction information (reaction names, exchange reaction names) - *infoCom*;
- An index structure corresponding to *infoCom*;
- The individual organisms' biomass reaction identification;
- The upper and lower bound of all reactions.

By default, SteadyCom uses the *simple guessing for bounds* algorithm, followed by MATlab's *fzero* to calculate flux bound values.

3.3.3 *micom*

The individual organisms' GSM models, in the SBML format, are required to build a microbial community model using micom's class *Community*.

Alike what was performed in MATlab, the individual models' metabolite, reaction and compartment IDs were changed when necessary, to match in all the used models. Such changes had to be performed in the SBML files prior to its input in micom.

The community structure is created automatically, following a compartmentalized approach. Thus, a common compartment mediating the exchange between the different organisms and the extracellular environment is created. Additionally, micom automatically sets the different organisms abundance in the community. For instance, for case study A, each strain's abundance was set to 0.25, as the community is composed of four organisms, thus representing equal proportions in the whole community. For the remaining cases, each strain's abundance was set to 0.5, as the consortia were composed of two organisms. These abundances are considered in all the created community reactions, as the community fluxes are scaled by organism abundance.

Additionally, the community objective function was automatically defined by micom. It represents the sum of the individual's biomass value, scaled by the respective abundance in the community.

All lower and upper-bounds of community exchange reactions and transport reactions connecting the individual organisms to the common compartment are set to -100 and 100 mmol/gDW/h by default. Thus, these were constrained to agree with the case studies' aforementioned environmental conditions.

After properly defining environmental conditions, pFBA (which is available as a COBRApy package function) was used to obtain community growth rate values, cross-feeding of metabolites information and community uptake/secretion fluxes.

3.3.4 *SMETANA*

SMETANA is a Python package available via command line interface. It takes as input genome-scale models in the SBML format and calculates metrics that describe and identify the potential for cross-feeding interactions between community members. Thus, it is not a simulation method.

As the used genome-scale models were reconstructed via different approaches, each genome-scale model file has its respective SBML format, as shown in Table 9.

Table 9: Used organisms and the respective genome-scale file SBML format in SMETANA.

Model	Format
E. coli (iAF1260)	BIGG
B. caccae	BIGG
LGG	BIGG
S. thermophilus	BIGG
L. bulgaricus	BIGG
B. adolescentis	Custom
F. prausnitzii	Custom
D. vulgaris	Custom
M. maripaludis	Custom
E. coli (iJR904)	BIGG
S. cerevisae	BIGG

After specifying the desired environment/extracellular media information, which consists in introducing IDs of metabolites present in the desired environmental conditions, the following metrics were calculated:

- The dependence of growth of a given organism on the presence of another, given by the species coupling scores (SCS)
- An organism's growth dependency on a metabolite produced by the other community member, given by the metabolite uptake score (MUS)
- An organism's ability to produce a certain metabolite, given by a binary score: the metabolite production score (MPS).

If one of the individuals requires a certain metabolite to grow and the other present organism has the metabolic capability to produce it, such metabolite is present in the results, with the corresponding scores of that possible interaction.

Finally, the communities' SMETANA score is calculated. It is given by the sum of all interspecies dependencies under the specified nutritional environment. Among the calculated scores, only results regarding the SMETANA score were analysed, as this score translates the strength of metabolic interaction between the organisms.

3.3.5 *DFBAlab and DyMMM*

DFBAlab and DyMMM are available in MATLAB code and provide dynamic community simulations. Both tools were used to study the potential of *S. cerevisiae* and *E. coli* ZSC113

co-cultures for efficient consumption of glucose/xylose mixtures and *S. cerevisiae* and *E. coli* co-cultures with glucose as substract.

- **DFBAlab**

As the considered communities were grown in batch fermentation, both dilution rates, in and out of the reactor, were constrained to o $\boldsymbol{h^{-1}}$.

It is necessary to provide the tool with the indices of the exchange fluxes of the metabolites we want to study. For instance, the exchange fluxes of glucose, xylose and oxygen are of interest in both case studies. Therefore, the corresponding reaction indices in *E. coli* and *S. cerevisiae*'s GSM models have to be inserted in DFBAlab .

Initial biomass concentration of each microbe was set to 0.05 g/L and glucose, xylose and ethanol concentrations were set to 16, 8 and 0 g/L, respectively, in both cases.

The addition of extracellular mass balance equations is required to turn a steady-state flux balance model into dynamic model. Thus, equations 9 and 10 represent the variation of *S. cerevisiae*'s and *E. coli*'s biomass concentration, while equations 11 to 13 represent glucose, xylose, and ethanol concentration variation over time.

$$\frac{dX_y}{dt} = \mu_y X_y \tag{9}$$

$$\frac{dX_c}{dt} = \mu_c X_c \tag{10}$$

$$\frac{dG}{dt} = -v_g X_y \tag{11}$$

$$\frac{dZ}{dt} = -v_z X_c \tag{12}$$

$$\frac{dE}{dt} = v_{e,y} X_y + v_{e,c} X_c \tag{13}$$

Where

- y and c represent *S. cerevisiae* and *E. coli*, respectively;
- μ_y and μ_c are the organisms' growth rates;
- X_y and X_c are the organisms' biomass;
- G, Z, and E represent the concentrations of glucose, xylose, and ethanol;
- $v_{e,y}$ and $v_{e,c}$ are the ethanol exchange fluxes for the two microbes;
- v_g and v_z are glucose and xylose uptake rates, respectively.

Additionally, kinetic expressions, which are necessary to calculate the glucose and xylose uptake rates' lower bounds, to limit their consumption, were added in the tool. Upon introduction of the equations, DFBAlab automatically updated the models' lower bounds.

$$v_g = v_{g,max} \frac{G}{K_g + G} \frac{1}{1 + \frac{E}{K_{ie}}} \tag{14}$$

$$v_z = v_{z,max} \frac{Z}{K_z + Z} \frac{1}{1 + \frac{E}{K_{ie}}} \tag{15}$$

Where

- g and z represent the concentrations of glucose and xylose;
- $v_{g,max}$ and $v_{z,max}$ are the maximum uptake rates of each substrate;
- K_g and K_z are corresponding saturation constants;
- K_{ieg} and K_{iez} are ethanol inhibition constants.

.

Ethanol plays an inhibitory effect in *E. coli* and *S. cerevisiaes*' growth. Therefore, ethanol inhibition constants were taken into account when defining the kinetic expressions for the respective sugar uptake rates. Additionally, as the wild-type *E. coli* strain does not consume xylose, the respective kinetic expression is $v_z = 0$.

To simulate the variation of metabolites over time, equations 9-10 were introduced as ODEs in DFBAlab . Additionally, the tool defined the biomass variation equations automatically.

As the uptake rates (equations 14 to 15) depend on substrate uptake parameters values, the respective values, retrieved from Hanly *et al*. (33), were introduced and are described in Table 10.

Finally, the tool defines biomass yield maximization as the objective function in both organisms.

- **DyMMM**

It is necessary to provide the IDs of the exchange fluxes of the metabolites we want to study in DyMMM. Thus, the exchange reaction of glucose and xylose IDs were provided.

Regarding initial reactor values, the initial biomass concentration of each microbe was set to 0.05 g/L and both dilution rates in and out of the reactor were constrained to 0 h^{-1}. Nevertheless, DyMMM requires metabolites concentration to be in *mmol* units. Thus, the previously used glucose and xylose concentration values of 16 and 8 g/L were converted to 88.8 and 53.28 mmol, respectively.

Table 10: Uptake parameters used to simulate the dynamic model for both case studies. $v_{g,max}$ and $v_{z,max}$ (mmol/gdw/h) are the maximum uptake rates for glucose, and xylose. K_g and K_z are saturation constants for glucose and xylose. $K_{i,e}$ (g/L) represents the ethanol inhibition constant.

Parameter	*E. coli* ZSC113	*S. cerevisiae*	*E. coli*
$\mathbf{V_{g,max}}$	-	25.9 or 10*	10*
$\mathbf{V_{z,max}}$	9	-	-
$\mathbf{K_g}$	-	0.5	0.5**
$\mathbf{K_z}$	0.01	-	-
$\mathbf{K_{i,e}}$	8	10	8

* $V_{g,max}$ was set to a value of 10 mmol/gDW/h specifically for this analysis purpose, not regarding experimental values.
** K_g was set to a value of 0.5 g/L mmol/gDW/h specifically for this analysis purpose, not regarding experimental values.

Similar to DFBAlab, ODEs are used to simulate the variation of biomass and metabolite concentration over time. Nevertheless, DyMMM automatically defines all the equations. Thus, equations 14 and 15 will not account for ethanol inhibition and the ethanol inhibition effect over glucose and xylose uptake from the medium was not considered.

Finally, the tool defines biomass yield maximization as the objective function in both organisms.

3.4 PERFORMANCE ASSESSMENT

For case studies A to E, static FBA approaches (pFBA - using COBRA Toolbox and micom's community models - and SteadyCom) results were compared with each other. Specifically,

- The structure of the different created communities;
- The individual organisms' metabolite uptake and secretion from and into the community environment;
- Information regarding metabolite cross-feeding predictions, was qualitatively compared to information available in literature.

Regarding case studies F and G, dynamic simulations, DFBAlab and DyMMM's performances were compared in terms of metabolite secretion/consumption and biomass yield dynamics.

Finally, SMETANA's performance was assessed for case studies A, B, C, F and G.

4

RESULTS AND DISCUSSION

The different tools' results regarding community reaction fluxes, growth rates and inter-species interactions were compared to available data from previously published studies.

4.1 Case Study A

The co-growth of four *E. coli* mutant strains was studied. As aforementioned, four different strains were considered:

- Ec1: auxotrophic for L-lysine and L-methionine, unable to export L-phenylalanine;
- Ec2: auxotrophic for L-arginine and L-phenylalanine, unable to export L-methionine;
- Ec3: auxotrophic for L-arginine and L-phenylalanine, unable to export L-lysine;
- Ec4: auxotrophic for L-lysine and L-methionine, unable to export L-arginine.

Since each mutant produces amino acids for the other mutants, growth from all strains is required due to cross-feedings. Thus, the following fluxes were analysed:

- Ec1's secretion of L-arginine and consumption of L-lysine and L-methionine;
- Ec2's secretion of L-lysine and consumption of L-arginine and L-phenylalanine;
- Ec3's secretion of L-methionine and consumption of L-arginine and L-phenylalanine;
- Ec4's secretion of L-phenylalanine and consumption of L-lysine and L-methionine.

Additionally, as the main carbon source is glucose, its consumption by each individual strain was also analysed.

Community models were generated in MATlab and Python, using COBRA Toolbox and micom's methods, respectively.

The resulting models topological information, available in Table 11, shows that both models have the same number of reactions and metabolites. However, the community model created with the COBRA Toolbox does not contain information regarding community genes.

Table 11: Topological information regarding the community models created for the four *E. coli*'s case study. Both micom and COBRA Toolbox were used. The number of genes is not available (N/A) in the COBRA Toolbox's model format.

	micom	**COBRA Toolbox**
Genes	1102	N/A
Reactions	9831	9831
Metabolites	6971	6971

4.1.1 *Computational results*

Results regarding the individual strains' growth in the community, glucose consumption and and amino acids cross-feeding, namely L-phenylalanine, L-arginine, L-methionine and L-lysine, were analysed.

- **pFBA in MATlab**

The FBA simulations performed in MATlab predicted a community growth rate of 0.572 h^{-1}. Such value is given by the sum of the individual organisms' growth rate.

The calculated fluxes regarding biomass, amino acids and glucose production and consumption are described in Table 12. It was predicted that strain 1 and 4 did not display growth under the specified environmental conditions, whereas strain 2 and 3 displayed a growth rate of 0.528 and 0.044 h^{-1}, respectively.

Table 12: Results regarding a pFBA performed in MATlab, for a community composed of four *E. coli* mutant strains (Ec1, Ec2, Ec3 and Ec4). Community and individual biomass formation rates (h^{-1}) are described, in addition to the studied amino acids and glucose production/consumption rates (mmol/gDW/h). A negative flux represents consumption of the metabolite, while a positive flux its production. For each flux, + and - represent expected secretion and consumption of the metabolite, respectively; * represents an expected reaction flux of 0 mmol/gDW/h. N/A represents the lack of the reaction in the respective GSM model.

Exchange reaction	**Community**	**Ec1**	**Ec2**	**Ec3**	**Ec4**
Biomass	0.572	0/+	0.528/+	0.044/+	0/+
L-phenylalanine	0	0/*	-0.093/-	-0.008/-	0.101/+
L-arginine	0	0.169/+	-0.156/-	-0.013/-	0/*
L-methionine	0	0/-	0/*	0/+	0/-
L-lysine	0	0/-	0/+	0/*	0/-
Glucose	-8	-0.678/-	-5.500/-	-0.988/-	-0.834/-

The obtained results predict that neither Ec1 nor Ec4 grow under the specified conditions. Additionally, Ec2 and Ec3 are auxotrophic for of L-arginine and L-phenylalanine,

which are secreted by Ec1 and Ec4, respectively. Even though it is not growing, Ec1 is still secreting L-arginine at a rate of 0.169 mmol/gDW/h and consuming glucose at 0.678 mmol/gDW/h, while Ec4 is secreting L-phenylalanine and consuming glucose at rates of 0.101 mmol/gDW/h and 0.834 mmol/gDW/h, respectively.

Even though these results do not match what was theoretically expected, they are biologically feasible, as metabolically active resting (nongrowing) *E. coli* cells have been shown to secrete metabolites in previous studies (43).

- **SteadyCom**

SteadyCom's simulations predicted growth from all *E. coli* strains, resulting in a total community growth rate of 0.735 h^{-1}. Specifically, strain 1, 2, 3 and 4's growth rate was 0.186, 0.239, 0.136 and 0.174 h^{-1}, respectively. The obtained fluxes for biomass, amino acids and glucose production and consumption are described in Table 13.

Table 13: SteadyCom simulation results regarding a community composed of four *E. coli* mutant strains (Ec1, Ec2, Ec3, Ec4). Community and individual biomass formation rates (h^{-1}) are described, in addition to the studied amino acids and glucose production/consumption rates (mmol/gDW/h). A negative flux represents consumption of the metabolite, while a positive flux represents its production.

Exchange reaction	Community	Ec1	Ec2	Ec3	Ec4
Biomass	0.735	0.186/+	0.239/+	0.136/+	0.174/+
L-phenylalanine	0	0/*	-0.042/-	-0.024/-	0.066/+
L-arginine	0	0.111/+	-0.071/-	-0.040/-	0/*
L-methionine	0	-0.029/-	0/*	0.056/+	-0.027/-
L-lysine	0	-0.064/-	0.124/+	0/*	-0.060/-
Glucose	-8	-2.026/-	-2.597/-	-1.480/-	-1.897/-

SteadyCom's results agree with what was expected when designing the community. Each organism is secreting one amino acid while consuming the ones it is auxotrophic for. For example, Ec1 is consuming L-methionine and L-lysine, while secreting L-arginine, which in turn is consumed by Ec2 and Ec3, that are auxotrophic for L-arginine.

It is important to note that no amino acids were supplied to the community. Thus, all amino acid consumption by the individual strains comes from its availability in the environment, granted by the strains who are able to secrete it. This co-dependency is in agreement with the presented results.

- **pFBA in Python**

Using pFBA as a simulation method, under the specified environmental constraints, micom's model did not display any growth. Specifically, each of the strains, Ec1, Ec2, Ec3 and Ec4's growth rates were 0 h^{-1}.

Thus, essential exchange reactions, which indicate which metabolites are required in the medium for the community to grow, were found by creating a function in Python, where:

- Each metabolite was individually removed from the medium, by setting the lower bound of the respective exchange reaction to 0. The remaining exchange reactions' lower bounds were set to -1000, to ensure that the blocked metabolite is the only limiting factor in the simulation;
- For each removal, a pFBA simulation was performed, with the maximization of community growth as the objective function;
- If the performed pFBA's solution was 0 h^{-1}, the reaction, and consequently the metabolite, was found to be essential.

Nevertheless, the analysis revealed that all metabolites which presence is mandatory for growth were already supplied to the community environment.

- **SMETANA**

As SMETANA requires the input of medium information to calculate possible interactions in the consortium, the metabolites present in Table 4 were introduced. Nevertheless, the studied amino acids (L-lysine, L-methionine, L-phenylalanine and L-arginine) were not provided as they are not supplied to the community but produced by the different strains when growing in co-culture.

Given the GSM models and environmental conditions, even though cross-feeding of the four studied amino acids was expected, SMETANA was not able to calculate any interactions between the organisms.

SMETANA does not run simulations, it calculates metrics based on the reactions the GSM models contain (95). As the studied auxotrophies are represented as constraints in the models, the four models contain the exact same reactions. It is likely that SMETANA would require the removal of the deactivated reactions to properly represent the mutations in SMETANA.

4.1.2 *Performance assessment*

It is important to note that pFBA and SteadyCom's simulations in MAtlab, were performed using the same community model, created by COBRA Toolbox's methods.

In pFBA's simulation, growth from two organisms in the community was predicted, which resulted in a total community growth rate of 0.572 h^{-1}. The objective function was set as the maximization of the community growth rate. Thus, the higher solution for pFBA

is given by the growth of two organisms. Nevertheless, such could not happen in SteadyCom, as the tool imposes a community steady-state condition in whichan organism cannot secrete any amino acid if it is not growing simultaneously.

SteadyCom's results predicted a total community growth rate of 0.735 h^{-1}, given by the sum of the four individual growth rates. The discrepancy between this value and the pFBA's is a result of the addition of a different constraint in SteadyCom. While in pFBA the predicted flux distribution needed to fulfill the ATPM requirement for four units of biomass (one for each *E. coli* strain), in SteadyCom it needs to satisfy the ATPM requirement for one unit of biomass instead.

Finally, the pFBA simulation using micom's model did not predict growth from any of the organisms.

On a concluding note, considering the three used methods, SteadyCom's performance was the most accurate, when comparing the results to what was theoretically expected (8).

4.2 Case Study B

Merging the individual models of *Bacteroides caccae* and *Lactobacillus rhamnosus* GG resulted in a new community model, which comprises a common space through which feeding and exchange of metabolites occurs, and a separate extracellular space for each model. This joint model created in COBRA Toolbox, consists of 4138 reactions and 3090 metabolites while micom's joint model is composed of 4136 reactions and 3089 metabolites, as described in Table 14.

Such difference in the number of genes and metabolites is caused by the consideration of *biomass* as a community metabolite in COBRA Toolbox's model. Thus, the respective community exchange reaction for the metabolite was also created. Additionally, micom's model lacks the presence of *octadecanoate* and its respective exchange reaction in the LGG compartment of the community.

The community model created by micom is composed of 1432 genes, while the community model generated by COBRA toolbox does not have information regarding genes. Only reaction and metabolites information is used to build the stoichiometric matrix, which is used to run the simulations.

Table 14: Topological information regarding the LGG and *B. caccae*'s community models created with micom or COBRA toolbox functions. N/A represents non-available information.

	micom	**COBRA Toolbox**
Genes	1432	N/A
Reactions	4136	4138
Metabolites	3089	3090

Previous *in silico* results (57) showed possible cross-feeding between LGG and *B. caccae* (Figure 6), where LGG provides lactate to *B. caccae* and the latter supplies LGG with L-alanine.

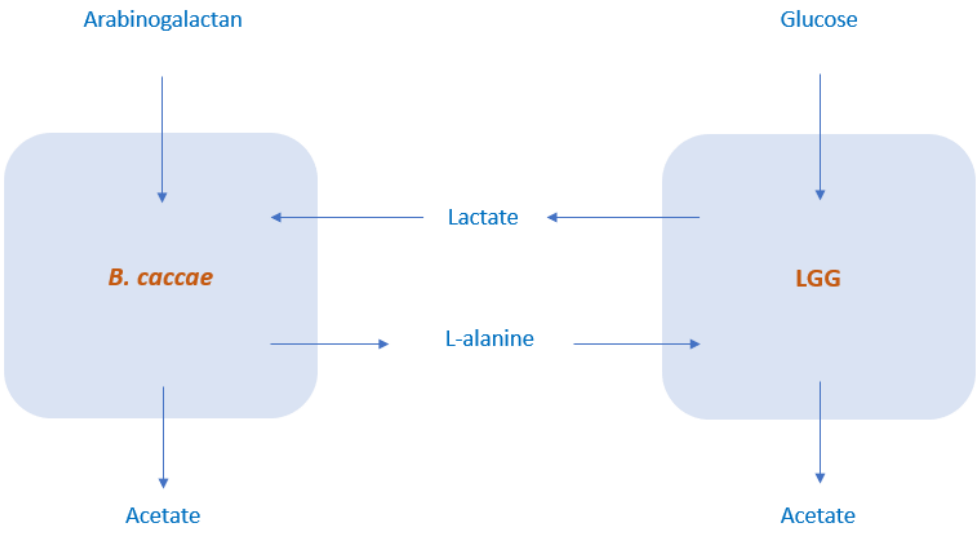


Figure 6: *B. caccae* and LGG interactions when growing together on DMEM 6429 medium supplemented with arabinogalactan.

This interaction takes place when glucose is present in the medium, which is metabolized in glycolysis, resulting in pyruvate production by means of "pyruvate kinase" (EC 2.7.1.40) (52; 53). *Lactate Dehydrogenase (LDH)* (EC 1.1.1.27) then converts pyruvate into lactate, which is released by LGG. When lactate is available, *B. caccae* strains are able to metabolize it to produce more pyruvate, through the activity of the enzyme LDH (96). Subsequently, pyruvate can be metabolized by *B. caccae*.

Reactions representing LDH's activity, as well as exchange reactions for lactate, were identified in *B. caccae*'s model, thus correctly representing such process.

In vivo results have shown that *B. caccae* is able to secrete L-alanine. However, its consumption by LGG was only predicted *in silico*, as L-alanine is an essential component of the biomass formation reaction in LGG' model.

Pyruvate is partly metabolized into acetyl-*Coenzyme A (CoA)* and formate by *Pyruvate-formate Lyase (PFL)* (EC 2.3.1.54). Formate is metabolized as a percursor for purine biosynthesis, while acetyl-CoA is metabolized, leading to acetate and energy (in the form of ATP) production. The resulting acetate, is then secreted by both *B. caccae* and LGG.

Arabinogalactan is degraded into three different carbon-based molecules: galactose, L-arabinose and glucoronate, which can subsequently be used as carbon sources.

4.2.1 *Computational results*

Results regarding the individual strains growth in the community, consumption and secretion of the studied metabolites are presented and analysed next.

- **pFBA in MATlab**

With the maximization of both organisms biomass formation reactions as the objective function, FBA simulations using the community model created by COBRA Toolbox (Table 15) predicted growth from both strains. Specifically, the predicted biomass formation by *B. caccae* and LGG is 0.815 and 0.172 h^{-1}, respectively, resulting in a total community growth rate of 0.987 h^{-1}.

Regarding carbon sources, LGG consumes all the glucose supplied to the medium, while *B. caccae* consumes arabinogalactan.

The simulation also predicted acetate secretion by both organisms, resulting in its accumulation in the community common compartment, which consequently represents its accumulation in the environment.

Additionally, the simulated production of L-alanine by *B. caccae* and its consequent consumption by LGG, agrees with the initial predictions. As all the secreted L-alanine is consumed, np accumulation of the metabolite in the community's common compartment occurs.

Finally, the cross-feeding of lactate is predicted, as *B. caccae* is consuming the metabolite at the same rate LGG is secreting it.

Table 15: pFBA simulation results for *B. caccae* and LGG's community, performed using COBRA Toolbox's model. Results regarding biomass formation (h^{-1}), L-alanine, arabinogalactan, glucose, acetate and lactate exchange fluxes (mmol/gDw/h) in the community are considered. A negative flux represents consumption of the metabolite and a positive flux represents its production. For each flux, + and - represent expected secretion and consumption of the metabolite, respectively; * represents an expected reaction flux of 0 mmol/gDW/h. N/A represents the lack of the reaction in the respective GSM model.

Exchange reaction	Community	*B. caccae*	LGG
Biomass	0.987	0.815/+	0.172/+
L-alanine	0	0.084/+	-0.084/-
Arabinogalactan	-0.094	-0.094/-	N/A
Glucose	-4.5	0	-4.5/-
Acetate	16.484	16.313/+	0.171/+
Lactate	0	-10.372/-	10.372/+

- **SteadyCom**

SteadyCom was not able to calculate a steady-state solution for this community. Thus, the following steps were followed to identify the cause of the problem:

- Set all community and individual organisms' upper and lower bounds to 1000 and -1000, respectively;
- Setting the ATP maintenance (ATPM) reaction's lower bounds to 0, to understand if the problem lied in this reaction.

As aforementioned, SteadyCom requires adding different variables to the community structure, to run a simulation. Specifically, the extracellular metabolites, the corresponding exchange reactions and the uptake rates have to be specified. The definition of such variables was performed following the authors' supplementary code regarding their published case study. Nevertheless, the code was not robust enough to work correctly in the present case study, resulting in an unfeasible solution.

- **pFBA in Python**

With the maximization of both organisms' growth reactions as the objective function, pFBA simulations using the community model created by micom, predicted growth from both strains. Specifically, the predicted biomass formation by *B. caccae* and LGG is 0.823 and 0.188 h^{-1}, respectively, resulting in a total community biomass of 0.506 h^{-1}. The calculated production and secretion values are present in Table 16.

Table 16: pFBA simulations for *B. caccae* and LGG, performed using micom's community model. Results regarding biomass formation (h^{-1}), L-alanine, arabinogalactan, glucose, acetate and lactate exchange fluxes (mmol/gDw/h) in the community are described. A negative flux represents consumption of the metabolite and a positive flux represents its production. For each flux, + and - represent expected secretion and consumption of the metabolite, respectively; * represents an expected reaction flux of 0 mmol/gDW/h. N/A represents the lack of the reaction in the respective GSM model.

Exchange reaction	Community	*B. caccae*	LGG
Biomass	1.011	0.823/+	0.188/+
L-alanine	0	0.091/+	-0.091/-
Arabinogalactan	-0.001	-0.002/-	N/A
Glucose	-4.5	-4.5	-4.5/-
Acetate	3.433	6.375/+	0.492/+
Lactate	0	-12.352/-	12.352/+

The community's internal exchange fluxes take the abundance of the respective submodel into account. Thus, in this two-organisms consortium, micom scales each strain's internal fluxes to 50%.

L-alanine and lactate's cross-feeding between the two organisms was correctly predicted by micom's model, as LGG is consuming L-alanine at the same rate *B. caccae* is secreting it.

Additionally, glucose is consumed by the whole community, by means of the common compartment, at a rate of 4.5 mmol/gDW/h, while both individual organisms are consuming it at a rate of 4.5 mmol/gDW/h. Specifically, both organisms are consuming it at the same rate and the following mathematical calculation is performed: 4.5 x 0.5 + 4.5 x 0.5 = 4.5. The final result represents the total community's uptake.

Similarly, consumption of arabinogalactan by the community is 0.001 mmol/gDW/h, while *B. caccae* is consuming 0.002 mmol/gDW/h, which agrees with the previous explanation.

Finally, acetate is secreted by both organisms. The final community' secretion is 3.433 mmol/gDW/h, which is given by the mean of both organisms' secretion values (6.375 and 0.492 mmol/gDW/h).

- **SMETANA**

SMETANA determines what metabolites might be produced by one organism and consumed by another in the community, creating a dependency between the organisms. Its results for *B. caccae* and LGG's interactions, with DMEM 6429 medium supplemented with arabinogalactan (57) as environmental conditions, are described in Table 17.

Table 17: Metrics calculated by the SMETANA's framework. *Receiver* is the organism who benefits from the production of a specific *Metabolite* from another organism in the community, the *Donor*. SMETANA is a score that indicates how feasible the interaction is. Its value varies between 0 and 1.

Receiver	Donor	Metabolite	SMETANA
LGG	*B. caccae*	L-alanine	1.0
LGG	*B. caccae*	Fructose	0.47
LGG	*B. caccae*	Galactose	0.35
LGG	*B. caccae*	Nicotinate	1.0

The highest SMETANA score concerns the cross-feeding of L-alanine, which, as aforementioned, can be secreted by *B. caccae* and is mandatorily consumed by LGG, due to the presence of the metabolite in the respective GSM model biomass formation reaction.

Aditionally, SMETANA predicted a dependency regarding nicotinate. The metabolite is not present in medium information introduced in SMETANA. Nevertheless, in this section's previously described simulations, it was supplied to the medium as it was required for minimal growth of LGG. Thus, if nicotinate is not supplied in the medium and *B. caccae* is able to produce it, a possible cooperative interaction can occur. Such interaction was predicted *in silico* by Magnúsdóttir *et al.* (2017).

Additionally, fructose and galactose were also identified as possible cross-feeding interactions between the organisms. Such is due to the fact that *B. caccae* can metabolize them after the breakdown of arabinogalactan, as previously explained. Xylose isomerase is present in *B. caccae*'s GSM model. Therefore, it has the potential to metabolize glucose into fructose, which explains the obtained result by SMETANA.

SMETANA did not predict a dependency regarding lactate cross-feeding. However, such lack of dependency does not eliminate the potential benefit of this interaction, as aforementioned.

4.2.2 *Performance assessment*

pFBA was performed using COBRA Toolbox and micom's models. Even though both simulations agreed with what was theoretically expected, they follow different mathematical representations of a community. While the first considers the community as a sum of the individual members secretion/consumption values, micom contemplates the community fluxes as mean values of both individuals.

Both tools predicted the expected metabolites cross-feeding, secretion and consumption. Nevertheless, the flux values differ. As both approaches follow different calculations, one can only state that one performs better or worse than the other if experimental values are used as reference.

Additionally, SteadyCom was unable to find a feasible solution for the present case study, which might indicate a possible lack of universality by the tool.

Regarding SMETANA, even though not being a simulation method, and therefore its results cannot be compared with the other methods' results, it still proved to be useful as it indicated possible and realistic metabolic interactions.

4.3 **Case Study C**

Combining the individual GSM models of *Streptococcus themophilus* LMG18311 and *Lactobacillus bulgaricus* using micom and COBRA Toolbox's methods resulted in two different community models for the present case study. These models differ in number of genes, reactions and metabolites (Table 18). COBRA Toolbox's model comprises 2185 reactions and 2133 metabolites, while micom's model comprises 688 genes, 2144 reactions and 1932 metabolites. Information regarding genes is not present in COBRA Toolbox's model.

The number of reactions and metabolites in the two models does not coincide. COBRA Toolbox's model has an extra metabolite, which is community biomass. Thus, a community exchange reaction for biomass was created. Additionally, *S. thermophilus* GSM model has extracellular metabolites for which no exchange reaction exists. Thus, when generating a

community model, COBRA Toolbox creates a new community exchange reaction for those metabolites.

On the other hand, for a specific metabolite, micom creates community exchange reactions and transport reactions between the common compartment and the individual organisms' compartments exclusively for the existing exchange reactions in the original individual GSM models. Therefore, the difference between both procedures explains the mismatch in the number of reactions between the community models.

Table 18: Topological analysis of the *S. thermophilus* and *L. bulgaricus* community models created with micom and COBRA toolbox functions. N/A represents non-available information.

	micom	**COBRA Toolbox**
Genes	688	N/A
Reactions	2144	2185
Metabolites	1932	1933

The *L. bulgaricus* and *S. thermophilus* consortium is characterized by mutualistic interactions between the two organisms in milk fermentation (81; 83). A schematic representation of these interactions is available in Figure 7.

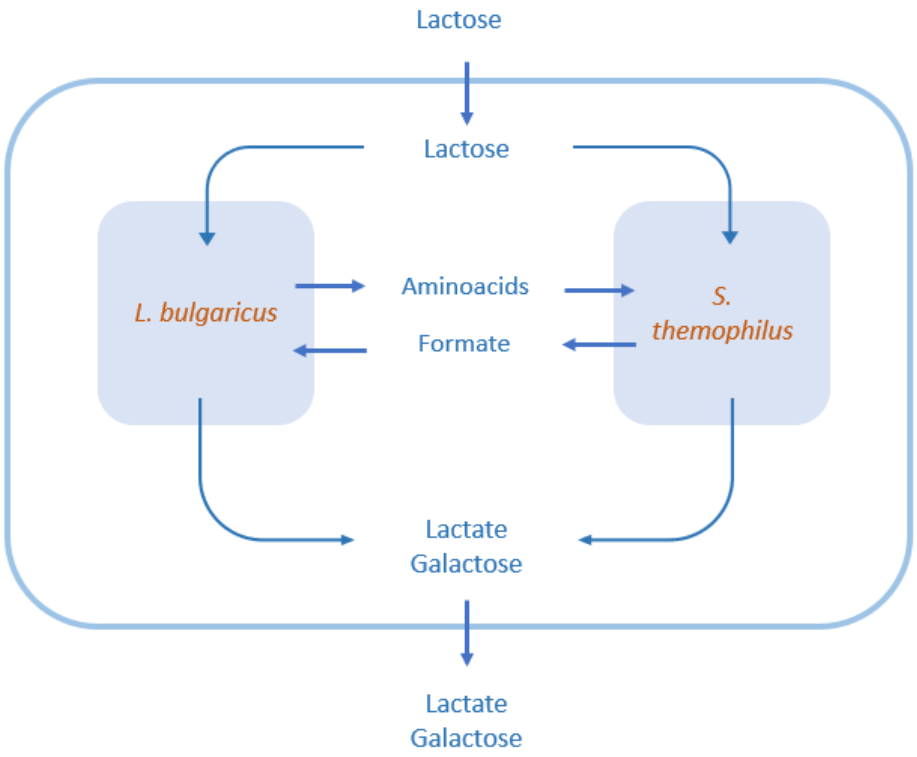


Figure 7: *L. bulgaricus* and *S. thermophilus* interactions when growing together on milk.

As previously explained, *L. bulgaricus* has proteolytic activity, result of the presence of a cell wall-associated proteinase, *PrtB*, which enables the degradation of caseins (39).

S. thermophilus LMG18311 cannot degrade proteins such as casein and is auxotrophic for L-histidine (66), requiring its presence in the environment in the form of a free amino acid. Due to *L. bulgaricus*'s proteolytic activity, the organism can degrade casein into smaller peptides and supply *S. thermophilus* with such peptides, which can be hydrolysed into amino acids, such as L-histidine. Thus, the cross-feeding of a "milk peptide", as a result of its secretion by *L. bulgaricus*, was analysed.

Additionally, the cross-feeding regarding production of formate by *S.thermophilus* and its consumption by *L. bulgaricus* has been indicated by experimental data (32).

S. thermophilus' model has a reaction representing PFL activity. Thus, by using pyruvate and CoA, acetyl-CoA and formate are produced. However, *L. bulgaricus* lacks PFL (84), relying on formate secretion by *S. thermophilus*.

Formate is involved as a cofactor in purine biosynthesis (84). Thus, when formate is supplied by *S. thermophilus*, by using "formate-tetrahydrofolate ligase" (EC 6.3.4.3), formate, ATP and tetrahydrofolate are consumed. Consequently, ADP, phosphate and 10-Formyltetrahydrofolate are produced. The latter metabolite is then used in purine biossynthesis, by activity of "phosphoribosylaminoimidazolecarboxamide formyltransferase" (EC 2.1.2.3).

Both organisms consume lactose, which is degraded into glucose and galactose. Glucose goes through glycolysis, being converted into pyruvate. Subsequently, pyruvate is metabolized into lactate by LDH. As lactate cannot be further metabolized, it is secreted by both organisms. Galactose is used in antiport transport of lactose by *L. bulgaricus* and *S. thermophilus* (2; 9). Subsequentely, its secretion is expected, as both organisms cannot grow on it.

4.3.1 *Computational Results*

Results regarding the individual strains' growth in the community, consumption and secretion of the studied metabolites are presented and analysed here.

- **pFBA in MATlab**

pFBA results, using the community model created using COBRA Toolbox's methods predicted growth from both organisms, resulting in a community growth rate value of 0.157 h^{-1}. Specifically, *S. thermophilus* and *L. bulgaricus* biomass production rate is 0.050 and 0.107 h^{-1}, respectively, as represented in Table 19.

L. bulgaricus' proteolytic activity is correctly predicted. Casein is supplied to the community by means of the common compartment. *L. bulgaricus* begins by consuming it from the

Table 19: pFBA simulation results, regarding the *S. thermophilus* and *L. bulgaricus*'s community model created using COBRA Toolbox. Results regarding biomass formation (h^{-1}), casein, milk peptide, lactose, galactose, lactate and formate's exchange fluxes (mmol/gDw/h) in the community are described. A negative flux represents consumption of the metabolite and a positive flux represents its production. For each flux, + and - represent expected secretion and consumption of the metabolite, respectively; * represents an expected reaction flux of 0 mmol/gDW/h.

Exchange reaction	Community	*L. bulgaricus*	*S. thermophilus*
Biomass	0.157	0.050/+	0.107/+
Casein	-10	-10/-	0/*
Milk peptide	9.995	9.999/+	-0.003/+
Lactose	-10	-5/-	-5/-
Galactose	10	5/+	5/+
Lactate	19.366	10.277/+	9.089/+
Formate	0.375	-0.044/-	0.418/+

community environment at a rate of 10 mmol/gDW/h. Subsequently, the breakdown of casein results in the synthesis of a "milk peptide", which is secreted into the community environment at a rate of 9.999 mmol/gDW/h, while 0.001 mmol/gDW/h are consumed by the organism.

It is important to note that the described proteolytic activity is the only source of milk peptide in the community environment, as it is not directly supplemented in the community environment.

Given the presence of milk peptide in the common compartment, *S. thermophilus* can consume it and further degrade it into amino acids. Thus, *S. thermophilus* is consuming 0.003 mmol/gDW/h of milk peptide, resulting in the accumulation of 9.995 mmol/gDW/h in the community common compartment.

Regarding the carbon source consumption, 5 mmol/gDW/h of lactose are consumed by both organisms. As a result, 5 mmol/gDW/h of galactose are secreted by both *S. thermophilus* and *L. bulgaricus*. It is important to note that the consumption of lactose and secretion of galactose's values should be the same, as galactose is expected to be used by antiporter in exchange for lactose intake (2).

As aforementioned, *L. bulgaricus* uses galactose in antiport for lactose (2). Thus, to correctly model galactose secretion, the upper bound of the "galactose kinase" (EC 2.7.1.6) reaction was constrained to 0 in *L. bulgaricus*' model. The lack of such constraint would lead to an inaccurate usage of galactose as carbon source by the organism.

As expected, both organisms secrete lactate. *L. bulgaricus* and *S. thermophilus* calculated secretion lactate fluxes are 10.277 and 9.089 mmol/gDW/h, respectively. As a result, the total lactate secretion by the community is 19.366 mmol/gDW/h.

Formate production and its consequent secretion by *S. thermophilus* is predicted at a rate of 0.418 mmol/gDW/h. Subsequently, 0.044 mmol/gDW/h of the metabolite are consumed by *L. bulgaricus*.

In the original *L. bulgaricus*' model, the *Methenyltetrahydrofolate Cyclohydrolase (MTHFC)* (EC 3.5.4.9) reaction was unconstrained, which resulted in 10-Formyltetrahydrofolate production without the metabolization of formate. Thus, to simulate *L. bulgaricus*' consumption of formate, this reaction's upper bound of was constrained to 0.

As *L. bulgaricus* is not consuming all of the available formate, its secretion by means of the common compartment is predicted at a rate of 0.375 mmol/gDW/h.

- **SteadyCom**

SteadyCom was not able to obtain a feasible solution for this community. Different steps were followed to determine the origin of the problem, as performed in section 4.2.1. Similar to that section's results, the lack of universality regarding the variables of the community model, resulted in an unfeasible solution.

- **pFBA in Python**

pFBA simulations using micom's community model predicted growth from both organisms in the community. The predicted growth rates for *L. bulgaricus* and *S. thermophilus* are 0.149 and 0.294 h^{-1}, respectively, resulting in a total community growth rate of 0.222 h^{-1}. Additionally, predictions regarding consumption and production of the studied metabolites are described in Table 20.

The obtained predictions show that *S. thermophilus*'s secretes 4.236 mmol/gDW/h of formate, which is not consumed by *L. bulgaricus*. As a result, the total community's secretion of formate is 2.118 mmol/gDW/h.

Similar to the what was performed in MATlab, using COBRA Toolbox's model, the reaction MTHFC was constrained. Such alteration resulted in an infeasible solution, making it necessary to keep the reaction unconstrained.

Regarding the lactate secretion, only *L. bulgaricus* is secreting it while *S. thermophilus* is neither consuming or secreting the metabolite. Such is possible as the latter organism is secreting acetate, as a result of pyruvate metabolization.

To simulate PFL activity in *S. thermophilus*, the upper bound of the acetate exchange reaction in *S. thermophilus* was constrained to 0, limiting its secretion. Nevertheless, such alteration resulted in an infeasible solution, due to the model's inability to maintain a steady-state. Thus, the reaction was kept unconstrained.

Similar to what was described in the previous simulation, both organisms consumption of lactose is 5 mmol/gDW/h. As a result, 5 mmol/gDW/h of galactose should be secreted by *S. thermophilus* and *L. bulgaricus*. However, only *S. thermophilus* is secreting galactose.

Galactose kinase's activity was not constrained in *L. bulgaricus* as such constrain would result in an infeasible solution.

It is also important to note that a total of 5 mmol/gDW/h of lactose are consumed by the community. This value is given by the mean of the individual organisms' lactose consumption.

Table 20: pFBA simulation results, regarding the *S. thermophilus* and *L. bulgaricus*'s community model created using micom. Results regarding biomass formation (h^{-1}), casein, milk peptide, lactose, galactose, lactate, acetate and formate's exchange fluxes (mmol/gDw/h) in the community are described. A negative flux represents consumption of the metabolite and a positive flux represents its production. For each flux, + and - represent expected secretion and consumption of the metabolite, respectively; * represents an expected reaction flux of 0 mmol/gDW/h.

Exchange reaction	Community	*L. bulgaricus*	*S. thermophilus*
Biomass	0.222	0.149/+	0.294/+
Casein	-0.002	-0.004/-	0/*
Milk peptide	0	0/*	-0.00066/-
Lactose	-5	-5/-	-5/-
Galactose	2.5	0/+	5/+
Lactate	10.227	20.455/+	0/+
Acetate	5.252	0/*	10.504/*
Formate	2.118	0/-	4.236/+

S. themophilus' individual GSM model lacked exchange reactions for L-lactate and NH_4. As micom does not create exchange reactions which are not already present in the original individual models, they have to be manually added if required. Specifically, the *S. thermophilus*' L-lactate exchange reaction was not present in the community model. As lactate is the main product of this organism's metabolis (66), the respective exchange reaction was manually added to the GSM model, to simulate its secretion by the organism, in the community.

Additionally, the lack of exchange reaction for NH_4 in the *S.thermophilus* compartment of the community model was causing an infeasible solution for pFBA, as the model was not able to maintain a steady-state due to the inability to secrete the metabolite, which was being produced inside the network, as a consequence of amino acid metabolism.

- SMETANA

SMETANA's metrics predict interactions between *L. bulgaricus* and *S. thermophilus* under the specified environmental conditions (Table 21). A medium simulating milk was defined. Thus, lactose, casein, biotin and ions were supplied to the community.

The SMETANA score was calculated for all the calculated interspecies interactions. Its maximum value is 1, and it represents how important that interaction is for the survival of the "Receiving" species (i.e. *S. thermophilus*).

An interaction regarding "milk peptide" cross-feeding was calculated. *L. bulgaricus* is the donor and *S. thermophilus* the consumer. The respective interaction's SMETANA score is 1, which indicates that the latter organism depends on such exchange to grow. This result agrees with what is expected. Due to *L. bulgaricus*' proteolytic activity, casein is degraded into milk peptide, which is subsequently used by *S. thermophilus* to produce amino acids, required for its growth.

It is possible that *L. bulgaricus* secretes L-arginine, L-serine and/or L-threonine after degrading the milk peptide. Nevertheless, *S. thermophilus* has the equivalent ability to degrade the peptide. Therefore, *S. thermophilus*' growth does not depend on this interaction. The calculated SMETANA score for the interactions regarding the three amino acids is 0.25. Such score indicates that the interactions are not binding, which corroborates what is theoretically expected.

Table 21: SMETANA's metrics for the *L. bulgaricus* and *S. thermophilus*' community. *Receiver* is the organism who benefits from the production of a specific *Metabolite* from another organism in the community, the *Donor*. SMETANA is a score that indicates how feasible the interaction is. It varies from 0 to 1, where 0 represents no interaction and 1 represents a mandatory interaction.

Metabolite	Donor	Receiver	SMETANA
Milk peptide	*L. bulgaricus*	*S. thermophilus*	1
L-arginine	*L. bulgaricus*	*S. thermophilus*	0.25
L-serine	*L. bulgaricus*	*S. thermophilus*	0.25
L-threonine	*L. bulgaricus*	*S. thermophilus*	0.25

As aforementioned, SMETANA calculates possible interactions based on the reactions present in the GSM models. As it is not a simulation method, upper and lower bounds' constraints do not affect the result. Hence, an interaction resulting from formate consumption by *L. bulgaricus* is not predicted, as this was previously achieved by constraining MTHFC.

4.3.2 *Performance assessment*

SteadyCom requires the addition of variables to the community model. Nevertheless, the published process to add such variables does not work for every GSM model. Thus, the inability to find a feasible solution using SteadyCom's methods indicates a lack of universality from the tool.

Community exchange reactions and individual organisms' transport reactions, connecting the respective compartment to the common community compartment, are created by micom, basing on the individual GSM models' drains, while COBRA Toolbox creates these based on the individual GSM models' extracellular metabolites. That is, if a GSM model has a particular extracellular metabolite for which no exchange reaction exists, micom does not create a community exchange reaction, whereas COBRA Toolbox does.

As an organism's metabolism is different when it is part of a community (i.e. due to the presence of different compounds in the environment), the lack of certain exchange or transport reactions can lead to different results or infeasible solutions. Thus, COBRA Toolbox's community model performs better, as it provides transport and exchange reactions for all the metabolites.

SMETANA's metrics were also calculated. Its predicted interaction regarding the "milk peptide" exchange is corroborated by the other performed simulations and theoretical information.

4.4 Case Study D

Figure 8 illustrates the interaction between *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*. This potential interaction is based on the production of acetate by *Bifidobacterium* and its consumption by *F. prausnitzii* (19).

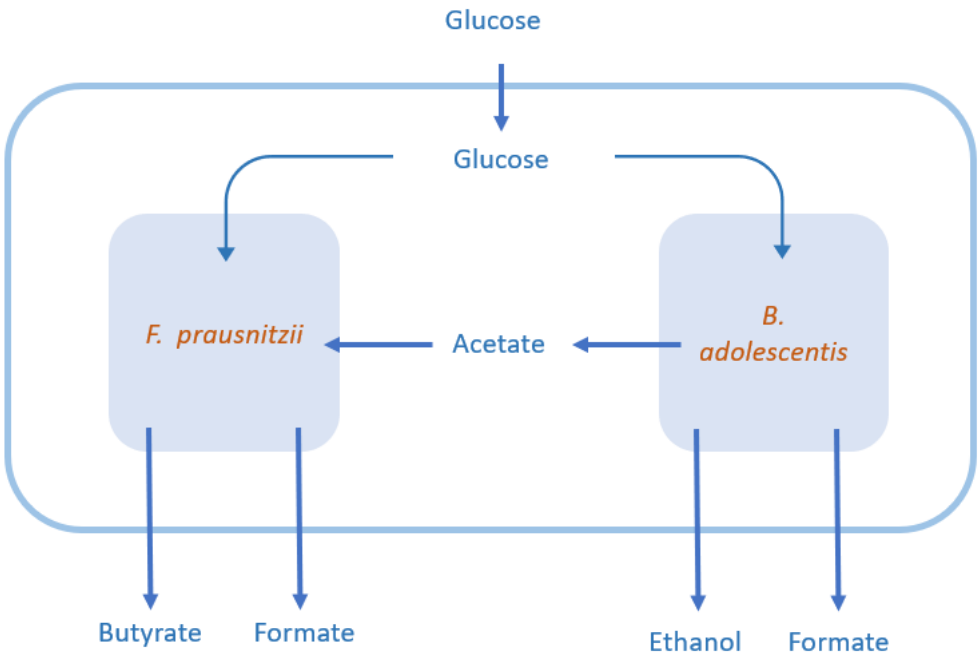


Figure 8: *B. adolescentis* and *F. prausnitzii* interactions when growing together on glucose.

Experimental data has shown that *B. adolescentis* produces formate, acetate and butyrate. In the presence of glucose in the medium, *Bifidobacterium* metabolizes it into glucose-6-phosphate, which is then metabolized into fructose-6-phosphate. The organism has the "fructose-6-phosphate phosphoketolase" enzyme (4.1.2.22), which subsequently converts fructose 6-phosphate into erythrose 4-phosphate and acetyl phosphate. The latter metabolite is consequently metabolized to acetate, generating ATP (19).

Additionally, the model is expected to regenerate NAD^+ through the production of ethanol (19), resulting in its secretion.

When acetate is present in the medium, *F. prausnitzii* can use it to secrete butyrate. It is normally formed from two molecules of acetyl-CoA, yielding acetoacetyl-CoA, which is then converted to butyryl-CoA (15). Finally, the GSM model converts butyryl-CoA into butyrate through "acetyl-CoA:butyrate-CoA transferase" (EC 2.8.3.4).

Community models were created using both COBRA Toolbox and micom's methods, and the results regarding the topology of these new models are described in Table 22.

While the community created using micom is composed of 1539 reactions and 1540 metabolites, the one created in COBRA Toolbox has 1752 reactions and 1568 metabolites. Such difference is explained by the fact that micom's community does not bear a common compartment, while COBRA Toolbox's does. Thus, exchange community reactions and transport reactions connecting the individual organisms to the common compartment were considered in the latter model's construction. The same explanation applies to the different number of metabolites, as community metabolites were added to COBRA Toolbox's model.

Table 22: Topological information regading the *B. adolescentis* and *F. prausnitzii* community models, created with micom and COBRA toolbox functions. N/A represents non-available information.

	micom	**COBRA Toolbox**
Genes	N/A	N/A
Reactions	1539	1752
Metabolites	1540	1568

4.4.1 *Computational results*

Results regarding the individual strains growth in the community, consumption and secretion of the studied metabolites were analysed and are presented next.

- **pFBA in MATlab**

pFBA simulations, performed using the community model created in COBRA Toolbox, show growth from the both strains. Specifically, *B. adolescentis* and *F. prausnitzii*'s growth

rate is 0.102 and 0.072 h^{-1}, respectively. In addition, the predicted consumption and secretion of metabolites in the *B. adolescentis* and *F. prausnitzii* consortium is described in Table 23.

Acetate cross-feeding is predicted as seen by the fact that its secretion by *B. adolescentis* is 2.802 mmol/gDW/h, followed by the consumption of exact same ammount by *F. prausnitzii*.

Regarding carbon source consumption, it is predicted that each organism is consuming 1 mmol/gDW/h of glucose, as expected.

Ethanol was expected to be secreted by *B. adolecentis*, as its production is a way of regenerating NAD^+. Nevertheless, that was not predicted as the model is regenerating the metabolite by production of glycerol, by the enzyme "glycerol dehydrogenase" (EC 1.1.1.6). Thus, in the present simulation, *B. adolescentis* is secreting 1.802 mmol/gDW/h of glycerol, instead of ethanol.

Additionally, *F. praunsnitzii* is secreting 3.228 mmol/gDW/h of butyrate. As it is not consumed by *B. adolescentis*, it is consequently secreted by the community.

Both organisms are secreting formate, as expected. Specifically, *B. adolescentis* and *F. prausnitzii*'s secretion rates are 1.622 and 3.489 mmol/gDW/h, respectively, summing to a total community secretion value of 5.120 mmol/gDW/h .

Table 23: pFBA simulation results for *B. adolescentis* and *F.prausnitzii*'s community, performed in MATlab. The present results describe biomass formation (h^{-1}) and metabolites exchanges rates (mmol/gDw/h). Acetate, ethanol, butyrate, formate and glucose are considered. A negative value represents consumption of a metabolite and a positive value represents its production. For each flux, + and - represent expected secretion and consumption of the metabolite, respectively; * represents an expected reaction flux of 0 mmol/gDW/h. N/A represents the lack of the reaction in the respective GSM model.

Exchange reaction	Community	*B. adolescentis*	*F. prausnitzii*
Biomass	0.174	0.102/+	0.072/+
Acetate	0	2.802/+	-2.802/-
Ethanol	0	0/+	N/A
Glucose	-2	-1/-	-1/-
Butyrate	3.228	N/A	3.228/+
Formate	5.120	1.622/+	3.489/+

- **SteadyCom**.

SteadyCom predicted that only *B. adolescentis* grows, with a growth rate of 0.148 h^{-1}. It consumes 1 mmol/gDW/h of glucose, while 6.397, 2.600 and 2.521 mmol/gDW/h of acetate, glycerol and formate are secreted.

As no growth was predicted for *F. prausnitzii*, no cross-feeding was simulated by SteadyCom.

- **pFBA in Python**

A common compartment, in which exchanges between the organisms occur, was not created in micom. Thus, the secretion and uptake of metabolites takes place directly in the extracellular compartment of the individual models.

The community common compartment is created, based on the individual organisms' exchange reactions, by micom.

Usually, an exchange reaction has a metabolite as a reactant and no products. An example of a drain's structure is presented in equation 16:

$$metabolite(extracellular) <=> \tag{16}$$

In the present GSM models, the directionality of these reactions changes and is not constant through the models. As micom relies on exchange reactions to create the community environment, the community problems arise from such inconsistency.

pFBA predicted growth for both organisms. Specifically, *B. adolescentis* and *F. prausnitzii* growth rates are 0.039 and 0.033 h^{-1}, respectively, resulting in a total community growth rate of 0.036 h^{-1}.

Production and secretion rates of metabolites (Table 24) all agree with what was theoretically expected. Namely, both organisms consume 1 mmol/gDW/h of glucose. Additionally, *B. adolescentis*' secretion of acetate, ethanol and formate is visible. The respective obtained secretion values are 1.732, 0.467, and 0.849 mmol/gDW/h. Additionally, part of the secreted acetate is consumed *F. prausnitzii* (1.387 mmol/gDW/h), which then secretes 1.602 and 1.754 mmol/gDW/h of butyrate and formate, respectively.

Table 24: pFBA simulations for *B. adolescentis* and *F. prausnitzii*'s community, performed using micom's model. Results regarding biomass formation (h^{-1}) acetate, ethanol, butyrate, formate and glucose exchange rates (mmol/gDw/h) in the community are presented. A negative rate represents consumption of a metabolite and a positive flux represents production. For each flux, + and - represent expected secretion and consumption of the metabolite, respectively; * represents an expected reaction flux of 0 mmol/gDW/h. N/A represents the lack of the reaction in the respective GSM model.

Exchange reaction	B. adolescentis	F. prausnitzii
Biomass	0.039/+	0.033/+
Acetate	1.732/+	-1.387/-
Ethanol	0.467/+	N/A
Butyrate	N/A	1.601/+
Formate	0.849/+	1.754/+
Glucose	-1/-	-1/-

- **SMETANA**

As SMETANA did not support the models' format, no metrics were obtained for the present community.

4.4.2 *Performance assessment*

SteadyCom and pFBA simulation results were obtained for the model created using COBRA Toolbox. Similarily, pFBA was also performed for the community model created by micom's methods.

SteadyCom's simulation predicted growth from only one organism in the consortium, not predicting cross-feeding between the organisms. Such result does not agree with what was theoretically expected, as both organisms had the metabolites required for its growth supplied in the community environment. It is possible that, for SteadyCom, the maximum community growth rate was achieved for *B. adolescentis* isolated growth.

Additionally, both pFBA simulations, performed using the different models, agreed with what was theoretically expected. Nevertheless, micom's model does not accurately represent a community environment, as it did not create a community common compartment. Such inaccuracy can be explained by the GSM models' structure. These models' exchange reactions direction is not consistent through the stoichiometric matrix, which caused the lack of a common compartment representation by micom.

Nevertheless, when comparing both created community models, COBRA Toolbox's model performs more realistically. COBRA Toolbox creates community common compartment based on the individual GSM models extracellular metabolites. Therefore, micom's dependence on the GSM models exchange reactions's structure points to the tool's lack of robustness.

Subsequently, more realistic results were obtained by the performed pFBA using COBRA Toolbox's community model.

SMETANA did not calculate any metrics, as the GSM models did not match any of its supported formats. Therefore, its performance was not assessed in this case study.

4.5 **Case Study E**

D. vulgaricus and *M. maripaludis* stoichiometric models (50) were used to create a community model in MATlab, using COBRA Toolbox's methods.

As aforementioned, the used models are only composed of one compartment. Thus, micom's methods were not available to create the respective community structure. Additionally, even though a community structure was created in MATlab, it was also not compatible with SteadyCom's simulation methods, due to the lack of representation of extracellular en-

vironment in the individual models. At last, SMETANA's metrics were also not calculated for the same reason.

Regarding the resulting community model, it comprises 216 reactions and 194 metabolites. It is important to note that no common compartment was created for this community and that the individual stoichiometric matrices were merged into a single one. Thus, community uptakes could not be constrained, as they were not represented. Only metabolite uptakes from the individual models could be constrained, directly in the internal reactions of the new community.

Using pFBA as simulation method, growth from both organisms was predicted. Specifically, *D. vulgaris* and *M. maripaludis*' growth rates are 1.819 and 0.138 h^{-1}, respectively, resulting in a total community growth rate of 1.957 h^{-1}.

Theoretically, it was expected that*D. vulgaris* would consume lactate and secrete acetate, carbon dioxide, hydrogen or/and formate, while *M.maripaludis* would consume hydrogen or/and formate, carbon dioxide, secreting methane (50). The respective production/secretion fluxes are described in Table 25.

Table 25: pFBA simulation results for *D. vulgaris* and *M. maripaludis*' community. Results regarding, biomass exchange (h^{-1}), lactate, acetate, CO_2, H_2, formate and methane exchange rates (mmol/gDw/h) are presented. A negative value represents consumption of a metabolite while a positive represents production. For each flux, + and - represent expected secretion and consumption of the metabolite, respectively; * represents an expected reaction flux of 0 mmol/gDW/h. N/A represents the lack of the reaction in the respective GSM model.

Exchange reaction	*D. vulgaris*	*M.maripalus*
Biomass	1.819/+	0.138/+
Lactate	0/+	0/*
Acetate	100/+	N/A
CO_2	100/+	-28.927/-
H_2	0/*	-100/-
Formate	7.790/+	0/*
Methane	N/A	23.229/+

Even though lactate consumption was not constrained, *D. vulgaris* did not use it as substrate. Nevertheless, acetate, CO_2 and formate were secreted. *M. maripaludis* consumes part of the secreted CO_2. Additionally, it is consuming H_2 even though it is not secreted by *D. vulgaris*. Such can be explained by the lack of a common compartment. Thus, as the lower-bound for H_2 exchange in *M. maripaludis* is unconstrained, it can consume H_2 , even if the other organism is not secreting it. Finally, *M. maripaludis* is secreting methane, as expected.

While COBRA Toolbox's methods showed a broader range regarding the creation of multiple models, the obtained structure does not accurately represent a community environment, as there is no compartment where cross-feeding between the organisms occurs.

4.6 Case Studies F and G

S. cerevisiae and *E. coli* co-culture was designed to study efficient aerobic consumption of glucose/xylose mixtures. As S. cerevisiae only consumes glucose and engineered *E. coli* strain ZSC113 only consumes xylose, the individual biomass and both glucose and xylose's dynamics were studied. Additionally, *S. cerevisiae* produces ethanol during glucose fermentation, which inhibits its own and *E. coli*'s growth. Thus, ethanol variation over time is also included in the results.

As aforementioned in section 3.3.6, iND750 and iJR904 GSM models were used in two different case studies to study the community's dynamics:

- The co-culture of *S. cerevisiae* and the engineered *E. coli* strain ZSC113;
- The co-culture of *S. cerevisiae* and a non-engineered *E. coli* strain.

Regarding the first case, glucose and xylose's concentration is expected to decrease over time, while ethanol's is expected to increase. Simultaneously, *S. cerevisiae* and *E. coli*'s biomass concentration values are expected to increase.

Regarding the second case study, only the glucose concentration is expected to decrease, as both organisms are consuming it. Thus, xylose concentration value should not vary. Additionally, both organisms biomass is also expected to increase over time. As *S. cerevisiae* is growing, ethanol is secreted. Subsequently, its concentration in the environment is expected to increase as well.

4.6.1 *Computational results*

DFBAlab, DyMMM and SMETANA's results are presented here.

- **DFBAlab 's results**

Figure 9 depicts glucose, xylose, ethanol and both organisms biomass' concentration variation over time, using DFBAlab 's simulation methods for *S. cerevisiae* and *E. coli* ZSC113 consortium.

Both sugars' concentration is decreasing over time. Nevertheless, glucose is completely depleted while xylose is still in the medium. Such fact can be explained by a faster glucose uptake rate from *S. cerevisiae*, in comparison to xylose's consumption rate by *E. coli*.

It is visible that *S. cerevisiae* reaches the stationary growth phase when glucose is no longer present in the environment. Consequently, ethanol concentration stabilizes as it is no longer being secreted.

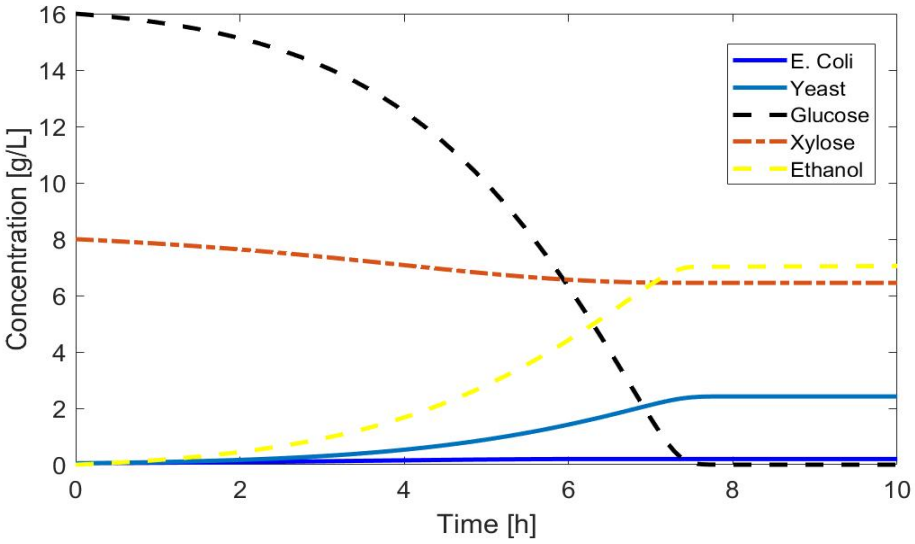


Figure 9: DFBAlab's predictions regarding metabolite, *E. coli* ZSC113 and *S. cerevisiae*'s (yeast) biomass variation over time.

Even though there is still xylose available for consumption in the environment, *E. coli*'s has reached the stationary growth phase and does not consume it. This can be explained by the inhibition associated with the presence of ethanol.

Figure 10 portrays glucose, xylose, ethanol and both organisms' biomass concentration over time, using DFBAlab 's simulation methods for *S. cerevisiae* and wild-type *E. coli* consortium.

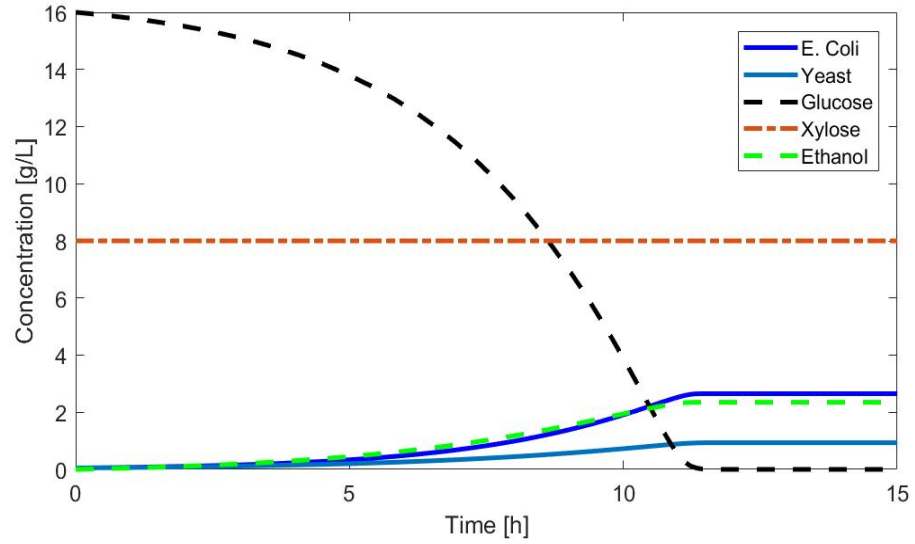


Figure 10: DFBAlab 's predictions regarding metabolite, wild-type *E. coli* and *S. cerevisiae*'s (yeast) biomass variation over time.

Xylose concentration is not varying over time, as expected. Additionally, glucose depletes at approximately 11h and both organisms reach stationary growth phase, which confirms that glucose consumption by both strains is limiting their growth. As a consequence from *S. cerevisiae*'s growth stabilization, ethanol secretion also reaches a stationary phase, as depicted the figure.

After analysing both consortia's simulation results using DFBAlab , it is possible to conclude that the tool can correctly predict accurate results, given the environmental conditions and community dynamics information, such as maximum velocity uptake rates for metabolites, inhibition and saturation constants and mass balance equations in the form of ODEs.

- **DyMMM 's results**

Figure 11 depicts DyMMM simulation results regarding biomass variation over time, for both *S. cerevisiae* and *E. coli* ZSC113. On the same note, glucose, xylose and ethanol's concentration variation over time is also represented.

Both organisms biomass concentration is increasing. Nevertheless, At approximately 2 and 6.5h, *S. cerevisiae* and *E. coli* reach the stationary growth phase, respectively. Such is caused by the depletion of the carbon sources. Specifically, glucose depletes first, followed by xylose.

Ethanol concentration is not varying over time, which is inaccurate as it should be produced as *S. cerevisiae* grows. Such result is a consequence of an incorrect mass balance equation representing ethanol variation over time, as it was not possible to correct it in DyMMM.

Figure 12 represents DyMMM simulation results regarding *S. cerevisiae* and wild-type *E. coli*'s biomass variation over time. Glucose, xylose and ethanols concentration variation over time is also depicted.

Firstly, xylose concentration is not varying over time, agreeing with theoretical expectations. Additionally, glucose depletes at approximately 5h and both organisms reach stationary growth phase at the same time, which confirms that glucose consumption by both strains is limiting its growth.

Lastly, ethanol concentration is not varying over time, which is once again, inaccurate. It was secretion was expected to match *S. cerevisiae* growth, stopping when the organism reached stationary growth phase.

After analysing both case studies' results, it is possible to affirm that DyMMM correctly predicted sugar consumption, and consequently, growth, by both organisms. Nevertheless, the inability to edit the tool's ODEs, specifically the one representing ethanol's concentration variation over time, generated inaccurate results regarding the metabolite's dynamic in both consortia.

- **SMETANA**

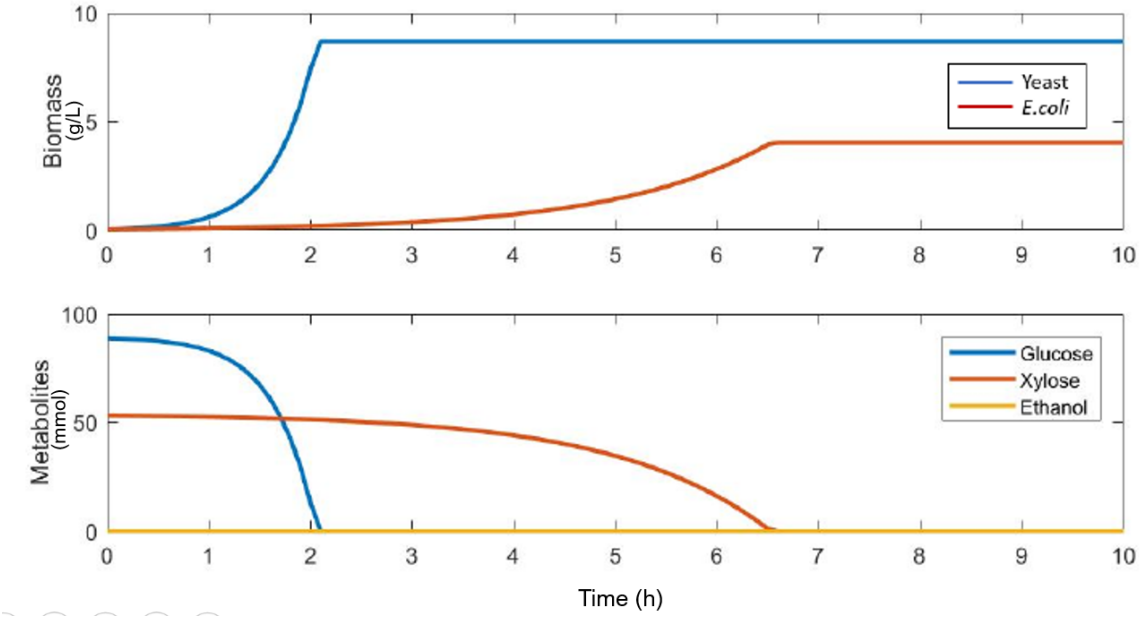


Figure 11: DyMMM's predictions regarding metabolite concentration (mmol), *E. coli* ZSC113 and *S. cerevisiae*'s (yeast) biomass (g/L) variation over time (h).

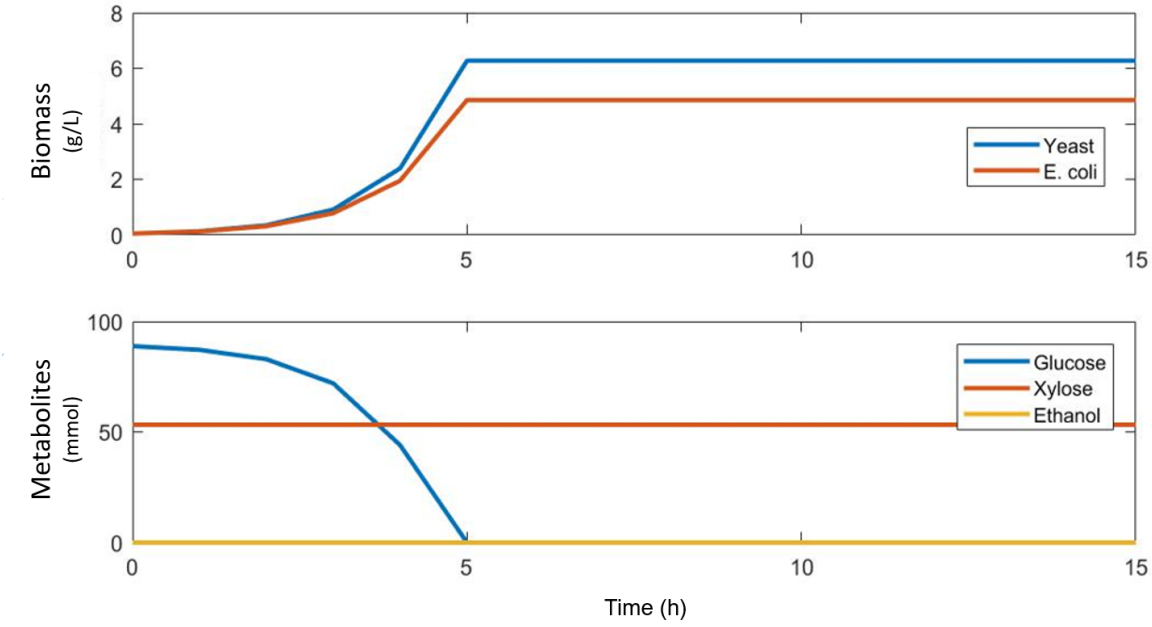


Figure 12: DyMMM's predictions regarding metabolite concentration (mmol), wild-type *E. coli* and *S. cerevisiae*'s (yeast) biomass (g/L) variation over time (h).

SMETANA's metrics were calculated using iND750 and iJR904 GSM models as input. Additionally, a minimal medium, which SMETANA creates by default, was used to specify

environmental conditions. Such medium is created based on the minimal requirements for growth of any introduced organisms' models.

SMETANA results regarding dependencies between the organisms in the community, when the environmental conditions are constrained to a minimal medium, are described in Table 26.

The results suggest that *S. cerevisiae* can provide *E. coli* with several amino acids. Specifically, the higher SMETANA scores are for L-cysteine and L-threonine's (0.83 and 0.58, respectively) cross-feeding.

Additionally, it is predicted that *S. cerevisiae* can provide citrate, fumarate and malate, which are compounds involved in tricarboxylic acid cycle.

The highest SMETANA score was calculated for phosphate secretion by *S. cerevisiae* and consequent consumption by *E. coli*. However, such interaction is improbable as the metabolite is commonly available in the culture medium, and each organism can naturally consume it.

4.6.2 *Performance assessment*

The scope of this study was to understand the consortium metabolites and biomass' dynamics. Unlike the previous case studies, the goal was not identifying the cross-feeding between community members. Thus, static FBA methods, such as pFBA and SteadyCom were not used and community models were not built.

After analysing DFBAlab and DyMMM's predictions, it is noticeable that the first tool's results are more meticulous. Such is caused by misrepresentation of ethanol's production over time, in DyMMM.

The lack of variation of ethanol's concentration over time affected the results regarding biomass formation, in both organisms.

On one hand, DFBAlab calculated lower biomass values. For instance, regarding *S. cerevisiae* and *E. coli* ZNC113's consortium, when the organisms reach the stationary growth phase, biomass concentration is approximately 2 and 0.5 g/L, respectively. On the other hand, DyMMM predicted values are approximately 9 and 5 g/L, respectively. This difference can be explained by the lack of ethanol inhibition effect over both strains growth, which leads to higher biomass formation, in DyMMM.

Additionally, regarding the *S. cerevisiae* and wild-type *E. coli* consortium, DFBAlab's results show a higher biomass concentration for *E. coli* than for *S. cerevisiae*, whereas DyMMM predicts the opposite. As the ethanol inhibition constant over *S. cerevisiae*'s growth is higher than *E. coli*'s (10 and 8 g/L, respectively, described in Table 10), it results in a lower biomass concentration for *S. cerevisiae*. Thus, the difference between the tools' results is explained by DyMMM's lack of ethanol inhibition effect over both strains.

Table 26: SMETANA's metrics for *E. coli* and *S. cerevisiae* co.culture. *Receiver* is the organism who benefits from the production of a specific *Metabolite* from another organism in the community, the *Donor*. SMETANA is a score that indicates how feasible the interaction is. It varies from 0 to 1, where 0 represents no interaction and 1 represents a mandatory interaction.

Receiver	Donor	Metabolite	SMETANA
E. coli	*S. cerevisiae*	L-alanine	0.05
E. coli	*S. cerevisiae*	L-arginine	0.04
E. coli	*S. cerevisiae*	L-Asparagine	0.04
E. coli	*S. cerevisiae*	Citrate	0.31
E. coli	*S. cerevisiae*	L-cysteine	0.83
E. coli	*S. cerevisiae*	Fumarate	0.1
E. coli	*S. cerevisiae*	L-glutamate	0.12
E. coli	*S. cerevisiae*	Glycine	0.01
E. coli	*S. cerevisiae*	Glycerol	0.17
E. coli	*S. cerevisiae*	Guanine	0.06
E. coli	*S. cerevisiae*	L-histidine	0.03
E. coli	*S. cerevisiae*	L-leucine	0.02
E. coli	*S. cerevisiae*	L-lysine	0.01
E. coli	*S. cerevisiae*	L-malate	0.32
E. coli	*S. cerevisiae*	Phosphate	1.0
E. coli	*S. cerevisiae*	L-proline	0.03
E. coli	*S. cerevisiae*	D-Sorbitol	0.12
E. coli	*S. cerevisiae*	L-serine	0.09
E. coli	*S. cerevisiae*	L-threonine	0.58
E. coli	*S. cerevisiae*	L-tryptophan	0.14
E. coli	*S. cerevisiae*	L-tyrosine	0.02
E. coli	*S. cerevisiae*	L-valine	0.02

Nevertheless, both tools correctly predicted the organisms behaviour regarding sugar consumption, thus having similar results in that aspect.

Regarding SMETANA's calculations, even though not particularly relevant to the purpose of this study, which is to analyse sugar consumption, they showed some potential cooperative interactions between the organisms, when in co-culture.

5

CONCLUSION AND FUTURE WORK

In this work, different tools' performances were compared. In addition, the respective tools' advantages and disadvantages were assessed.

Regarding the used methods to build a community model, COBRA Toolbox's proved to be more reliable as the built community structures accurately represent exchange between the organisms and the environment. Additionally, it is more universal regarding the creation of a community model. Specifically, in *D. vulgaris* and *M. maripaludis*' case study, the respective community model was created, even though the individual GSM models did not have information regarding an extracellular compartment.

Additionally, the fact that micom creates community reactions based on the GSM models' drains, rather than the extracellular metabolites, can lead to connectivity issues. Thus, a certain curation might be required, regarding the addition of certain exchange reactions. On the other hand, COBRA Toolbox creates community reactions based on the presence of the extracellular metabolites,being more reliable and universal regarding the creation of a community model.

Regarding simulation methods, SteadyCom presented more realistic results in the four *E. coli*'s case study. Specifically, it does not allow an organism to secrete metabolites to benefit the growth of another entity in the community, if the organism itself is not growing. Nevertheless, SteadyCom was not able to compute feasible solutions for all case studies, not allowing its proper comparison to other tools. Additionally, such inability points to a lack of universality, as the procedure to use the tool is not compatible with every GSM model's structure.

Furthermore, a feasible solution was found for all case studies, using pFBA. These solutions were in agreement with the available theoretical and/or experimental information. Thus, pFBA can be used to understand microbial community interactions.

Regarding dynamic simulations, they are very useful to predict how a specific metabolite concentration changes over time. Nevertheless, to perform a simulation, the reactions which will be analysed have to be pre-determined. Thus, dynamic methods are not the best option if the purpose of the study is to identify unknown interactions. Additionally, specific organism information, such as its maximum uptake velocities, is required. Such

information is not always available as it is determined experimentally, raising a limitation to this tools' usage.

SMETANA proved to be useful at identifying possible interactions. It poses an advantage regarding the required time for its usage and result analysis. Specifically, the tool's results can be more rapidly obtained than pFBA or SteadyCom's, as reaction bounds do not need to be individually constrained. Nevertheless, for it to accurately calculate its metrics, the metabolites available in the environment in which a consortium grows needs to be well specified.

Finally, in addition to how accurate a tool is, the quality of the used individual GSM models directly affects the results. For instance, if such a model is not properly curated, misleading interactions might be identified. Thus, high quality GSM models with extensive manual curation, based on literature, biological databases, organism-specific and related-organisms information, should be used to make the best use of the tools capabilities.

Further work should include μubialSim (68). It is a dFBA-based tool whose performance could be assessed and compared to DyMMM and DFBAlab's. Additionally, other simulation methods, such as FVA, could be explored on a community level.

To conclude, systems biology approaches allows the understanding of how different species interact and affect the environment in which a studied consortium is found. Nevertheless, the community modelling field still requires the development of a standard procedure to analyse community behaviour. As a starting point, GSM models' reaction and metabolite IDs should be standardized to facilitate the process of building a desired community model. Nowadays, it seems that the compartmentalized approach best recreates a community and the environment it exists in. Additionally, a standard approach to simulate a community's behaviour should also be able to mimic a biological feasible solution.

BIBLIOGRAPHY

[1] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology 215*, 3 (1990), 403–410.

[2] Arioli, S., Della Scala, G., Remagni, M. C., Stuknyte, M., Colombo, S., Guglielmetti, S., De Noni, I., Ragg, E., and Mora, D. Streptococcus thermophilus urease activity boosts lactobacillus delbrueckii subsp. bulgaricus homolactic fermentation. *International journal of food microbiology 247* (2017), 55–64.

[3] Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., et al. Kbase: the united states department of energy systems biology knowledgebase. *Nature Biotechnology 36*, 7 (2018).

[4] Benthin, S., Nielsen, J., and Villadsen, J. A simple and reliable method for the determination of cellular rna content. *Biotechnology Techniques 5*, 1 (1991), 39–42.

[5] Biggs, M. B., Medlock, G. L., Kolling, G. L., and Papin, J. A. Metabolic network modeling of microbial communities. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine 7*, 5 (2015), 317–334.

[6] Bosi, E., Bacci, G., Mengoni, A., and Fondi, M. Perspectives and challenges in microbial communities metabolic modeling. *Frontiers in genetics 8* (2017), 88.

[7] Burgard, A. P., Pharkya, P., and Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering 84*, 6 (2003), 647–657.

[8] Chan, S. H. J., Simons, M. N., and Maranas, C. D. Steadycom: Predicting microbial abundances while ensuring community stability. *PLoS computational biology 13*, 5 (2017), e1005539.

[9] Chervaux, C., Ehrlich, S. D., and Maguin, E. Physiological study of lactobacillus delbrueckii subsp. bulgaricus strains in a novel chemically defined medium. *Appl. Environ. Microbiol. 66*, 12 (2000), 5306–5311.

[10] Derzelle, S., Bolotin, A., Mistou, M.-Y., and Rul, F. Proteome analysis of streptococcus thermophilus grown in milk reveals pyruvate formate-lyase as the major upregulated protein. *Appl. Environ. Microbiol. 71*, 12 (2005), 8597–8605.

[11] Dias, O., and Rocha, I. Systems Biology in Fungi. In *Mucormycosis. Food and Water Borne Mycotoxigenic and Mycotic Fungi*. 2015, ch. 6, pp. 69–88.

[12] Dias, O., Rocha, M., Ferreira, E. C., and Rocha, I. Reconstructing genome-scale metabolic models with merlin. *Nucleic acids research 43*, 8 (2015), 3899–3910.

[13] Diener, C., and Resendis-Antonio, O. Micom: metagenome-scale modeling to infer metabolic interactions in the microbiota. *bioRxiv* (2018), 361907.

[14] Duarte, N. C., Herrgård, M. J., and Palsson, B. Ø. Reconstruction and validation of saccharomyces cerevisiae ind750, a fully compartmentalized genome-scale metabolic model. *Genome research 14*, 7 (2004), 1298–1309.

[15] Duncan, S. H., Barcenilla, A., Stewart, C. S., Pryde, S. E., and Flint, H. J. Acetate utilization and butyryl coenzyme a (coa): acetate-coa transferase in butyrate-producing bacteria from the human large intestine. *Appl. Environ. Microbiol. 68*, 10 (2002), 5186–5190.

[16] Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. Cobrapy: constraints-based reconstruction and analysis for python. *BMC systems biology 7*, 1 (2013), 74.

[17] Eddy, S. R., et al. Multiple alignment using hidden markov models. In *Ismb* (1995), vol. 3, pp. 114–120.

[18] Edwards, J. S., and Palsson, B. O. Systems Properties of the Haemophilus influenzae Rd Metabolic Genotype. *Molecular Biology 274*, 25 (1999), 17410 –17416.

[19] El-Semman, I. E., Karlsson, F. H., Shoaie, S., Nookaew, I., Soliman, T. H., and Nielsen, J. Genome-scale metabolic reconstructions of bifidobacterium adolescentis l2-32 and faecalibacterium prausnitzii a2-165 and their interaction. *BMC systems biology 8*, 1 (2014), 41.

[20] Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. Predicting subcellular localization of proteins based on their n-terminal amino acid sequence. *Journal of molecular biology 300*, 4 (2000), 1005–1016.

[21] Feist, A. M., and Palsson, B. Ø. The growing scope of applications of genome-scale metabolic reconstructions using escherichia coli. *Nature biotechnology 26*, 6 (2008), 659.

[22] Feist, A. M., Zielinski, D. C., Orth, J. D., Schellenberger, J., Herrgard, M. J., and Palsson, B. Ø. Model-driven evaluation of the production potential for growth-coupled products of escherichia coli. *Metabolic engineering 12*, 3 (2010), 173–186.

[23] Ferrer, J., Prats, C., and López, D. Individual-based modelling: an essential tool for microbiology. *Journal of biological physics 34*, 1-2 (2008), 19–37.

[24] Fleming, R. M., Thiele, I., and Nasheuer, H. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to escherichia coli. *Biophysical chemistry 145*, 2-3 (2009), 47–56.

[25] Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., Altermann, U., Angerer, P., Ansorge, S., Balasz, K., et al. Loctree3 prediction of localization. *Nucleic acids research 42*, W1 (2014), W350–W355.

[26] Gomez, J. A., Kai, H., and Barton, P. I. DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis. 1–10.

[27] Gottstein, W., Olivier, B. G., Bruggeman, F. J., and Teusink, B. Constraint-based stoichiometric modelling from single organisms to microbial communities. *Journal of the Royal Society Interface 13*, 124 (2016), 20160627.

[28] Green, M., and Karp, P. Genome annotation errors in pathway databases due to semantic ambiguity in partial ec numbers. *Nucleic acids research 33*, 13 (2005), 4035–4039.

[29] Gudmundsson, S., and Thiele, I. Computationally efficient flux variability analysis. *BMC bioinformatics 11*, 1 (2010), 489.

[30] Hanemaaijer, M., Olivier, B. G., Röling, W. F., Bruggeman, F. J., and Teusink, B. Model-based quantification of metabolic interactions from dynamic microbial-community data. *PloS one 12*, 3 (2017), e0173183.

[31] Hanemaaijer, M., Röling, W. F., Olivier, B. G., Khandelwal, R. A., Teusink, B., and Bruggeman, F. J. Systems modeling approaches for microbial community studies: from metagenomics to inference of the community structure. *Frontiers in microbiology 6* (2015), 213.

[32] Hanemaaijer, M. J. *Exploring the potential of metabolic models in the study of microbial ecosystems*. Amsterdam: Vrije Universiteit, 2016.

[33] Hanly, T. J., Urello, M., and Henson, M. A. Dynamic flux balance modeling of s. cerevisiae and e. coli co-cultures for efficient consumption of glucose/xylose mixtures. *Applied microbiology and biotechnology 93*, 6 (2012), 2529–2541.

[34] Heinonen, M., Osmala, M., Mannerström, H., Wallenius, J., Kaski, S., Rousu, J., and Lähdesmäki, H. Bayesian metabolic flux analysis reveals intracellular flux couplings. *arXiv preprint arXiv:1804.06673* (2018).

[35] Hellweger, F. L., Clegg, R. J., Clark, J. R., Plugge, C. M., and Kreft, J.-U. Advancing microbial sciences by individual-based modelling. *Nature Reviews Microbiology 14*, 7 (2016), 461.

[36] Henry, C. S., Bernstein, H. C., Weisenhorn, P., Taylor, R. C., Lee, J.-Y., Zucker, J., and Song, H.-S. Microbial community metabolic modeling: a community data-driven network reconstruction. *Journal of cellular physiology 231*, 11 (2016), 2339–2345.

[37] Henson, M. A., and Hanly, T. J. Dynamic flux balance analysis for synthetic microbial communities. *IET systems biology 8*, 5 (2014), 214–229.

[38] Herbert, D., Phipps, P., and Strange, R. Chapter iii chemical analysis of microbial cells. 209–344.

[39] Herve-Jimenez, L., Guillouard, I., Guedon, E., Boudebbouze, S., Hols, P., Monnet, V., Maguin, E., and Rul, F. Postgenomic analysis of streptococcus thermophilus cocultivated in milk with lactobacillus delbrueckii subsp. bulgaricus: involvement of nitrogen, purine, and iron metabolism. *Appl. Environ. Microbiol. 75*, 7 (2009), 2062–2073.

[40] Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., et al. The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics 19*, 4 (2003), 524–531.

[41] Izard, J., and Limberger, R. J. Rapid screening method for quantitation of bacterial cell lipids from whole cells. *Journal of microbiological methods 55*, 2 (2003), 411–418.

[42] Jeske, L., Placzek, S., Schomburg, I., Chang, A., and Schomburg, D. Brenda in 2019: a european elixir core data resource. *Nucleic acids research 47*, D1 (2018), D542–D549.

[43] Julsing, M. K., Kuhn, D., Schmid, A., and Bühler, B. Resting cells of recombinant e. coli show high epoxidation yields on energy source and high sensitivity to product inhibition. *Biotechnology and bioengineering 109*, 5 (2012), 1109–1119.

[44] Kanehisa, M., and Goto, S. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research 28*, 1 (2000), 27–30.

[45] Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., et al. The ecocyc database: reflecting new knowledge about escherichia coli k-12. *Nucleic acids research 45*, D1 (2016), D543–D550.

[46] Khandelwal, R. A., Olivier, B. G., Röling, W. F., Teusink, B., and Bruggeman, F. J. Community flux balance analysis for microbial consortia at balanced growth. *PloS one 8*, 5 (2013), e64567.

[47] Kim, T. Y., Sohn, S. B., Kim, Y. B., Kim, W. J., and Lee, S. Y. Recent advances in reconstruction and applications of genome-scale metabolic models. *Current opinion in biotechnology 23*, 4 (2012), 617–623.

[48] King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., and Lewis, N. E. Bigg models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic acids research 44*, D1 (2015), D515–D522.

[49] Klitgord, N., and Segre, D. Environments that Induce Synthetic Microbial Ecosystems.

[50] Koch, S., Benndorf, D., Fronk, K., Reichl, U., and Klamt, S. Predicting compositions of microbial communities from stoichiometric models with applications for the biogas process. *Biotechnology for biofuels 9*, 1 (2016), 17.

[51] Koch, S., Kohrs, F., Lahmann, P., Bissinger, T., Wendschuh, S., Benndorf, D., Reichl, U., and Klamt, S. Redcom: A strategy for reduced metabolic modeling of complex microbial communities and its application for analyzing experimental datasets from anaerobic digestion. *PLoS computational biology 15*, 2 (2019), e1006759.

[52] Koponen, J., Laakso, K., Koskenniemi, K., Kankainen, M., Savijoki, K., Nyman, T. A., de Vos, W. M., Tynkkynen, S., Kalkkinen, N., and Varmanen, P. Effect of acid stress on protein expression and phosphorylation in lactobacillus rhamnosus gg. *Journal of proteomics 75*, 4 (2012), 1357–1374.

[53] Laakso, K., Koskenniemi, K., Koponen, J., Kankainen, M., Surakka, A., Salusjärvi, T., Auvinen, P., Savijoki, K., Nyman, T. A., Kalkkinen, N., et al. Growth phase-associated changes in the proteome and transcriptome of lactobacillus rhamnosus gg in industrial-type whey medium. *Microbial biotechnology 4*, 6 (2011), 746–766.

[54] Letort, C., Nardi, M., Garault, P., Monnet, V., and Juillard, V. Casein utilization by streptococcus thermophilus results in a diauxic growth in milk. *Appl. Environ. Microbiol. 68*, 6 (2002), 3162–3165.

[55] Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., et al. Omic data from evolved e. coli are consistent with computed optimal growth from genome-scale models. *Molecular systems biology 6*, 1 (2010).

[56] Machado, D., Andrejev, S., Tramontano, M., and Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic acids research 46*, 15 (2018), 7542–7553.

[57] Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature biotechnology 35*, 1 (2017), 81.

[58] Mahadevan, R., Edwards, J. S., and Doyle III, F. J. Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophysical journal 83*, 3 (2002), 1331–1340.

[59] Mahadevan, R., and Henson, M. A. Genome-based modeling and design of metabolic interactions in microbial communities. *Computational and structural biotechnology journal 3*, 4 (2012), e201210008.

[60] Mahadevan, R., and Henson, M. A. Genome-based modeling and design of metabolic interactions in microbial communities. *Computational and structural biotechnology journal 3*, 4 (2012), e201210008.

[61] McCloskey, D., Palsson, B. Ø., and Feist, A. M. Basic and applied uses of genome-scale metabolic network reconstructions of escherichia coli. *Molecular systems biology 9*, 1 (2013).

[62] Mitri, S., and Richard Foster, K. The genotypic view of social interactions in microbial communities. *Annual review of genetics 47* (2013), 247–273.

[63] Mo, M. L., Jamshidi, N., and Palsson, B. Ø. A genome-scale, constraint-based approach to systems biology of human metabolism. *Molecular Biosystems 3*, 9 (2007), 598–603.

[64] Oberhardt, M. A., Palsson, B. Ø., and Papin, J. A. Applications of genome-scale metabolic reconstructions. *Molecular systems biology 5*, 1 (2009).

[65] Orth, J. D., Thiele, I., and Palsson, B. Ø. What is flux balance analysis? *Nature biotechnology 28*, 3 (2010), 245.

[66] Pastink, M. I., Teusink, B., Hols, P., Visser, S., de Vos, W. M., and Hugenholtz, J. Genome-scale model of streptococcus thermophilus lmg18311 for metabolic comparison of lactic acid bacteria. *Appl. Environ. Microbiol. 75*, 11 (2009), 3627–3633.

[67] Patil, K. R., Rocha, I., Förster, J., and Nielsen, J. Evolutionary programming as a platform for in silico metabolic engineering. *BMC bioinformatics 6*, 1 (2005), 308.

[68] Popp, D., and Centler, F. μbialsim: constraint-based dynamic simulation of complex microbiomes. *BioRxiv* (2019), 716126.

[69] Rajendran, N., Subramaniam, S., Lotha, R., Pemaiah, B., and Sivasubramanian, A. Isolation and characterization of flavonoids and flavone glycosides from the ethnic traditional medicinal plant cotoneaster bacillaris wall. ex lindl. *Der. Pharm. Lett. 8*, 5 (2016), 321–324.

[70] Raman, K., and Chandra, N. Flux balance analysis of biological systems: applications and challenges. *Briefings in bioinformatics 10*, 4 (2009), 435–449.

[71] Rau, M. H., and Zeidan, A. A. Constraint-based modeling in microbial food biotechnology. *Biochemical Society Transactions 46*, 2 (2018), 249–260.

[72] Reed, J. L., Vo, T. D., Schilling, C. H., and Palsson, B. O. An expanded genome-scale model of escherichia coli k-12 (i jr904 gsm/gpr). *Genome biology 4*, 9 (2003), R54.

[73] Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J. P., Nielsen, J., Patil, K. R., Ferreira, E. C., and Rocha, M. Optflux: an open-source software platform for in silico metabolic engineering. *BMC systems biology 4*, 1 (2010), 45.

[74] Rocha, M., Maia, P., Mendes, R., Pinto, J. P., Ferreira, E. C., Nielsen, J., Patil, K. R., and Rocha, I. Natural computation meta-heuristics for the in silico optimization of microbial strains. *BMC bioinformatics 9*, 1 (2008), 499.

[75] Rocha, M., Maia, P., Mendes, R., Pinto, J. P., Ferreira, E. C., Nielsen, J., Patil, K. R., and Rocha, I. Natural computation meta-heuristics for the in silico optimization of microbial strains. *BMC bioinformatics 9*, 1 (2008), 499.

[76] Rodriguez, J., Kleerebezem, R., Lema, J. M., and van Loosdrecht, M. C. Modeling product formation in anaerobic mixed culture fermentations. *Biotechnology and bioengineering 93*, 3 (2006), 592–606.

[77] Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Federhen, S., et al. Database resources of the national center for biotechnology information. *Nucleic acids research 40*, D1 (2011), D13–D25.

[78] Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., et al. Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2. 0. *Nature protocols 6*, 9 (2011), 1290.

[79] Schotten, S. Coaggregation of Gut Bacteroides & Parabacteroides with Probiotic Lactobacillus Rhamnosus GG.

[80] Segre, D., Vitkup, D., and Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences 99*, 23 (2002), 15112–15117.

[81] Settachaimongkon, S., Nout, M. R., Fernandes, E. C. A., Hettinga, K. A., Vervoort, J. M., van Hooijdonk, T. C., Zwietering, M. H., Smid, E. J., and van Valenberg, H. J. Influence of different proteolytic strains of streptococcus thermophilus in co-culture with lactobacillus delbrueckii subsp. bulgaricus on the metabolite profile of set-yoghurt. *International journal of food microbiology 177* (2014), 29–36.

[82] Shlomi, T., Berkman, O., and Ruppin, E. Proc Natl Acad Sci USA 2005 Shlomi. 1–6.

[83] Sieuwerts, S. Microbial interactions in the yoghurt consortium: current status and product implications. *SOJ Microb. Infect. Dis 4* (2016), 1–5.

[84] Sieuwerts, S., De Bok, F. A., Hugenholtz, J., and Van Hylckama Vlieg, J. E. Unraveling microbial interactions in food fermentations: From classical to genomics approaches. *Applied and Environmental Microbiology 74*, 16 (2008), 4997–5007.

[85] Stalidzans, E., Seiman, A., Peebo, K., Komasilovs, V., and Pentjuss, A. Model-based metabolism design: constraints for kinetic and stoichiometric models. *Biochemical Society Transactions 46*, 2 (2018), 261–267.

[86] Standaert, A. R., Poschet, F., Geeraerd, A. H., Uylbak, F. V., Kreft, J.-U., and Van Impe, J. F. A novel class of predictive microbial grown models: Implementation in an individual-based framework. *IFAC Proceedings Volumes 37*, 3 (2004), 183–188.

[87] Stephanopoulos, G., Aristidou, A. A., and Nielsen, J. *Metabolic engineering: principles and methodologies*. Elsevier, 1998.

[88] Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., Lie, T. J., Leigh, J. A., and Stahl, D. A. Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology 3*, 92 (2007), 1–14.

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

[90] Thiele, Ines; Palsson, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc 5*, 1 (2010), 93–121.

[91] Wintermute, E. H., and Silver, P. A. Emergent cooperation in microbial metabolism. *Molecular systems biology 6*, 1 (2010).

[92] Wintermute, E. H., and Silver, P. A. Emergent cooperation in microbial metabolism. *Molecular systems biology 6*, 1 (2010).

[93] Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J., et al. Psortb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics 26*, 13 (2010), 1608–1615.

[94] Zeidan, A. A., Rådström, P., and van Niel, E. W. Stable coexistence of two caldicellulosiruptor species in a de novo constructed hydrogen-producing co-culture. *Microbial cell factories 9*, 1 (2010), 102.

[95] Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., and Patil, K. R. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences 112*, 20 (2015), 6449–6454.

[96] Zhao, C., Dong, H., Zhang, Y., and Li, Y. Discovery of potential genes contributing to the biosynthesis of short-chain fatty acids and lactate in gut microbiota from systematic investigation in e. coli. *NPJ biofilms and microbiomes 5*, 1 (2019), 1–8.

[97] Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R., and Lovley, D. R. Genome-scale dynamic modeling of the competition between rhodoferax and geobacter in anoxic subsurface environments. *The ISME journal 5*, 2 (2011), 305.

[98] Zomorrodi, A. R., Islam, M. M., and Maranas, C. D. d-optcom: dynamic multi-level and multi-objective metabolic modeling of microbial communities. *ACS synthetic biology 3*, 4 (2014), 247–257.

[99] Zomorrodi, A. R., and Maranas, C. D. Optcom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS computational biology 8*, 2 (2012), e1002363.