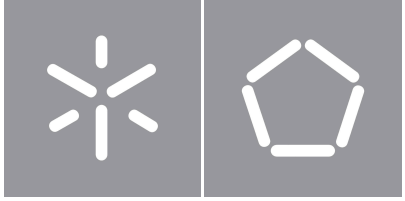
**University of Minho**
School of Engineering

Alexandre Areias Castro

# Development of microbial community simulation methods to characterize and analyze the effects of metal concentrations on pathogen silencing/promotion

january 2024



**University of Minho**
School of Engineering

Alexandre Areias Castro

# Development of microbial community simulation methods to characterize and analyze the effects of metal concentrations on pathogen silencing/promotion

Masters Dissertation
Masters in Bioinformatics

Dissertation supervised by
**Vitor Manuel Sá Pereira**
**Sarela Garcia Santamarina**

january 2024

# Copyright and Terms of Use for Third Party Work



## License granted to users of this work:

# Acknowledgements

To my supervisors, Sarela and Vítor, for providing all the conditions and support to make this into a finished research project. Through all the challenges, difficulties and time constraints, your belief in the conclusion of this project, even when it seemed impossible, only made it more enjoyable. I feel extremely honored to have completed this task with the support and counseling of such great researchers and professionals.

To my family, for their sacrifices and support allowed me to pursue this education and follow my dreams. Thanks to you and your support, I was able to not only finish but to take every step that has led me here. May life allow me to give back all you have given me.

To my friends, for their company and reassurance in every one of the times we were together helped me gather the strength to move on. Every laugh, joke or every time you have listened to me rant about how difficult life was was truly a therapy, and makes me realize that friends are truly one of life's biggest gifts.

Lastly, to my girlfriend, Matilde, for your love and unconditionall support. No one believed in me more than you, and most times, that belief overcame my own. Words cannot express how proud I am to have had on my side every step of the way, and hope to continue to have so. Life is truly special by your side.

# 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.

University of Minho, Braga, january 2024

Alexandre Areias Castro

# Abstract

Genome-scale metabolic modeling has become a widespread methodology for analyzing microbial metabolism. The mathematical modeling of these networks has proven itself as an indispensable tool for understanding microbes by predicting their behavior and systematically allowing the testing of different hypotheses. On microbial communities, it further allows the prediction of interactions at the level of metabolites and metabolic reactions, therefore providing valuable insights on the inner working of the vast networks they constitute. Current extensions of this methodology have considered the addition of enzymatic data to single models in order to further constraint reaction fluxes and produce far more accurate results.

The goal of the research presented in this dissertation is to combine different methodologies currently in existence in order to create a methodology that allows the effective and accurate simulation of the behaviour of gut microbial species inside a communal space. These simulations will ultimately aim to predict growth under low levels of transition metals in situations that resemble those of infection, where the host limits the availability of iron and other elements essential to all life forms in order to fight pathogenesis.

Through the combination of the enhancement of models with enzymatic data with the creation of compartmentalized models under different simulation methods, the devised work aimed to extend both existent methodologies and produce results that were improved throughout the entirety of the work by basing and aiming to replicate provided results obtained from *in vitro* conditions. The obtained results reveal not only that the enhancement of models with enzyme constraints provide community stability but also the extension of the capabilities of community simulation methods previously established.

The results of this work constitute a novel approach to the simulation of gut microbial communities that revealed the impact the enhancement process has on both the way models behave in singular settings and also in providing community stability. Moreover, the devised approach and subsequent results provide the oportunity of extension and improvement with ever more data in the application to similar or other study cases.

**Keywords** Gut microbial communities, Metabolism, Community modeling, Genome-scale metabolic models, Transition metals

# Resumo

A simulação metabólica de modelos em escala genômica tornou-se uma metodologia proeminente na análise do metabolismo microbiano. A modelação matemática destas redes tem-se provado continuamente como uma ferramenta indispensável na compreensão do comportamento de microorganismos ao prever o seu comportamento, permitindo assim a sistematização da testagem de diferentes hipóteses. Em comunidades microbianas, esta permite ainda a previsão das interações ao nível dos metabolitos e reações, conferindo assim uma visão mais alargada do funcionamento interno destas vastas redes por eles constituídas. As extensões para esta metodologia atualmente utilizadas como a adição de restrições enzimáticas têm como objetivo o aumento das restrições aplicadas a cada modelo, de modo a que estes produzam resultados cada vez mais precisos.

O objetivo deste projeto de investigação apresentado nesta dissertação passa pela combinação de diferentes metodologias atualmente existentes de modo a proceder à criação de uma nova metodologia que permita a simulação do comportamento de espécies microbianas do trato intestinal inseridas num espaço comunitário. Estas simulações terão como objetivo final a capacidade de prever o seu crescimento sobre baixos níveis de metais de transição em situações que se assemelham ás de infeção, onde o hospedeiro limita a disponibilidade de ferro e outros elementos essenciais à vida de modo a combater a patogénese.

Através da combinação do melhoramento de modelos pela adição de restrições enzimáticas com a criação de modelos compartimentalizados sobre diferentes métodos de simulação, o present trabalho visou a extensão de ambas as metodologias e produzir resultados que foram sistematicamente aprimorados por se basearam e emularem resultados obtidos em condições *in vitro*. Os resultados obtidos revelam que não só o melhoramento dos modelos com restrições enzimáticas promovem a estabilidade comunitária, mas também que extendem as capacidades de métodos previamente estabelecidos.

Os resultados deste trabalho consitutem uma nova abordagem para a simulação de comunidades microbianas intestinais que revelam o impacto que o processo de melhoramento teve na maneira como os modelos se comportam, assim como na promoção da estabilidade das comunidades. Adicionalmente,

a abordagem desenvolvidade e subsequentemente, os seus resultados, são passíveis de extensão e melhoramento para ainda mais dados de modo a serem aplicados a estudos similares ou extendidos a outros casos de estudo.

**Palavras-chave** Comunidades microbianas intestinais, Metabolismo, Modelação de comunidades, Modelos metabólicos em escala genómica, Metais de transição

# Contents

# List of Figures

# List of Tables

# Acronyms

**CP** Calprotectin.

**FBA** Flux Balance Analysis.

**GEM** Genome-Scale Metabolic Models.

**GPR** Gene-Protein-Reaction.

**IB** Individual-based.

**LV** Lotka-Volterra.

**MA** MacArthur.

**SBML** Systems Biology Markup Language.

**TB** Trait-based.

# Glossary

**GECKO** Method for enhancement of GEMs with Enzymatic Constraints using Kinetic and Omics data.

# Part I

# Introductory material

# Chapter 1

# Introduction

## 1.1 Motivation

Microbial communities are the most natural form of occurrence of microbes. The interactions that occur among them as well as those that occur between them and their host can have various repercussions that can largely contribute to the overall health of humans and many other forms of life. The power to understand their effects and more importantly the underlying constitution, function and interactions that happen amidst these communities, as well as how they are affected by factors such as their diversity and other outer factors can unlock a new way of understanding microbial ecology.

The study of microbial communities and their functioning inevitably comprehends the understanding of the emergent properties that surge within the communities. When the sheer number of interactions is put into perspective, one can understand how such microscopic interactions can produce macroscopic properties. These properties can be defined as any pattern or function that cannot be deduced as the sum of the properties of the constituent parts [van den Berg et al., 2022] and are typically a byproduct of the size of the community, as well as the connections established between the members of these complex microbial networks. Finding the threshold that triggers these properties can prove to be an ambiguous and mostly situational task due to the specificity and uniqueness that characterize every single community as well as their intrinsic properties. Due to that, most characterizations of the thresholds and the emergent properties from which they are inherent are mostly done in quantitative forms that often lack precision and reliability.

Transition metals are essential nutrients to all forms of life. Their importance and effects are directly influenced by their presence and concentration. However, it is crucial to understand that both the deficiency and excess of these metals are the basis for the development of numerous pathologies and represent a major public health burden [Van Gossum and Neve, 1998]. The interaction with these elements alters the physio-chemical properties of proteins, thereby promoting catalysis of enzymatic reactions, stabilizing protein structure, and/or facilitating electron transport [Murdoch and Skaar, 2022], and can affect many crucial cellular processes such as respiration, transcription, signal transduction, and proliferation [Andreini et al., 2006]. Organisms maintain appropriate levels of metals by controlling uptake, use, storage, and excretion at the cellular and systemic levels. Tissue metal levels in vertebrate hosts are mainly controlled by absorption from dietary sources in the intestinal tract, whereas bacterial pathogens obtain metal from the extracellular and intracellular environments of host tissues during infection [Lopez and Skaar, 2018].

Unlocking new mechanisms that enable a better understanding of how emergent properties are triggered and their functioning and the effect of different metal concentrations on the intestinal tract microbial communities can result in new forms of comprehending the appearance of a great number of pathologies while providing new insights on new approaches for that matter.

## 1.2 Goals

Through the creation of new mechanisms of metabolic modelling, the research here presented aims to capture the effects of fluctuating levels of transition metals in gut microbial communities and the several mechanisms these trigger, including both pathogen colonisation resistance and commensalist interactions. Examples of these are seen in the utilization of siderophores produced by pathogens under Fe limitation carried out by commensal species such as *Bacteroides thetaiotaomicron* as reported in Zhu et al. [2020]. By combining the use of **Genome-Scale Metabolic Models (GEM)**, the enzymatic enhancement of these models and the use of various simulation approaches for community models like **Flux Balance Analysis** and SteadyCom [Chan et al., 2017], the results of the this work will constitute a novel approach to not only microbial community modelling in general but also to a more accurate comprehension of the mechanisms that are intrinsic to the interactions that occur within the gut microbial space and at the host–pathogen interface.

## 1.3 Structure

The present document is structured in different chapters, each focusing on a specific subject related to the task at hand.

- **Chapter 2:** Metabolic Modelling - This chapter serves as introductory material to metabolic modeling by exploring the most important aspects of this approach. The chapter continues by presenting introductory concepts to community models as well as some of the most widely used types of models. Afterward, the topic of enzymatic constraints enhancement of models in which this approach is introduced, as well as some of its core advantages.

- **Chapter 3:** Microbial Communities - Here, a brief introduction to microbial communities is made while addressing some of its core characteristics and concepts like their structure and function. Afterward, further notions are presented on the several interactions that occur both on the internal and external interfaces of these biological systems.

- **Chapter 4:** Transition Metals - This chapter addresses the general introductory concepts of transition metals, focusing on their structure and function, while also approaching the various control mechanisms exerted by host and microbial organisms, as well as the interactions promoted by the presence of these elements.

- **Chapter 5:** Methods - In this chapter the established workflow is described, along with the performed tasks, the work's timeline, main methods, and tools used.

- **Chapter 6:** Results and Discussion - This chapter presents the obtained results, their discussion and limitations.

- **Chapter 7:** Conclusions - Here the main conclusions, contributions and propositions for future work, as well as the main dificulties encountered are presented and described.

# Chapter 2

# Metabolic Modelling

## 2.1 Metabolic Modelling of Communities

Understanding the relationship between genotype and phenotype is a crucial part of understanding how biological systems work. To that end, the comprehensive knowledge of the complex metabolic pathways, genetic information, and how to mathematically model them has surged as a major requirement for biologists. The use of these mathematical models has become an established tool for systematic analyses of metabolism for a wide variety of organisms, having proved themselves as being crucial not only to the understanding of mechanisms underlying complex human diseases but also to other applications such as model-driven development of efficient cell factories [Domenzain et al., 2022].

The interactive biological groups that exist in nature hold the secret to understanding how individuals or groups of individuals affect population dynamics, resource availability, and even diversity. Communities' biotic composition and functions are extremely variable and being able to associate these factors with the role of singular forms of life creates the possibility to not only understand their functioning but also create ways to implement dynamics that allow communities to attain certain structures and functions. For biologists, it has become a major goal to create models and simulation methods that allow the obtaining of results that can encompass interactions and co-factors close to those that occur in natural environments.

There is currently a wide range of types of models that for the creation of simulation methods under many approaches and desired capabilities. Globally, the formulation methods of models can be divided into two main groups: steady-stake and kinetic models.

The creation of community models has presented itself as an effective mechanism for the resolution of that problem. These models can be defined as agglomerates of singular models that take into account reactions that occur within each individual whilst also accounting for interactions that take place amidst and between other constituent parts of these communities. This mathematical approach, allied with the ever-growing computational power, has enabled the creation of simulation methods that can emulate growth rates and other important factors such as enzymatic usage, according to desired parameters, while also allowing the creation of desired mutants that can better display desired traits for each desired effect.

In the next sub-chapter, a list of the most common types of community models is presented, alongside their functioning, mathematical counterparts, as well as their viability in terms of their application to the creation of microbial community models.

### 2.1.1 Types of Models

#### Lotka–Volterra models - Kinetic

The **Lotka-Volterra (LV)** model was introduced in the 20th century and is used to describe predator-prey systems and their dynamics. It is one of the first mathematical approaches to describe complex ecological dynamics [Nedorezov, 2016], consisting of a set of non-linear, coupled, first-order differential equations. The classic model does not capture interactions involving more than two species, so its relevance to the study of microbial communities lies in the Generalized **Lotka-Volterra** model, which can accommodate any number of species. The generic **LV** system for $n$ populations or species $x_i$ takes the form:

$$\frac{dB_i}{dt} = B_i(\alpha_i + \sum_{j=1}^{n} \beta_{ij} x_j) \tag{2.1}$$

where the non-negative parameter $\alpha_i$ is the growth rate of species $i$ and each real-valued interaction parameter $\beta_{ij}$ quantifies the type and strength of the effect of species $j$ on species $i$, if $i \neq j$. If $i = j$, $\beta_{ij}$ reflects intraspecies interactions.

In the case of the presence of only a single species ($n = 1$), the **LV** model simplifies to the logistic equation:

$$\frac{dB_i}{dt} = \alpha_i + \beta x^2 \tag{2.2}$$

The two-variable case ($n = 2$) of the Lotka-Volterra model includes three terms for each species: one growth term ($\alpha_i$), one intraspecies term ($\beta_{ii}$) and one interspecies interaction term ($\beta_{ij}$):

$$\frac{dx_1}{dt} = x_1(\alpha_1 + \beta_{11}x_1 + \beta_{12}x_2)$$

$$\frac{dx_2}{dt} = x_2(\alpha_1 + \beta_{21}x_1 + \beta_{22}x_2)$$

The sign of each interspecies interaction term represents the type of relationship between the two species, such as described in Table 1:

Table 1: Types of interspecies interactions and interaction coefficients signals.

| Interaction Type | $\beta_{ij}$ | $\beta_{ji}$ |
|---|---|---|
| Neutralism | 0 | 0 |
| Mutualism | + | + |
| Commensalism | + | 0 |
| Predation | + | - |
| Amensalism | 0 | - |
| Competition | - | - |

When contemplating the application of these models to microbial communities, however, some difficulties associated with the assumptions created when using the model arise. These assumptions presuppose that firstly, the interactions between populations occur while being static in time and space, meaning that events such as environmental changes, physiological adaptation, regulatory shifts or evolutionary changes are not taken into account [Voit et al.]. Secondly, positive feedback loops in the system can lead to unrestrained population growth. Consequently, fitting data to a **Lotka-Volterra** model may underestimate the prevalence of mutualism (Hoek et al. [2017], van den Berg et al. [2022]). Also, a third problem surges as the model postulates that there are no higher-order interactions meaning that the alteration of interactions between a pair as a consequence of the presence of a third species does not occur. Lastly, small interactions such as metabolite exchange or quorum sensing are not included in these models, which highly compromises the study of the emergent properties of a microbial community, and therefore also produces results and conclusions farther from practical scenarios.

### MacArthur consumer–resource models - Kinetic

This phenomenological model presents itself as an alternative approach to that proposed by the **Lotka-Volterra** group of models. The resource-explicit model encompasses the growth rate's dependence on chemical factors such as nutrients, signals or toxins, while also coupling their concentrations to population dynamics. The model allows the quantification of the way different species use resources: the bigger the resource overlap, the smaller the chance of coexistence. The growth rate of each species is directly proportional to its resource consumption, accounting for a non-parameterization of species interactions but rather their mediation being done by the deficiency in shared resources among the entire community or population. In **MacArthur**'s models, each consumer species is regulated by its resource harvest, which is linear in the resource densities, relative to its maintenance requirement:

$$\frac{dB_i}{dt} = r_i B_i(\sum_{\alpha} \Delta w_{i\alpha} C_{i\alpha} R_\alpha) - m_i B_i \tag{2.3}$$

$$\Delta w_{i\alpha} = w_\alpha - \sum_{\beta} D^i_{\beta\alpha} w_\beta \tag{2.4}$$

where $B_i$ is the density of consumer species $i$, $1 \leq i \leq n$ ; per unit time, a unit of consumer species $i$ harvests (utilizes) $C_{i\alpha}R_\alpha$ units of resource; per unit time, a unit of consumer species $i$ must harvest energy $r_i$ to maintain itself. Each resource species is subject to a growth function (Equation 2.3), which may involve not only self-regulation but also interactions with other resource species, and mortality by consumption ($m_i$). It is assumed that the resources don't interact with one another, but they can be utilized and consumed by intricate dynamics. The **MacArthur** consumer-resource equations also allow for resources and consumers to have different timescales. If the resource dynamics are fast, the system simplifies to dynamics similar to those presented by **Lotka-Volterra** models.

Its applicability to microbial communities, however, depends on its extension to far beyond more than two species and two resources. It also involves a series of adaptations to the model, like the ability to include cross-feeding, the assumption of fixed metabolism of community members, the absence of taxonomic and hierarchical metabolic structures, and the missing thermodynamic and biochemical detail, being the latter a common limitation amongst other ecological models. This is a classic ecological model that is deterministic in nature, in contrast with the stochastic biotic and abiotic influences on community dynamics [Zelezniak et al., 2015].

### Trait-based models - Kinetic

**Trait-based (TB)** models, unlike other types of models previously mentioned, define the focal variables as the phenotypic traits of individuals rather than phylogenetic groups. Members are defined by their traits and the models describe how trait combinations respond to environmental variables, as well as their influence over them. Also, unlike classic ecological models that generally assume a constant environment, trait-based models focus on community dynamics along major environmental gradients, which leads to a functional trait distribution that optimizes certain community-level functional properties. This allows the **TB** model's capability to better simulate the emergence of selection pressures and diversity patterns along modeled environmental gradients, which are important aspects in unlocking the understanding of evolutionary processes [Litchman et al., 2013]. The growth per population (Equation 2.5) and impact environmental factor (Equation 2.6) are calculated through the following equations

$$\frac{dB_i}{dt} = (r_i - m_i)B_i \tag{2.5}$$

$$r_i = \frac{r_{i\alpha}^{max} R_\alpha}{R_\alpha + K_{i\alpha}} e S_i (T - T_{ref}) \tag{2.6}$$

where $B_i$ is the population size of $i$, $r_i$ is the impact environmental factor of $i$ and $m_i$ represents the death of individuals belonging from population $i$. **Trait-based** models are, therefore extremely well suited to be applied to larger communities, being however dependent on their low local variations in environmental variables. This has proven itself as a problem to its application to gut microbiota systems, not only because these communities present high local variability, but also because even though **TB** models do not require the prior imposition of trait combinations, they do require the definition of which traits are included in the system.

### Individual-based models - Kinetic

With their first implementations having served the purpose of studying macrofaunal communities like the **MacArthur** and **Lotka-Volterra** models, more recent works have increasingly applied **Individual-based (IB)** models to microbial communities. These agent-based models take a bottom-up approach in the way that they model every member of the community as an individual interacting agent with a set of complex "rules", therefore presenting a contrast when compared to models that encompass top-down approaches that simulate adaptations at the population level and hence not capturing smaller evolutionary dynamics that are mediated by mutations that occur at an individual level. The system of equations for the individual-based models encompasses a number of rules per agent (Equation 2.7):

$$\frac{di}{dt} = r_i(\sum_{\alpha} w_{i\alpha} C_{i\alpha} R_{\alpha}) \tag{2.7}$$

Where $i$ is an agent, $r_i C_{i\alpha}$ is the consumption rate of a metabolite $R_{\alpha}$. If $i$ is bigger than the threshold, then $i$ can expand into the free neighboring space by some probability $P$. **IB** models present therefore a lot of capabilities and advantages concerning their use being applied to modeling microbial communities. Their ability to simulate the complete evolutionary process via inheritable mutations [Clark et al., 2011] allows the capturing of the emergence of community resilience and phenotypic complexity under perturbations that occur in the environment. Another major capability of **IB** models lies in the possibility to include a series of biophysical processes and thermodynamic constraints on community interactions and diversity [Gogulancea et al., 2019] that can not only increase model accuracy but also reduce computation time as a result of the reduction of the solution space [Gutiérrez et al., 2017].

Limitations to the use of **IB** models lie, however, in their complexity, as they are not only computationally expensive but also of low robustness. One solution to this problem can be the reduction of dimensional descriptions [Parise et al., 2015], as well as the lowering of the number of agents in each community to lower the community variety [Coyte et al., 2015]. Overall, **Individual-based** models are well suited to identify both spatially explicit and system-wide emergent properties in microbial communities, even though reaching of optimal robustness and stability can present itself as a great challenge.

### Genome-scale metabolic models - Steady-state

**Genome-Scale Metabolic Models (GEM)** hold the ability to mathematically represent almost the entirety of the biochemical reactions that constitute the metabolism of an individual by reconstructing them through the genetic information encoded in every single living being (Figure 1). These models can be generated either through semi-automatic or automatic genome-scale reconstruction tools, while also creating the possibility of enhancement with other forms of physiological, kinetic, and omics data, such as enzymatic constraints.

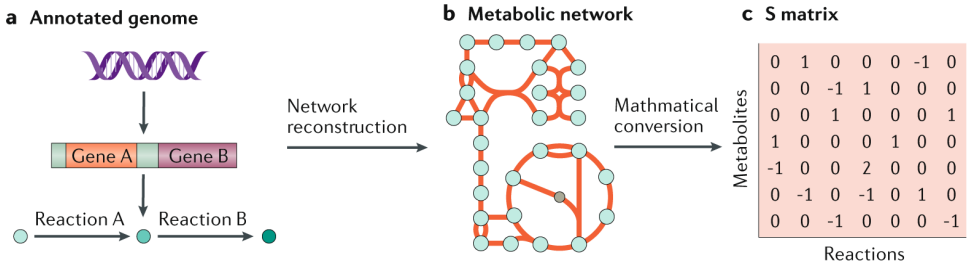



Figure 1: Construction process of a **GEM** from the genome of an individual.

Similarly to **Trait-based** models, these aim at capturing the near-complete metabolic and biosynthetic capabilities of all community members as encoded in their genomes. Their use enables the simulation of the metabolic phenotype of an organism, including its growth rate, intracellular reaction fluxes, and secretion/uptake rates, providing an insight into its response to both environmental and genetic perturbations. The most common simulation method, **Flux Balance Analysis (FBA)**, calculates the flow of metabolites through the metabolic network, thereby creating the possibility to predict the growth rate of an organism or the rate of production of a desired metabolite, according to its objective function. Other common simulation approaches, like SteadyCom [Chan et al., 2017] or OptCom [Zomorrodi and Maranas, 2012] have also proven themselves extremely advantageous in offering results that better enable the simulation of emergent properties of communities. Current limitations to **Genome-Scale Metabolic Models** include the absence of negative interactions beyond resource competition (toxins, antimicrobials) as well as the necessity of extensive amounts of curations to the models in order to obtain more precise results. Regarding the use of **GEM** models for creating microbial community models, even with the challenges previously stated, continue to constitute a great approach to their simulation mainly due to their capability of automated reconstruction and integration of (meta-) omics data [van den Berg et al., 2022].

### Compartmentalized models - Steady-state

Compartmentalized models can be defined as models in which each organism is a distinct compartment and exchangeable metabolites are transferred through an external compartment representing the extracellular environment (Figure 2A). The approach can be implemented by assigning reactions and metabolites to a network representing each organism, while explicit transport reactions account for the exchange of metabolites between organisms and the extracellular space. This approach is an ideal method to enable the understanding of how organisms perform a particular metabolic transformation [Taffs et al., 2009].

### Bag-of-reactions models - Steady-state

This approach encompasses the combination of the entirety of the metabolic reactions and metabolites from all community members in one shared compartment (Figure 2B). Furthermore, in cases where more than one organism catalyzes the reaction, only one reaction is assigned [Glöckler et al., 2022]. Its low necessity in terms of computational power can constitute an advantage that can overthrow its inability to specify which species use or produce a particular enzyme or metabolite [Taffs et al., 2009]. Overall, this type of model can constitute a useful tool to study the growth of a community but not to analyze the interactions that happen between its members.

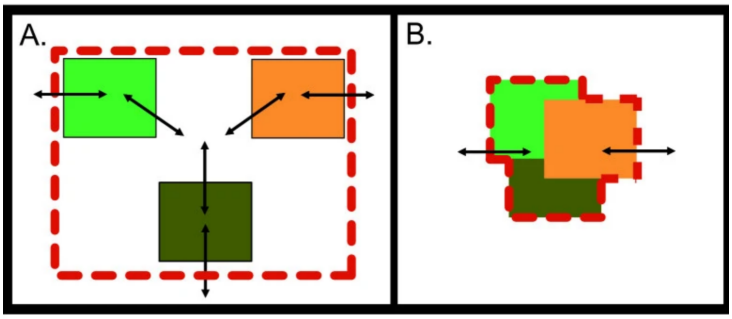



Figure 2: Conceptual depiction of the Compartmentalized (A) and Pooled (B) modelling approaches.

## 2.2 Enzymatic Constraints enhancement of models

**Genome-Scale Metabolic Models (GEM)** have increasingly been continuously proven as an established tool for systematic simulation and metabolic analysis. Their ability to capture the metabolic phenotype of an organism unlocks the power to understand the many reactions that are involved in the metabolic network of each organism. One limitation to **GEM**s however, lies in its inability to attain different flux distribution profiles, a property caused by the presence of network redundancies that provide organisms with robustness to environmental and genetic perturbations [Domenzain et al., 2022]. Oftentimes, experimental measurements of exchange fluxes are incorporated into the models as a way to remediate this problem. However, measurements are not readily available for a wide variety of conditions and organisms, as they are hard to capture and often not precise.

Presented as a solution to the problem stated above, the integration of enzymatic restrictions on metabolic reactions has proven itself successful at expanding the range of predictions of classical **FBA**, enabling the enlightenment on phenomena like overflow metabolism and cellular growth on diverse environments under several conditions. The development and implementation of **GECKO**, a method for enhancement of **GEM**s with Enzymatic Constraints using Kinetic and Omics data, has allowed not only the extension of the classical FBA approach (Equation 2.8) by incorporating a description of enzymatic requirements for all the corresponding metabolic reactions in a model (Equation 2.9), taking into account all types of enzyme-reaction relations, but also direct integration of proteomics abundance data as constraints for individual protein demands.

$$\text{(Metabolic-only model) Stoichiometry of reaction: } A \longrightarrow B \tag{2.8}$$

$$\text{(Enzyme-constrained model) Stoichiometry of reaction: } A + 1/\,k_{cat}\,E \longrightarrow B \tag{2.9}$$

The kinetic information, in the form of $k_{cat}$ values, is included as stoichiometric coefficients to convert the metabolic flux in $mmol/gDW^{-1}$ to the required enzyme usage in $mmol/gDW^{-1}$. The implementation of the method is achieved by extending the traditional stoichiometric matrix representing the whole metabolism in genome-scale modelling, where each reaction's stoichiometry is defined at the matrix's columns and the mass balance for each metabolite is indicated in each row. **GECKO** expands this matrix 3 by adding new rows to the matrix, therefore representing the enzymes while the added columns represent each enzyme's usage.

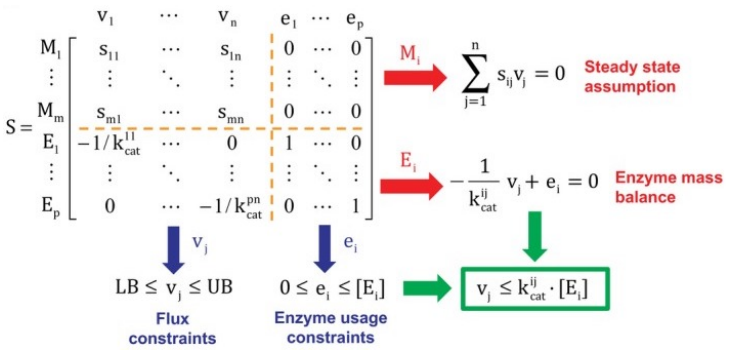



Figure 3: Expansion of the traditional genome-scale stoichiometric matrix.

In the above visual representation of the stoichiometric matrix, $M$ refers to metabolites, $E$ to enzymes, $v$ to metabolic fluxes, and $e$ to enzyme usage. The created division causes the appearance of 4 new submatrices inside the new stoichiometric matrix. The upper left submatrix is equivalent to the original stoichiometric matrix, the upper right submatrix has only zeros, the lower left submatrix has the kinetic information, and the lower right submatrix is the identity matrix.

Furthermore, the use of this method allows the simulation of models without the necessity to define an extremely accurate growth medium, while also dampening the growth of organisms to its enzymatic constraints. Both of these advantages are extremely useful when considering their use in community metabolic modelling, as these models encompass the sharing of the growth medium amongst various microorganisms. Current obstacles to this approach however, arise when considering the biggest advantage of using it, as the most notable improvements in flux predictions are achieved when experimental exchange rates are limited or unavailable [Robinson et al., 2020].

## 2.3 Chapter Summary

The above chapter served as an overview of metabolic modelling, with a special emphasis on community models being made. The most common types of models, along with their characteristics, advantages, and limitations were presented, as well as a dedication to address the approach of enzymatic constraints enhancement of models, having this sub-chapter focused on the overview of the method and its advantages and limitations.

# Chapter 3

# Microbial Communities

## 3.1 Introduction

Microbial communities are the naturally occurring form of microbes. These complex biological systems are the vessel for a vast number of host-microbe and microbe-microbe interactions whose understanding unlocks the possibility to comprehend how singular parts of these networks behave as a part of the whole, while also providing greater insights on how they interact at the host-microbe interface, with this latter possibility being crucial in the process of creating a bridge between the occurring interactions and the various mechanisms and triggers alike.

## 3.2 Structure and Function

Akin to other forms of biological communities, the structure of a microbial community is defined by its biotic composition and diversity. These are generally composed of a group of microorganisms including bacteria, archaea and microeukaryotes, while the specific community composition can be determined by a set of abiotic factors such as nutrient availability, temperature, or even through the presence of other microorganisms whose presence exerts pressure in other taxonomic groups.

The function of a community attains to its energy flow, resilience, and resistance. These properties are affected by a large range of factors and the full comprehension of the function can result in a higher understanding of its mechanisms. Composition and function are, as it can be concluded, not independent factors of the equation. Rather, both are affected and affect each other in ways that further postulate the sense of network they presuppose.

## 3.3 Internal Interactions

Microbe-microbe interactions account for an important part of the interactions that happen within communities. The mechanisms that are unlocked through the interacting between several populations can shape community dynamics and therefore give rise to emergent properties that can confer resilience, resistance to colonization and even promote pathogenesis. Pratical examples shown in studies have proven that microbiota can promote resistance to colonization by pathogenic species, including studies in which mice that were treated with antibiotics or bread in sterile environments have proven to be more prone to the infection by enteric pathogenic bacteria such as *Shigella flexneri*, *Citrobacter rodentium*, *Listeria monocytogenes* and *Salmonella enterica serovar Typhimurium* (Sprinz et al. [1961], Kamada et al. [2012], Zachar and Savage [1979]). Other studies have also shown that some microbiotas can also lead to the expansion or enhance the virulence of other pathogenic populations [Cameron and Sperandio, 2015]. Creating the link between the composition of the microbial communities, their interactions and function has therefore been pinpointed as a major focal point in the comprehension of microbial communities. Other cases of healthy individuals being transplanted with microbiota from infected individuals has also resulted in the virulence of those previously healthy individuals [Ghosh et al., 2011], therefore supporting the idea that the internal interactions play a major role not only in composition but also in the surging of emergent properties.

### 3.3.1 Auxotrophs

Microbial communities are composed of cells with a wide range of metabolic capacities, and regularly include auxotrophs. This group can be defined as microorganisms that present auxotrophy, a state in which an organism is unable to syntethise compounds that are essencial to their growth. The search for the role of auxotrophs in microbial communities has lead to the discovery of not only the prevalence of auxotrophs in microbial communities, but also that their metabolic cooperation unlocks a series of mechanisms that contribute to the emergence of properties that convey resistence to drugs not only to them, but also to the entirety of the community, by promoting a richer metabolic environment with high metabolite efflux [Yu et al., 2022].

## 3.4 Microbiota-pathogen-host interface

The whole range of interactions that occur at the host-pathogen interface mediate a arms race that is characterized by a range of mechanisms. In one front, the need to obtain compounds that are essential to the growth of microbial cells promotes a series of interactions that seek not only to obtain metabolites from the extracellular space but also promotes the theft of any of those elements as the hosts find ways to sequester many compounds in order to stop growth that might enable pathogenesis. Another form of interactions encompass the constant need for communities to evolve ways to develop resistence to drugs and other forms of antibiotics that might be introduced by the host's immune system. This promotes a series of populational dynamics that shapes adaptability of the communities either through selective pressure that results in mutants that promote the surging of emergent properties or through the change of metabolite uptake and secretion, in which a range of microbe-microbe interactions mediate a change in the community's function. Having that said, other forms of mutualistic relations between hosts and microbes give shape to virulence resistance, where non-pathogenic microbiota promote resistance to colonization by pathogenic species, as previously mentioned. Furthermore, the understanding of these dynamics creates a connection between many of the aspects previously mentioned.

The presence of pathogens in microbial communities constitutes the basis for many pathologies that affect hosts in many ways. While their presence in communities can be benefitial or not to them, the pressure exerted by the hosts, and in some cases, by other microbiota, leads to a number of interactions, therefore representing another complex form of interaction that contribute to the ever growth of the network in which microbial communities and their interactions can be represented. In further chapters, a review of some of those mechanisms, specially those involving the use of trace metals will be presented.

## 3.5 Chapter Summary

In this chapter, a brief introduction to microbial communities was made, addressing some of its core concepts and characteristics, as well as the presentation of some of the internal and external interactions that occur between microbiota and hosts alike.

# Chapter 4

# Transition Metals

## 4.1 Introduction

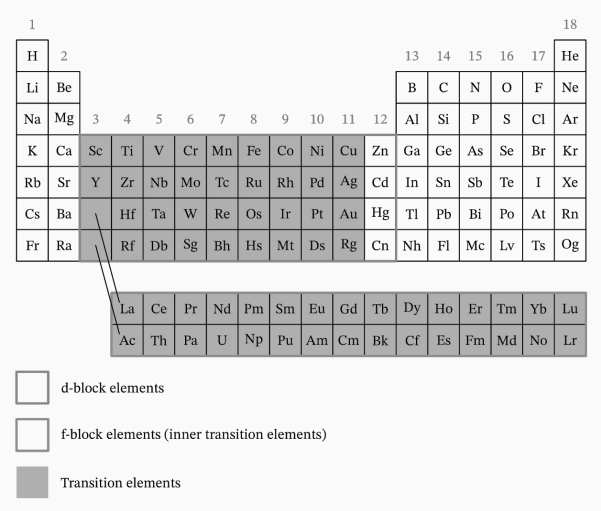



Figure 4: Transition metals in the Periodic Table of elements.

Transition metals are comprised of the chemical elements located in the central block (d-block) of the Periodic Table (Figure 4). All the elements are solids with the exception of mercury, Hg, which is a liquid at normal pressures and temperatures. Naturally occurring metals in biological systems include elements like iron (Fe), zinc (Zn), cobalt (Co) and copper (Cu). All these elements, in contrary to some of those present in the f-block of the periodic table (inner transition metals) are synthesized naturally and are crucial to several processes in biological systems. The absorption of these elements occurs mainly from dietary sources when considering animals, while microbial forms of life use many other forms of uptaking trace metals. Further discussion on the effects, homeostasis control and interactions that occur between hosts and pathogens will be made.

## 4.2 Effects

The presence of trace metals in organic systems translates into the ability to interact and cause a panoply of causes and effects. The most common lies in the interaction with a certain and widely common group of proteins: metalloproteins. This type of protein is characterized by its metal ion co-factor and metal-binding protein domains and functions in a diverse set of cellular processes, including respiration, transcription, signal transduction and proliferation [Andreini et al., 2006]. The interactions that happen between metalloproteins and transition metals produce a series of effects on processes like the promotion of catalysis of enzymatic reactions, stabilization of protein structure and/or even the mediation of electron transport. As it has been proven, the panoply of effects that the presence of trace metal produces are mostly essential to life forms of many kinds.

Excessive levels of these elements, however, have also been proven to be toxic to most life forms. One potential mechanism of toxicity resides in the products of Fenton chemistry by the redox-active transition metals iron, copper, and manganese.

$$Fe^{3+} + O_2{}^{\bullet -} \longleftrightarrow Fe^{2+} + O_2 \tag{4.1}$$

$$Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + HO + OH^- \tag{4.2}$$

$$O_2{}^{\bullet -} + H_2O_2 \longleftrightarrow O_2 + OH^{\bullet} + OH^- \tag{4.3}$$

In the Haber-Weiss reaction of Fenton chemistry, $Fe^{2+}$ reacts with hydrogen peroxide to form $Fe^{3+}$ (Equation 4.2), a hydroxyl radical, and hydroxide; $Fe^{3+}$ can then react with hydrogen peroxide to regenerate $Fe^{2+}$ in addition to a hydroperoxyl radical and a proton (Equation 4.1). These oxidative species inhibit cell growth by damaging proteins, DNA, and lipids.

Another potential mechanism of toxicity is mediated by a process known as protein mismetallation, mediated by the prevalence of metalloproteins in the bacterial cell and their relatively compatible binding ligands that require tight control of metal homeostasis for continued metabolism. In the uncontrolled levels of metals resides the origin of a shortage of protein binding sites and a subsequent excess of supply under a shortage of demand.

## 4.3 Control Mechanisms

In spite of all the necessity of trace metals exerted by cells to promote activity, the excess of these elements provokes toxicity. This phenomenon can be explained either by the excessive interaction with metalloproteins and consequent mismetallation, which nullifies their enzymatic capabilities or by the formation of redox-active molecules. To that effect and prevent toxicity, it is crucial for organisms to have mechanisms that allow them to maintain appropriate levels of these components.

### 4.3.1 Homeostasis Control

The maintenance of proper trace metals homeostasis is crucial to the survival of all organisms. To that effect, proteins dedicated to the metal homeostasis control import and export (metallotransporters), target the delivery of metals to specific enzymes (metallochaperones), are responsible for the metallocenter assembly, and regulate the expression of the other proteins in response to trace metal ions availability (metalloregulators). When the metal-trafficking process malfunctions, susceptibility to infection and many pathologies arrive. A more in-depth review is done of the main metal efflux system in Chapter 4.4.2, as well as their structure and functioning.

## 4.4 Host-Pathogen Interactions

Host metal concentrations and presence can drastically affect susceptibility to bacterial infections [Palmer and Skaar, 2016]. Other than the homeostasis control mechanisms described above, hosts have evolved to develop many other ways of controlling metal concentrations and harnessing their capabilities to fight the presence of bacterial pathogens. Some of those mechanisms include ways of creating metal withdraws from pathogens and causing a stoppage in metabolic pathways that enable cell growth, as well as others that can unleash toxic levels of metals to provoke pathogen cell death.

In the context of infection, bacterial and host metal metabolism determine disease pathogenesis under three emerging principles. The first one states that although the host represents a rich nutrient metal source for bacteria, the host immune system can block bacterial metal acquisition in a process denominated as termed nutritional immunity. Secondly, hosts have the ability to overload bacteria with toxic concentrations of metals. Finally, the deregulation of host metal homeostasis through genetic mutation or nutrition has an impact on the host's susceptibility to infection.

In order to obtain their required nutrient metals, bacteria have evolved three main classes of metal acquisition systems: elemental metal import, extracellular metal capture mediated by siderophores (low molecular weight, high-affinity $Fe^{3+}$ binding compounds secreted and imported by bacteria for iron acquisition and even theft from host), as well as metal acquisition from host proteins, being the latter a process in which metal theft from host nutritional immunity proteins under other mechanisms is also included. In response, mammalian hosts have evolved strategies to restrict bacteria from being successful in their metal acquisition process.

### 4.4.1 Metal-Sequestering Proteins

Metal bioavailability exerts a strong selective pressure at the infectious interface giving rise to an evolutionary arms race between host and pathogen that shapes metal sequestration strategies. As a response to microbial infection, many hosts evolved the deployment of metal-sequestering host-defence proteins to better control the availability of essential metal nutrients in the extracellular space. In human hosts, the case of **Calprotectin (CP)** serves as an example of these mechanisms. **CP** is an abundant antimicrobial heteroligomer of two S100 proteins (S100A8 and S100A9) (Fig. 5) released from neutrophils and epithelial cells at sites of infection that sequesters divalent first-row transition metal ions like $Fe^{2+}$ and $Zn^{2+}$ [Zygiel and Nolan, 2018], that, as it name highlights, has properties that include the binding with elements like those previoulsy mentioned and its subsequent antimicrobial activity. This is just one of the many examples of strategies enforced by hosts to inhibit the availability of essential trace metals to pathogens, in another attempt to fight their nefarious effects to its organism.

Figure 5: Crystal structure of human calprotectin (S100A8/S100A9).

### 4.4.2 Host-Imposed Metal Toxicity

#### Bacterial Metal Efflux Systems

Hosts have evolved mechanisms to deploy toxic levels of metal ions for bacterial control, and bacteria use a large span of metal exporters to resist host intoxication [Guilhen et al., 2013]. These bacterial exporters can be divided in five main groups:

- **Resistance-Nodulation-Cell division (RND)** type transporters;
- **P-type ATPase** family;
- **Cation Diffusion Facilitator (CDF)** family;
- **Transporter Mediating Manganese Export (MntX)** family;
- **Cobalt Resistance protein A (CorA)** family.

The **Resistance-Nodulation-Cell division (RND)** type transporters are integral membrane proteins mediating the efflux of a broad variety of substrates with a subset exporting metals. It is composed by a pump (annotated A in Fig. 6), an integral membrane protein with a hydrophilic periplasmic component, which is connected to a trimeric outer membrane factor (C in Fig. 6), that serves the purpose of allowing the efflux of metal into the extracellular space [Paulsen et al., 1997]. The third and final part of the RND transporter complex is composed of several units of a periplasmic membrane protein (B in Fig. 6)I, serving as a binder that creates a ring involving the outer membrane proteins and the pump, promoting the stabilization of the contact between the two other components (Murakami et al. [2002], Akama et al. [2004]). Several RND exporter systems have been identified to mediate the efflux of $Co^{2+}$, $Zn^{2+}$, $Cd^{2+}$, $Ni^{2+}$, $Cu^{+}$ and $Ag^{+}$ (Nies and Silver [1989], Liesegang et al. [1993], Stähler et al. [2006], Long et al. [2012]).

Members of the **P-type ATPase** family (Fig. 6) are present in eukaryotes and prokaryotes and couple metal transport to the hydrolysis of ATP, in a process where catalytic phosphorylation of the transporter occurs after binding of cytoplasmic metal to the trans-membrane metal-binding sites [Fagan and Saier, 1994]. Metals affected by the action of these transporters can be divided in two main groups: $Zn^{2+}/Cd^{2+}/Pb^{2+}$ or $Cu^{+}/Ag^{+}$ (Botella et al. [2011], Fu et al. [2013]).

**Cation Diffusion Facilitator (CDF)** family members are the mediators for the transport of a variety of metals in both prokaryotes and eukaryotes [Haney et al., 2005]. These transporters are usually constituted by six transmembrane domains followed by a metallochaperone-like cytoplasmic domain (Fig. 6) that regulates metal transport activity [Lu et al., 2009]. These transporters mediate the efflux of $Zn^{2+}$ $Cd^{2+}$ and $Fe^{2+}$, having however the transport of $Fe^{2+}$ having been proven less effective than the others [Hoch et al., 2012].

A subclass of CDF exporters, MntE, a $Mn^{2+}$ transporter has also been described. Several recent studies, however, have shown that MntE is not the only type of $Mn^{2+}$ exporter, having lead to the discovery of the existence of the **Transporter Mediating Manganese Export (MntX)** family [Li et al., 2011]. Although much is yet unknown about the structure of this family of exporters, studies regarding the secondary structure and topological predictions have suggested an inverted repeat of three transmembrane segments [Veyrier et al., 2011](Fig. 6). These transporters have all been described to export principally $Mn^{2+}$ with some lesser affinity for other divalent metals.

The fifth group of exporters (**Cobalt Resistance protein A (CorA)** family) were first identified as $Mg^{2+}$ transporters, in which some members are dedicated to the export of other divalent cations, principally $Zn^{2+}$, with the exception of some recorded cases of mediated transport of $Zn^{2+}$ and $Cd^{2+}$ [Worlock and Smith, 2002]. This protein harbours two transmembrane domains and a long cytoplasmic region (Fig. 6) that mediates the acquisition and subsequent delivery of cations to the transport channel.

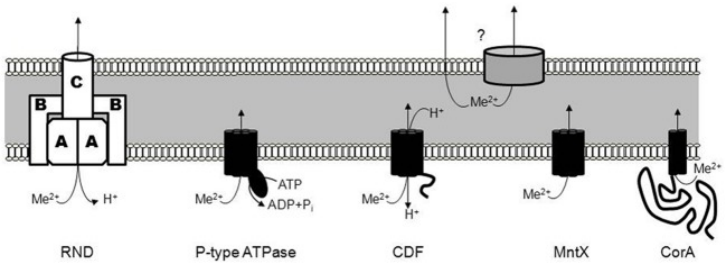



Figure 6: Different families of bacterial metal exporters.

### Pathogen Homeostatic Control

As previously stated, hosts restrict bacterial growth by depriving pathogens of essential levels of trace metals, and by definition, export activity has a rather unhelpful function for that matter. The maintenance of viable concentrations of metals under fluctuating levels imposed by hosts composes an interaction that further translates the fight happening in the front line of the host-pathogen interface.

According to the above description, most of the metal exporters from bacterial pathogens are dedicated to mediating the export of three transition metals (Zn, Cu, and Mn) that are present in substantial amounts in the human body. In observed cases of the influence of trace metals in virulence, however, it has been shown an increase of $Zn^{2+}$ levels during infection [Gaballa and Helmann, 2003], postulating the use of $Zn^{2+}$ export systems to face the rise to toxic levels of this metal.

Other cases have also demonstrated that animals use copper as an anti-microbial weapon by inducing oxidative stress [Haney et al., 2005]. It has also been established that the lack of $Cu^{+}$ exporters on pathogens is directly related to their decrease in virulence and ability to survive on the host organism.

This altogether reveals that although hosts harness the power of these metals to fight infection [Botella et al., 2012], pathogens explore the capabilities of their export systems to remediate the effects of this rise, enabling the conclusion that pathogens that lack metal exporters present impaired virulence.

## 4.5 Chapter Summary

In this chapter an overview of the characteristics, effects, homeostatic control mechanisms as well as the interactions that are shaped by the necessities exerted by organisms to obtain trace metals. A special emphasis on first-row trace elements was made, justified by their most common presence in biological systems, as well as the number of interactions mediated by their presence.

# Part II

# Core of the Dissertation

# Chapter 5

# Methods

This chapter presents and explores in detail every step taken along with the methods used to accomplish the present research. To that effect, a defined framework and worflow were created in order to not only ease the process of obtaining results for this reserach but also to make it extendable and usable for further research in a variety of contexts and applications.

Firstly, the subject of the model obtention and enhancement of models with enzymatic constraints is presented, with an overview of every step of the process as well as the used tools and frameworks. Next, an overview of the community simulations performed is also presented, with a focus on the growth media, simulation methods and samples used along this process. Finally, the entire workflow for the devised endeavour is displayed and explained to better illustrate the cyclical nature of improvement established.

The main goal of the established framework was to automate all the processes that span from the obtention and curation of the models to their enhancement with enzymatic data, and finally to the simulations made using community models. To achieve this, several tools and *Python 3* packages were used and extended to better fit this specific use case, including both previously written code and original code that can be used and further extended to other related research.

## 5.1 Experimental component

The entire methodology and comparison vessel of the work here presented was based on both *in vitro* results provided by the secondary counselor for this dissertation and published works regarding mechanisms of utilization of siderophores from *Enterobactericiae* species that promote the astraddle growth of *B. thetaiotaomicron* in Zhu et al. [2020].

The first set of results were comprised of both lists of both commensal and pathogen human gut bacterial isolates from which to obtain the metabolic models as well as optical density measurements. These provided insights on the various growth profiles for various gut microbial species as seen in Figure 7 and through which the various samples were built around, while the latter provided both various abundaces that mediated some results in Section 6.1.3 as well as providing a target mechanism to capture through the application of the formulated framework.

Overall, these results were of upmost importance to the concretization of this work and enabled, as will be further demonstrated, the constant improvement of the end results for this research.

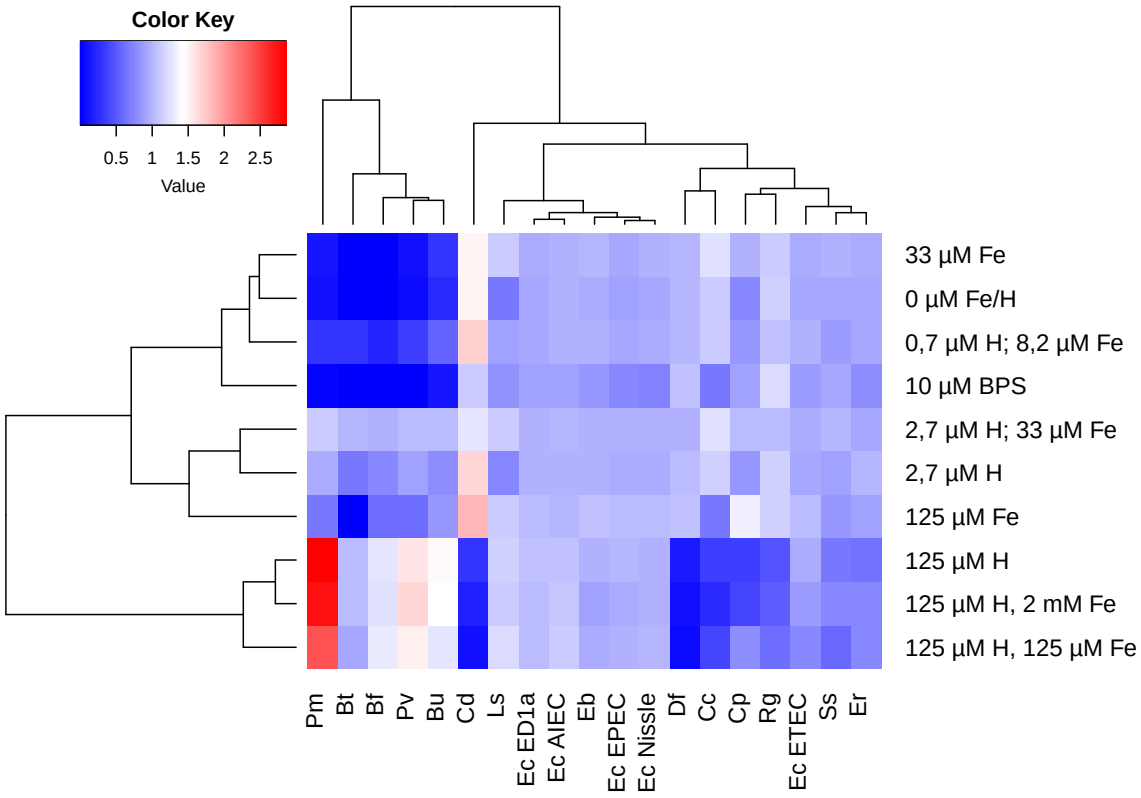



Figure 7: *in vitro* measurements of metabolic profiles

## 5.2 Model obtention

To enable the simulation of microbes and their communities, mathematical representations of the genome of individuals are created under the form of **Systems Biology Markup Language (SBML)** files. As previously mentioned, there are a large number of ways through which these models can be reconstructed and various reconstructions of several strains and species currently available.

For the present work, the AGORA repository was chosen as it contains a large number of semi-automatically generated genome-scale metabolic reconstructions for human gut microbes [Magnúsdóttir et al., 2017] with strains and species that match the target communities to study. From the entirety of the repository, 7 models were picked: *Bacteroides thetaiotaomicron* VPI 5482, *Bacteroides uniformis* ATCC 8492, *Escherichia coli* ED1a, *Fusobacterium nucleatum subsp. nucleatum* ATCC 25586, *Roseburia intestinalis* L1 82, *Streptococcus parasanguinis* ATCC 15912 and *Streptococcus salivarius* DSM 20560 (termed here, for convenience as Bt, Bu, Ec, Fn, Ri, Sp and Ss respectively) in order to match the strains present in the provided *in vitro* results with the objective to replicate them and further try to capture mechanisms and interactions that were observed in those same studies.

The desired target models were then queried and scraped, having then gone through a curation process to allow for better simulations in the present context. This process involved not only fixing formatting errors in the files, but also the fixing of **Gene-Protein-Reaction (GPR)** rules for each model, a byproduct of the semi-automatic nature of the model generation procedure through which they were created. When curated, the models were then subject to the addition of enzymatic constraints.

## 5.3 Enhancement of models with enzymatic constraints

The enzymatic enhancement of **GEM** further extends their capabilities by imposing constraints that better emulate the natural speed of reactions. To test its effects on each models' uptake profile and how it affects the overall community composition and dynamics of each sample, the models obtained from AGORA underwent a process through which they were enhanced with enzymatic constraints. This was achieved by scraping the annotations of the reactions present in the model, followed by a series of queries to biological databases such as ModelSEED, BIGG and many others, as well as the use of a Deep Learning model that predicts the Kcat of an enzyme from the reaction's substrates' SMILES and the protein sequence for each enzyme [Li et al., 2022]. The final process consisted mainly in the altering of each reaction's bounds, as well as their proteins' molecular weight through the mapping of reactions to their corresponding gene.

### 5.3.1 Preparation

The first step of this implementation involved, as previously mentioned, the scraping of the annotations of each reaction present in the model, in order to obtain as much ecNumbers and entry id's for the desired databases as possible as well as the mapping of each reaction to its corresponding gene by making use of the **Gene-Protein-Reaction** associations present in each of the reactions and genes. This was achieved by creating a simulation instance for the model and then iterating over every reaction of the model to obtain the available information. This step provided, when available, the ecNumbers for the reaction's enzyme, as well as entry id's for the ModelSEED and BIGG, which were crucial to the obtention of enzymatic data.

### 5.3.2 $k_{cat}$ extration

After obtaining the available information in the model through the previous step, several queries to a plethora of databases were made, including, in a primary stage, to ModelSEED. This database is a web-based resource for high-throughput generation, optimization and analysis of **Genome-Scale Metabolic Models** [Henry et al., 2010] and was used to obtain entry id's to the BIGG database, a collection of **GEM** reconstructions that links genome annotations and external databases [King et al., 2016]. This consequent query allowed the obtention of the ecNumbers correspondent to each entry and were mapped to each reaction to later be used to obtain each corresponding $k_{cat}$ and molecular weight values for each enzyme.

Following the scraping of id's and ecNumbers performed in the previous step, it was necessary to obtain information on each reactions substrates as it would reveal itself crucial in a further step, as it will be demonstrated. To that effect, it was first obtained each reactions substrate id's present in the model. Next, an automatic curation of these id's was performed in order to ensure that the maximum number of id's would match in the query to the PubChem API to be performed next. This query's end goal was to obtain the canonical SMILES for each substrate for further use.

With all the necessary information obatined from the model, it was then possible to proceed to the obtention of the necessary enzymatic data to then enhance the target model. To that effect, 2 different approaches were taken in order to both maximize the ammount of information gathered and remedy the lack of necessary information in some reactions. The first approach consisted of a query to the BRENDA database where, by making use of the ecNumbers obtained previously, the protein sequence, as well as the turnover numbers and molecular weights are obtained. In organisms where a higher number of available information is observed, the organism name is also provided to obtain more precise results. The other

encompassed the use of the Deep-Learning approach from DLKcat, a toolbox for prediction of $k_{cat}$ values and generation of ec**GEM** [Li et al., 2022]. The used model recieved as input the ecNumbers for each reaction, as well as its protein sequence and substrate names or SMILES. The model, after parsing the available information, predicts the $k_{cat}$ value for each of the reactions.

After obtaining the $k_{cat}$ and molecular weights for each enzyme, an average of all the obtained values per enzyme is estimated to better make use of all the available information.

Table 2: Coverage of enzymatic data per model

| Model | Coverage |
|---|---|
| Bt | 50.7% |
| Bu | 50.2% |
| Ec | 71.2% |
| Fn | 74.8% |
| Ri | 68.6% |
| Sp | 66.2% |
| Ss | 75.2% |

### 5.3.3 Model generation

To generate the models enhanced with enzymatic constraints, the enzymatic data obtained in the previous steps, and by making use of the `add_enzyme_constraints` function in the MEWpy package, a mapping of the genes and reactions in the model from the data followed by the separation of reactions by reversibility, adding the $k_{cat}$ to each reaction's stoichiometry, adding gene-protein species and drawing protein pseudo-reactions, as well as creating a protein pool for each model. Finally, each model is saved in the **SBML** format to then be calibrated and used.

The differences in the models before and after the generation process can be observed in Table 3. Here, major differences in various aspects of the models are noticeable, like those in the model's reactions (Table 3a), where the number of reeactions nearly doubled, which was mainly due to the separation of reactions in the model by reversibility. Another big change is visible regarding the difference in the number of metabolites (Table 3c), a byproduct of the addition of genes as pseudo-reactions.

Overall, the results of this process, even if noticeable through this observation, will only truly be impactful in the simulations and comparisons made against non-enhanced models displayed further in Chapter 6.

Table 3: Differences between models before and after enhancement with enzymatic constraints

(a) Reactions

| Model | Not enhanced | Enhanced |
|---|---|---|
| Bt | 2438 | 4740 |
| Bu | 2686 | 5523 |
| Ec | 3586 | 6916 |
| Fn | 1203 | 2310 |
| Ri | 1983 | 4974 |
| Sp | 1258 | 2764 |
| Ss | 939 | 1842 |

(b) Genes

| Model | Not enhanced | Enhanced |
|---|---|---|
| Bt | 823 | 823 |
| Bu | 821 | 821 |
| Ec | 1427 | 1427 |
| Fn | 503 | 503 |
| Ri | 853 | 853 |
| Sp | 899 | 899 |
| Ss | 525 | 525 |

(c) Metabolites

| Model | Not enhanced | Enhanced |
|---|---|---|
| Bt | 2041 | 2865 |
| Bu | 2204 | 3026 |
| Ec | 3119 | 4547 |
| Fn | 1135 | 1639 |
| Ri | 1691 | 2545 |
| Sp | 1231 | 2131 |
| Ss | 852 | 1378 |

(d) Exchange reactions

| Model | Not enhanced | Enhanced |
|---|---|---|
| Bt | 536 | 1072 |
| Bu | 459 | 918 |
| Ec | 556 | 1112 |
| Fn | 133 | 266 |
| Ri | 465 | 930 |
| Sp | 125 | 250 |
| Ss | 103 | 206 |

(e) Demand reactions

| Model | Not enhanced | Enhanced |
|---|---|---|
| Bt | 15 | 424 |
| Bu | 13 | 423 |
| Ec | 17 | 1031 |
| Fn | 13 | 383 |
| Ri | 16 | 594 |
| Sp | 10 | 598 |
| Ss | 9 | 400 |

### 5.3.4 Model calibration

The calibration of enzyme-constrained models is crucial to the obtention of valid models and relevant results. Therefore, when considering this important step, many approaches can be employed. In other similar tasks, strategies like the use of *in vitro* results or the use of proteomic measurements were used but, considering both the ammount of models being created and calibrated at the same time and the available omics data and *in vitro* studies, the best strategy involved the calibration of models by fitting them both according to the growth results obtained experimentally and making them viable for growth with one another.

To that effect, the first part of this task involved the calibration of each model's protein pool exchange $P$ value. This was performed by manually fitting the the predicted maximal growth rates for each model according both to the low iron conditions previously described and whose results were based off of experimentally obatined results and values fitted in other experiments and published enzyme-constrained models [Adadi et al., 2012, Carrasco Muriel et al., 2023].

Table 4: Maximum predicted growth under different $P$ values

| Protein pool $P$ | Bt | Bu | Ec | Fn | Ri | Sp | Ss |
|---|---|---|---|---|---|---|---|
| 0.1 | 0.000037 | 0.000082 | 0.000767 | 0.000195 | 0.000341 | 0.000144 | 0.000126 |
| 0.2 | 0.000074 | 0.000164 | 0.001535 | 0.000390 | 0.000683 | 0.000288 | 0.000252 |
| 0.3 | 0.000111 | 0.000245 | 0.002302 | 0.000584 | 0.001024 | 0.000431 | 0.000379 |
| 0.4 | 0.000148 | 0.000327 | 0.003070 | 0.000779 | 0.001366 | 0.000575 | 0.000505 |
| 0.5 | 0.000185 | 0.000409 | 0.003837 | 0.000974 | 0.001707 | 0.000719 | 0.000631 |
| 0.6 | 0.000222 | 0.000491 | 0.004604 | 0.001169 | 0.002048 | 0.000863 | 0.000757 |
| 0.7 | 0.000259 | 0.000573 | 0.005372 | 0.001364 | 0.002390 | 0.001006 | 0.000883 |
| 0.8 | 0.000296 | 0.000654 | 0.006139 | 0.001559 | 0.002731 | 0.001150 | 0.001010 |
| 0.9 | 0.000333 | 0.000736 | 0.006906 | 0.001753 | 0.003073 | 0.001294 | 0.001136 |
| 1.0 | 0.000370 | 0.000818 | 0.007674 | 0.001948 | 0.003414 | 0.001438 | 0.001262 |

Basing the values off of those previoulsy established in other models of *Escherichia coli* was the first step to determine the prtotein pool values for the rest of the models. Following the $P$ value determined in Adadi et al. [2012] of 0.095 $gdw^{-1}h^{-1}$, simulations were run in order to determine the feasibility of the model when put in a community setting with other models. After running the simulations, a value of 0.13 was deemed ideal as it provided most compatibility with the remaining models. Afterwards, in order to determine the values for the other models, adjustments were made based on both the ammount of enzymatic data was added in the models and how their protein pool should be in proportion to the already established value for *E. coli*. For that, various simulations were run on the samples in order to find a consensus that would make the simulations feasible and allowed growth, all while maintaning realistic values for all models. Moreover, this type of approach followed a more empyrical way of calibrating the models. After deliberating on the optimal values (Table 5), the models were then ready to be sampled.

Table 5: Chosen protein pool $P$ values for each model

| Model | Protein pool value ($gdw^{-1}h^{-1}$) |
|---|---|
| Bt | 0.075 |
| Bu | 0.075 |
| Ec | 0.13 |
| Fn | 0.046 |
| Ri | 0.078 |
| Sp | 0.082 |
| Ss | 0.048 |

This calibration process did not involve any kind of $k_{cat}$ adjustment as the lack of available enzymatic information across all models would create an imbalance in the models that would further cause troubles when the sampling and subsequent simulations were to occur. Nonetheless, the models were deemed prepared and the samples ready to be created, as will be further explained in Section 5.4.3.

## 5.4 Community simulations

### 5.4.1 Growth media

Growth media for the selected samples was chosen to portrait different levels of availability of both organic and inorganic iron in gut microbial communities, with the objective to ultimately observe how communities dynamics are affected during various stages of colitis when the host exerts metabolic pressure by reducing the availability of iron in the extracellular space, envisioning the observation of how this affects communal composition and promotes the adaption of the metabolic profiles of both indiviuduals and the community as a whole. The chosen growth media is represented in Table 6.

Table 6: Growth media for the simulations

(a) Iron availability

| Media | Constraints |
|---|---|
| M1 | $33\,\mu M$ Fe |
| M2 | $0\,\mu M$ Fe/H |
| M3 | $0.7\,\mu M$ H; $8.2\,\mu M$ Fe |
| M4 | $2.7\,\mu M$ H; $33\,\mu M$ Fe |
| M5 | $2.7\,\mu M$ H |
| M6 | $125\,\mu M$ Fe |
| M7 | $125\,\mu M$ H |
| M8 | $125\,\mu M$ H; $2\,mM$ Fe |
| M9 | $125\,\mu M$ H; $125\,\mu M$ Fe |
| M10 | No constraints |

(b) Composition

| Type | Compounds |
|---|---|
| Sugar | D-glucose, Fructose, Cellobiose, Maltose, Lactose, Inulin |
| Amino acids | L-Histidine, L-Isoleucine, L-Leucine, L-Methionine, L-Valine, L-Arginine, L-Cysteine, L-Glutamate, L-Phenylalanine, L-Proline, L-Asparagine, L-Aspartate, L-Glutamine, L-Serine, L-Threonine, L-Alanine, Glycine, L-Lysine, L-Tryptophan, L-Tyrosine |
| Nucleotides | Adenine, Guanine, Uracil, Xanthine |
| Salts and minerals | Magnesium Chloride |
| Vitamins and antioxidants | Biotin, Riboflavin, Folate, Niacin, Inositol, L-Glutathione reduced |
| Others | NAD |

The choosing of the above media reflects the need to emulate the conditions through which the *in vitro* experiments were conducted. This was achieved by emulating the **GMM+LAB plus Mucin** media [Tramontano et al., 2018] without mucin in which the *in vitro* experiments were conducted and defining a growth media with the available exchange reactions in the models. Through this, a comparison and

evaluation of the obtained results against those obtained experimentally can be made, allowing for the finer tuning of models in reaching the end goal of capturing observed mechanisms and interactions in those same studies.

### 5.4.2 Simulations

To predict the steady-state metabolic flux distributions in the selected microbial communities using genome-scale metabolic models, two approaches were employed: FBA and SteadyCom. In this section is present an overview of each one and their functioning, while explaining how they were applied to the current study as well as their advantages and disadvantages.

#### FBA

This approach is a direct extension of **Flux Balance Analysis (FBA)** (or joint **FBA**), which integrates metabolic reconstructions of individual microbial species into a multi-compartment model with a community compartment allowing for the exchange of metabolites between species. The optimization objective function is usually the sum of the biomass reactions of individual species (called the community biomass). Generally, this approach often requires additional constraints for capturing observed co-growth behavior. The following equation presents the linear problem through which this method operates.

Let **K** be the set of all organisms in the community. For an organism $k$ in **K**, the traditional FBA for predicting maximum growth can be stated as follows:

$$\begin{aligned} &\max \quad v^k_{\text{biomass}} \\ &\text{subject to} \sum_{j \in \mathbf{J}^k} S^k_{ij} v^k_j = 0, \quad \forall i \in \mathbf{I}^k \\ &LB^k_j \leq v^k_j \leq UB^k_j, \quad \forall j \in \mathbf{J}^k \end{aligned}$$

where $v^k_j$ is the flux of reaction $j$ (in $mmol\ gdw^{-1}h^{-1}$ for general metabolic reactions, in $gdw^{-1}h^{-1}$ for the biosynthesis of macromolecules, and in $h^{-1}$ for the biomass reaction), $S^k_{ij}$ is the stoichiometry for metabolite $i$ in reaction $j$, $LB^k_j$ and $UB^k_j$ are the lower bound and upper bounds for fluxes $v^k_j$ respectively, $\mathbf{I}^k$ and $\mathbf{J}^k$ are respectively the set of metabolites and reactions for organism $k$. All this allow the capturing of important features of microbial communities such as cross-feeding and competition.

Frameworks based on Flux Balance Analysis (FBA) exhibit, however, a fundamental omission. This stems from the fact that the biomass reaction flux not only serves as a sink for biomass constituents but

also quantifies the specific growth rate (in $h^{-1}$), thereby indicating the size of the system. In the case of a mono-culture, a single biomass flux is present, normalized by specific rates of consumption or production (in $mmol\ gdw^{-1}h^{-1}$). However, in the context of multiple organisms growing together, a unified FBA framework does not necessarily impose constraints on the growth rate of all participating members in the community. This absence of a consistent growth rate across all microbes within the community may result in metabolic fluxes that are inconsistent with a stable average community composition as the fastest growing organism will take over the population.

### SteadyCom

This approach surged as an answer to many of the problems stated before and is termed an extension of **Flux Balance Analysis** that takes into account the fact that to reach a stable composition the organisms need to grow at the same specific growth rate ($h^{-1}$), which means that the absolute growth rate ($gDW^{-1}/h^{-1}$) of each organism is proportional to its abundance at steady-state ($gDW^{-1}$). Using the following equation, the maximum community growth rate $\mu_{max}$ of a community satisfying the community steady-state can be found by solving the following non-linear optimization problem:

$$
\begin{aligned}
&\max^{\mu} \\
&\text{subject to} \quad \left[\begin{array}{ll}
\sum_{j\in \mathbf{J}^k} S_{ij}^k V_j^k = 0, & \forall i \in \mathbf{I}^k \\
LB_j^k X^k \le V_j^k \le UB_j^k X^k, & \forall j \in \mathbf{J}^k \\
V_{\text{biomass}}^k = X^k \mu & \\
X^k \ge 0 &
\end{array}\right] \quad \forall k \in \mathbf{K} \\
&u_i^c - e_i^c + \sum_{k\in\mathbf{K}} V_{ex(i)}^k = 0, \quad \forall i \in \mathbf{I}^{\text{com}} \\
&\sum_{k\in\mathbf{K}} X^k = X_0 \\
&, \quad e_i^c \ge 0, \quad \forall i \in \mathbf{I}^{\text{com}}
\end{aligned}
$$

For convenience the community export rates $e_i^c$ and uptake rates $u_i^c$ are normalized for one unit of total community biomass, therefore $X_0$ is set at 1 $gDW$ and $X^k$ is thus equal to the relative abundance of organism k. By setting $X^1$ = 1 and $X^k$ = 0 for k > 1, SteadyCom is reduced to the standard single-organism FBA model and the aggregate biomass flux coincides with the specific growth rate. Similarly to what happens in single-organism FBA, constraints on the system uptake rates $u_i^c$ are sufficient to guarantee a finite solution (i.e. finite $\mu_{max}$). Furthermore, predictions by SteadyCom are in general different from the predictions by joint FBA because of the constraints relating biomass, the bounds for specific rates, the aggregate fluxes and the community growth rate.

### 5.4.3 Sampling

The sampling step of the work here presented was of most importance, as it not only was a determinant in the whole workflow but also allowed for a better understanding of each organism's behaviour in the presence of different microbial organisms. The initial iteration of these samples involved the creation of groups of 3 models where 2 models prevailed in the entirety of the samples: the *Bacteroides thetaiotaomicron* VPI 5482 and *Bacteroides uniformis* ATCC 8492, with the third model being each one of the remnant 5 target models. This provided 5 initial samples from which simulations under many different conditions were conducted and results generated. From that it was observed that those communal compositions were not ideal, mainly due to the more resource demanding uptake profile of *B. uniformis* ATCC 8492 that hindered the growth of *B. thetaiotaomicron* VPI 5482 across all samples.

Considering the previously stated facts, 6 new samples were designed with the goal to study the growth of *B. thetaiotaomicron* VPI 5482 during colitis, with a focus of understanding the growth of this strain under the presence of other microbes with different consumption profiles observed in provided results from *in vitro* experiments and confirmed by means of running simulations in every single model under same abiotic conditions as those conducted experimentally. These new samples constituted the final target of this experiment and possibilitated both the corroboration of *in vitro* results in other studies and the observing of new results that further prove the precision and accuracy of these representations and simulations.

Table 7: Chosen samples (**S1** to **S6**) and present models

| **Organism** | **S1** | **S2** | **S3** | **S4** | **S5** | **S6** |
|---|---|---|---|---|---|---|
| *B. thetaiotaomicron* VPI 5482 | ● | ● | ● | ● | ● | ● |
| *B. uniformis* ATCC 8492 | ● | | | | | |
| *E. coli* ED1a | | ● | | | | |
| *F. nucleatum subsp. nucleatum* ATCC 25586 | | | ● | | | |
| *R. intestinalis* L1 82 | | | | ● | | |
| *S. parasanguinis* ATCC 15912 | | | | | ● | |
| *S. salivarius* DSM 20560 | | | | | | ● |

As displayed in Table 7, the chosen samples consist of community models that combine *B. thetaiotaomicron* VPI 5482 with each of the remaining models. This choice reflected the need to meet the end goal of infering on the growth of this strain under the presence of both commensal and pathogenic organisms under diferent concentrations of the chosen transition metals for this study case. Moreover, since all models present such different metabolic profiles when compared with *B. thetaiotaomicron*, with the exception of *B. uniformis* as it will be further demonstrated, it will aso provide valuable insights on the interactions that occur between 2 different microbes in conditions that induce metabolic stress.

## 5.5 Workflow

The creation and application of a workflow (Figure 8) that better suits the devised endeavour is vital to the overall success of the work. Here, an explanation of the entirety of the steps to be taken under the experiments made, with an emphasis on the methods and tools used along the entirety of the process.

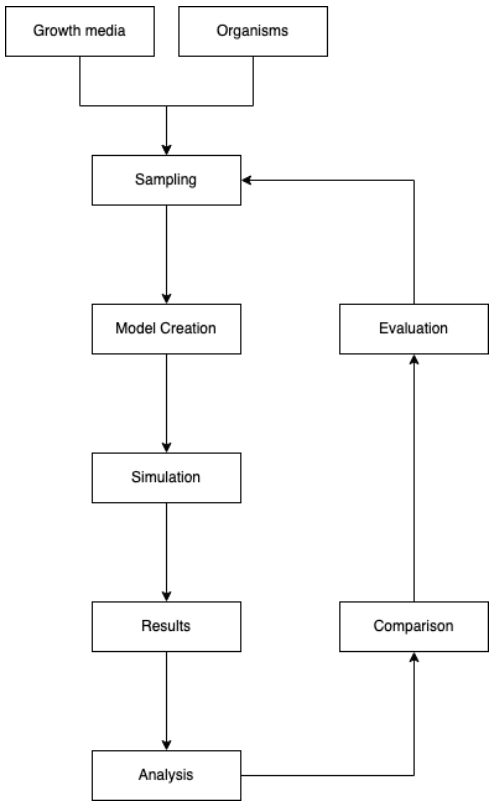



Figure 8: Visual representation of the established workflow.

When observing the representation of the established workflow in Figure 8 one can understand that the nature of the developed work followed a trial-error approach, therefore enabling the constant tuning of certain aspects in the search for improvement of the end result. Moreover, by implementing such plan, each iteration produced interesting results that, even if not optimal or considered worthy of mentioning, served as both a guideline and a blueprint for the results that shall be presented and discussed in Chapter 6.

The preliminary stage of sampling involved the creation of samples that constituted the study cases of this work. Firstly, desired species of microorganisms were selected according to *in vitro* studies performed parallel to this study. Secondly, the selected growth media were selected to better emulate dietary sources of many kinds and selected according to *in vitro* approaches. Finally, the samples were created from the combination of all these factors.

The creation of models encompassed the acquirement of **GEM**s of the selected organisms. The models were obtained as previously mentioned from the AGORA repository. The obtained models then underwent a curation process followed by an implementation of the **GECKO** method, in which they were be enhanced with enzymatic constraints using kinetic data supplied by *in vitro* results present in various databases, and also by making use of a Deep-Learning model that predicts turnover values based on substrates and enzymatic data. The entire process as well as the creation of community models from each singular **Genome-Scale Metabolic Models** was accomplished using both various methods and clasees from the MEWpy [Pereira et al., 2021] package and original code.

Simulations of singular and community models under the selected abiotic conditions was done using both **Flux Balance Analysis** and Steadycom [Chan et al., 2017], a simulation method that constitutes a generalization and an extension of FBA with the ability to predict the change in species abundance in response to changes in diets with minimal additional imposed constraints on the model. This method, by taking into account the fact that to reach a stable composition the organisms need to grow at the same specific growth rate ($1/h$), creates a direct proportionality between the absolute growth rate ($gDW/h$) of each organism and its abundance at steady-state ($gDW$).

The fourth step of the work plan consists on the obtention of results and analysis. This step was a successor of the simulations, as here the analysis of its results, as well as the creation of the first hypothesis surged. Finally, the ending step of each iteration of this cyclic workflow constituted a vital part of the work, as here comparisons were made between the obtained results and other results obtained both through previous simulations and those obtained using *in vitro* approaches under similar target conditions. Here, the precision of the encompassed method was also constantly evaluated, with the intention of assessing the

advantages and disadvantages of each method being employed, therefore seeking a balance that creates the possibility of developing a better method for the proposed work and leading to the development of better and improved ones. Finally, new samples were created and new changes were made to better enhance the method each time a cycle ended.

## 5.6 Code availability

The entirety of the code and obtained results is publicly available and can be used in its integrity or extended to conduct new studies. The source code for the developed project can be found at: https://github.com/areias03/tmCOM.git.

## 5.7 Chapter Summary

In this chapter, a summary of the steps devised for the execution of the intended experiments is presented, with a visual display of the planned work schedule also being highlighted.

# Chapter 6

# Results and Discussion

## 6.1 Model Analysis

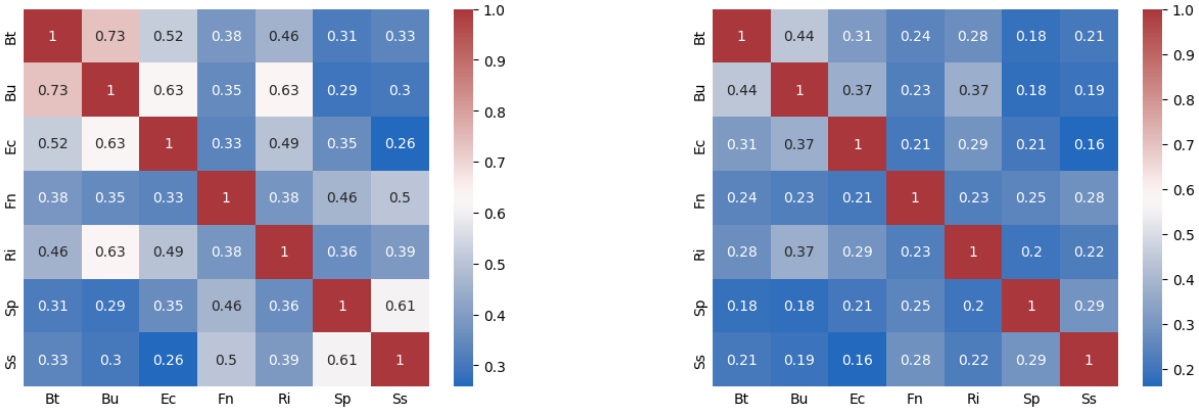
### 6.1.1 Similarity

The enhancement process to which the models were subjected was the origin of many changes in each model's metabolic profile. These changes, even if mainly affecting reaction's stoichiometry and protein molecular weights, caused also many changes in each model's naming of reactions and metabolites, as well as their organization by reversibility.

In Figure 9 it is possible to observe some of those changes. Regarding metabolite changes when considering the enhancement of models with enzymatic constraints through the process here employed, one must consider the addition of genes as pseudo-reactions a major factor that impacts the models. Specially, considering the sheer number of metabolites added this way causes, as observed, a decrease in similarity among the models.

In the case of the overlap of reactions in Figure 10, by taking into account the fact that mapped reactions's stoichiometries are changed and splitted by reversibility, it is expected that such changes in similarity occur after the process is concluded. Finally, in the case of the observed changes in resource overlap, no changes were seen as observed in Figure 11 as no changes made in the models as they were enhanced impacted the models' uptake reactions' bounds and therefore it did not cause a decrease in similarity in terms of uptake overlap.

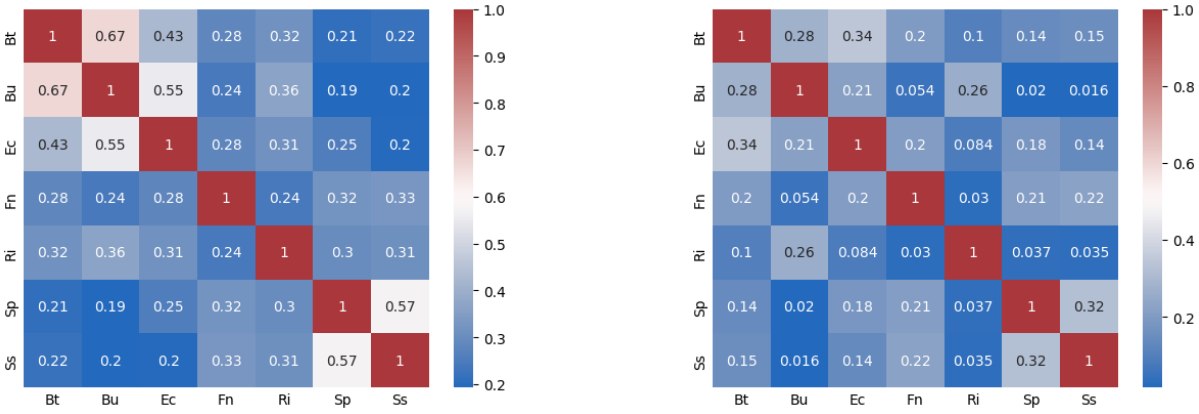
It is of most importance to note however, that the results here displayed may not be indicative of the real changes that occured in the models, as results displayed further in this section show many differences specially regarding the changes in uptake profiles of the various models.




Non-enhanced models

Enhanced models

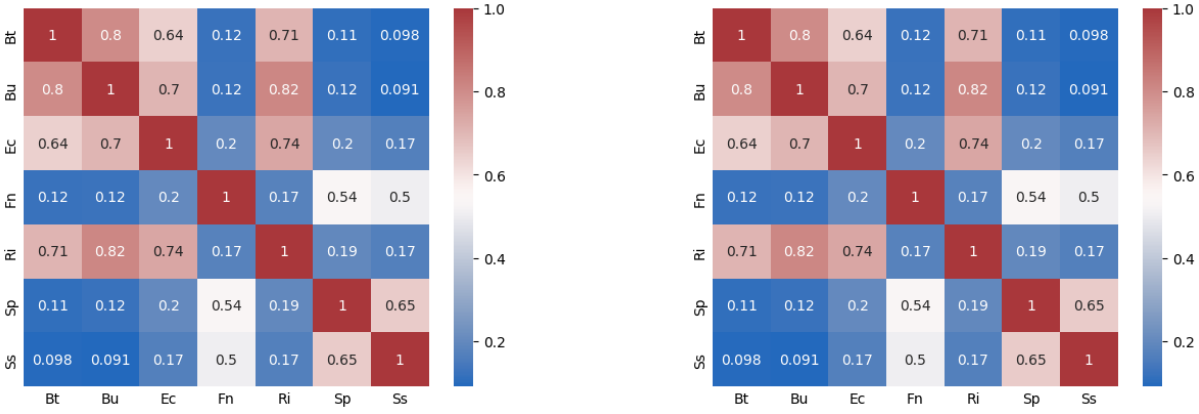
Figure 9: Metabolite overlap




Non-enhanced models

Enhanced models

Figure 10: Reaction overlap




Non-enhanced models

Enhanced models

Figure 11: Uptake overlap

### 6.1.2 Uptake profile

The understanding of the impact the enhancement process that every model went through had on each model's uptake profile was of serious importance to the entirety of the research. The obtained results can be seen in Figures 12, 13, 14 and 15.

The results observed in the growth media tests performed on the non-enhanced models reveals a rather visible uniformity that spans across almost all models independently from the growth media. These resulta can be attributed to the lack of constraints in the models that lead not only to the null profiling of both the media and models but also to the uncontrolled and unrealistic growth values here presented.

In comparison to the enzyme-enhanced models however, one can truly now start to observe that not only the growth values for the models appear much more realistic, but also a pattern that is befitting with the applied constraints in terms of both protein pool and the growth media. This characteristic reveals one of the many advantages of this implementation: the lack of necessity to extensively define a growth media. The enhancement of models, by applying constraints directly in the model, retains the ability to relax the definition of media *a posteriori*.

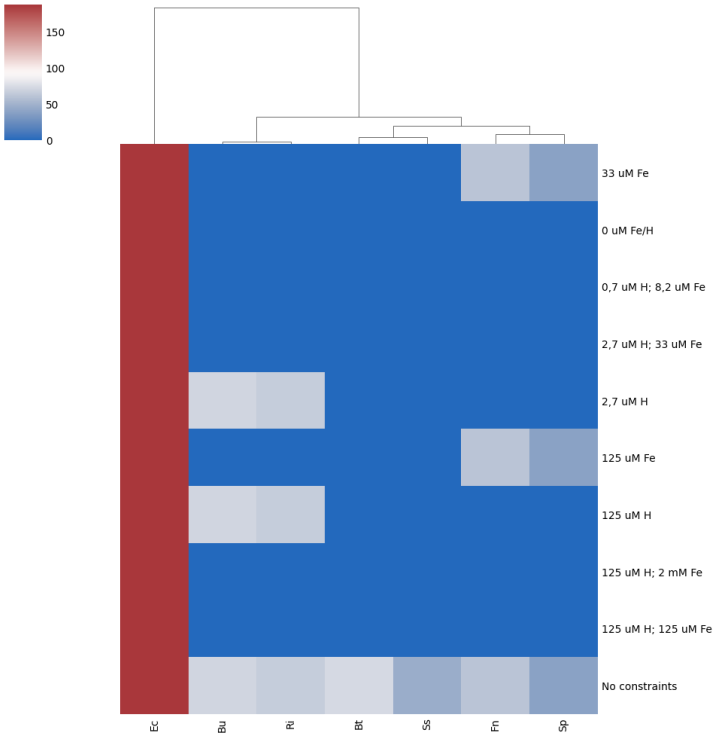



Figure 12: Growth of non-ec models through **Flux Balance Analysis**

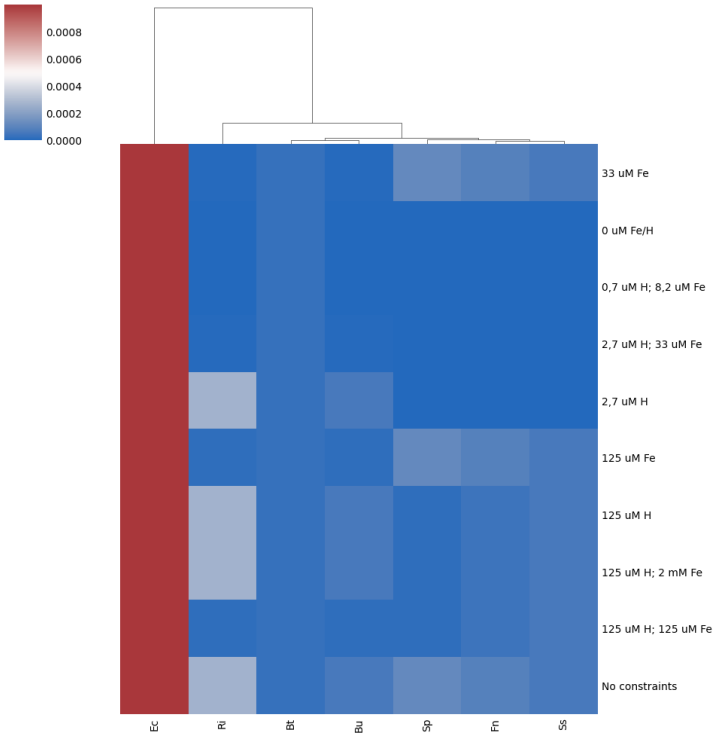



Figure 13: Growth of ec models through **Flux Balance Analysis**

The simulations performed in the community models under **Flux Balance Analysis (FBA)** further reinforced the ideas previously stated. In the non-enhanced samples, it is observable, as in the the singular models, a lack of profile and overall uniform growth across the samples that correlates, once again, to the lack of constraints. These samples, other than presenting unrealistic growths, do not possess any sort of differences in growth except when applied no constraints, as in that case, the maximum growth for each sample is obtained.

In enhaced samples, the exact contrary is observed. Other than attaining growth profiles, the samples display a larger similarity between samples, as is the case between samples 5 and 6, where *Bacteroides thetaiotaomicron* is paired with *Streptococcus parasanguinis* and *Streptococcus salivarius*, respectively. This sort of similarities, other than revealing of the profiling of communities, fall within the expected spectrum of results, as it would be expectable that samples with high similarity between species would retain similar uptake profiles.

Globally, the results obtained and presented in this section not only prove the success of the enhancement process but also further prove its importance in the obtention of results that are ever more precise and compatible with those obtained experimentally.

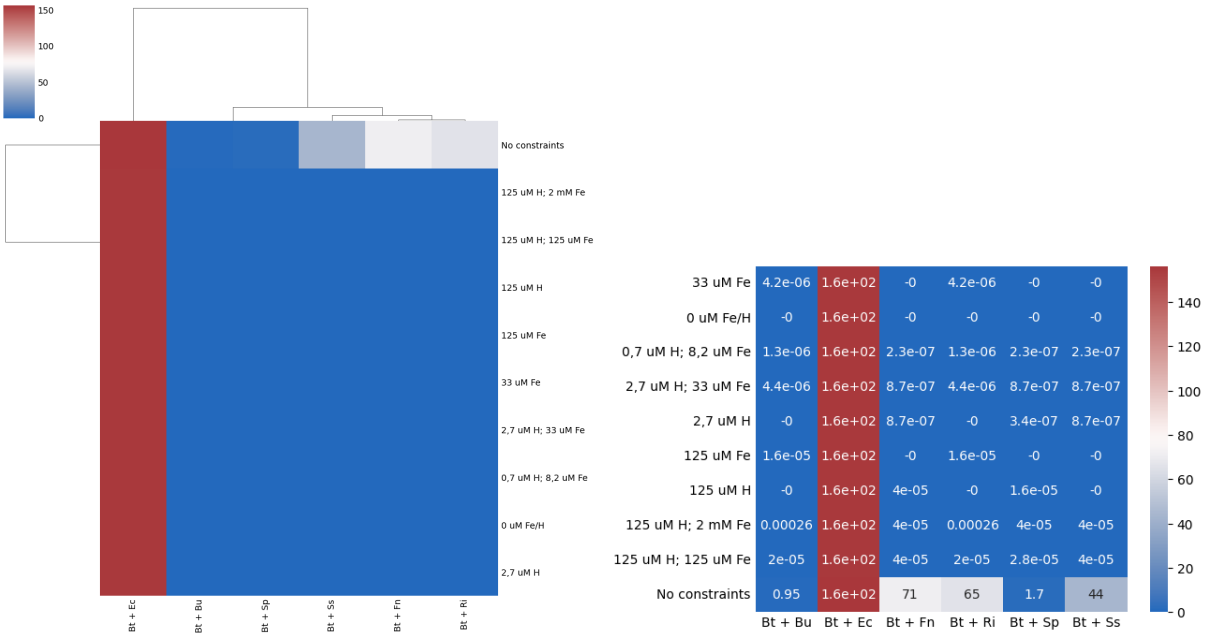



Figure 14: Growth of non-ec samples through **Flux Balance Analysis**

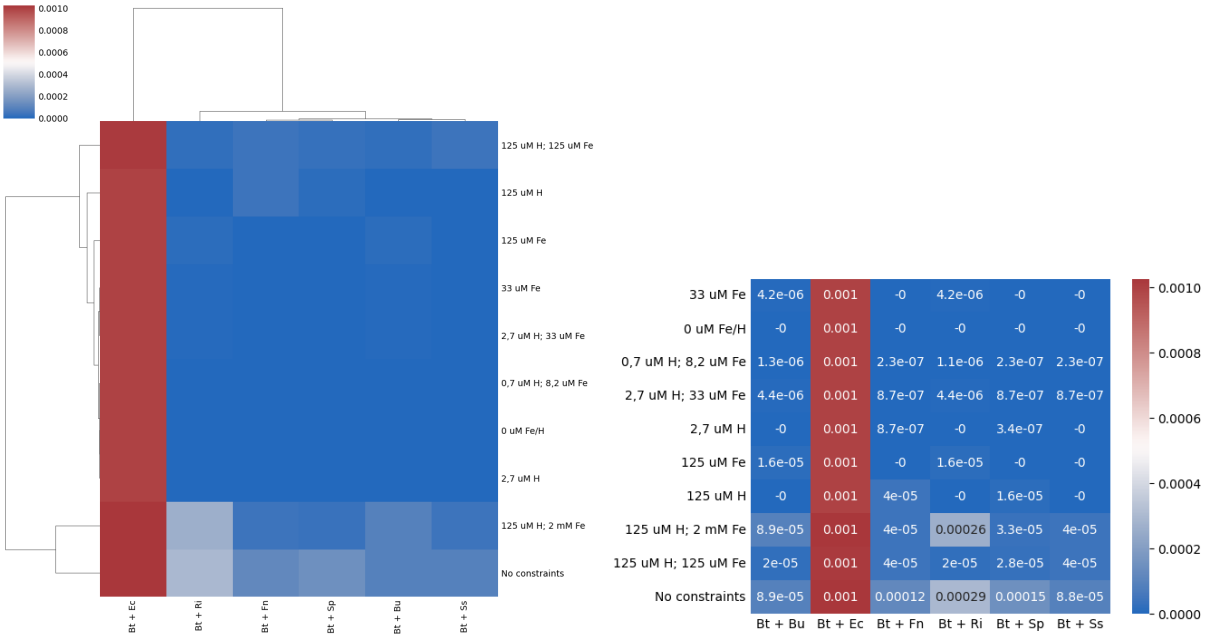



Figure 15: Growth of ec samples through **Flux Balance Analysis**

### 6.1.3 Community compostion and abundances

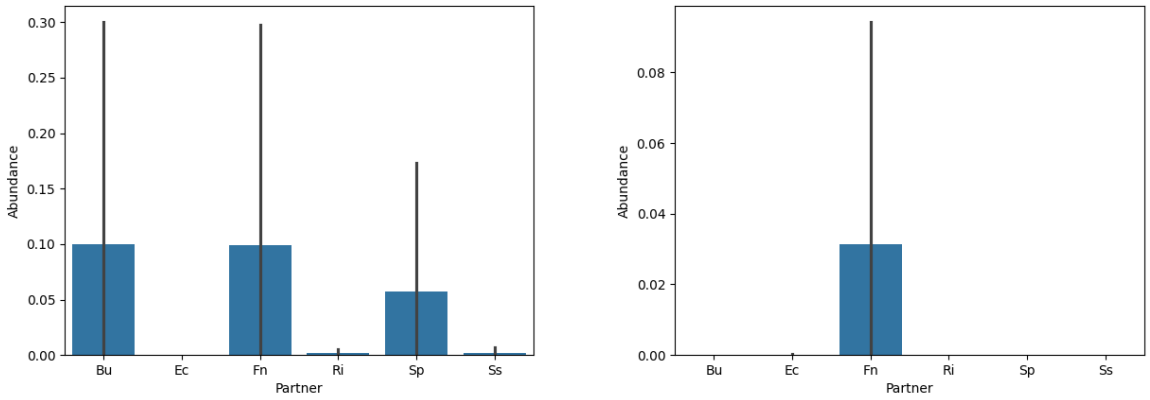
In order to infer on the achieved consortium on the selected samples regarding their composition and abundances, various approaches were applied. Firstly, abundances for each organism were obtained from the results in Tramontano et al. [2018], where afterwards the relative abundances were calculated for each organism from the selected models in relation to *B. thetaiotaomicron*. These abundances, present in Table 8 were then used to perform simulations with defined abundances for each sample with rather unsatisfactory results, especially regarding the obtained results when running simulations through SteadyCom where most samples would not present a plausible solution and abundances for more resource demanding organisms would be extreme.

Table 8: Relative abundaces from the selected models

| Model | Relative abundance |
|---|---|
| Bt | 1 |
| Bu | 4 |
| Ec | 0.5 |
| Fn | 1 |
| Ri | 1 |
| Sp | 1 |
| Ss | 1 |

To that effect, a new strategy consisiting of growing the samples under SteadyCom with no defined abundances was employed, with the objective of understanding the optimal abundaces necessary for each sample to achieve stability in the selected growth media. The results, displayed in Figure 16, further enforced the idea that in low levels of iron the *Bacteroides thetaiotaomicron* selected strain would have to be low in abundace in steady-state for the community to achieve balance.

By comparing the results obtained from enhaced and non-enhanced samples, one can observe that the abundance of *B. thetaiotaomicron* was higher in non-enhanced samples. While this fact is true, the community growth rates to be displayed further will show that, although the abundances are higher, there was little to no stability in the samples. The entirety of the results concerning the abundances can be seen in Tables 11 and 12 in the Details of Results appendix (Chapter 8).




Non-enhanced models

Enhanced models

Figure 16: Abundance of B. thetaiotaomicron under simulations through SteadyCom

## 6.2 Effects on community stability

Community stability and how to achieve it is an important aspect of microbial communities. When simulating the metabolism of microbial communities, one often finds that the multi compartment models denominated as community models seldomnly achieve such state as models often fail to capture the full extent of the metabolism of microbes. In this study, the creation of enzyme-constrained models aimed to achieve such point that allowed the samples to grow effectively.

In non-enhanced community models, it is often common to find that communities do not achieve a stable consortium as some members of the community consume most resources, therefore hampering the growth of less resource-requiring microbes. This happens mainly due to the fact that, in non-enhanced models, there is no constraints that stop the consumption of rersources by any reaction of any model. With the enhancement of models with enzymatic constraints, it was expected that such change should be seen, and it was.

When running simulations through **Flux Balance Analysis**, it was observed that across all samples, enzyme-constrained or not, there was growth in almost all media that maintained a reasonable level of organic or inorganic iron. This is mainly due to the fact that, as previously mentioned in Section 5.4.2, **FBA** fails to impose constraints on the growth of all models across the community, allied to the fact that the models do not possess any constraints *a priori*. Therefore, when employing this simulation method, growth of members of non-enhanced communities happened uncontrollably and in a manner that does not reflect a realistic way. In enzyme-enhanced communities, the obtained growth results reflect a much higher approximation to reality as the constraints imposed on models happen independently from the

simulation method itself. An interesting result surged, however, when running the simulations through SteadyCom. Here, due to the fact that the constraints are applied to all members of a community and the consistency in terms of growth rates amongst all microbes, it was observed that only enzyme-constrained were able to grow in low iron conditions. This observation further leads to the conclusion that the enzymatic enhancement of models further extended the capabilities of **Genome-Scale Metabolic Models (GEM)** by capping the overtaking of resources by species that possess a higher rate of consumption of metabolites. Moreover, it is also possible to conclude that this process confered stability to the communities as the hindering of growth for all models applied by both constraints employed by SteadyCom and the enzymatic enhancement of models further contributed to a more controlled growth of all members of each community in spite of their metabolic profile that promoted the favoring of organisms with more catalytically efficient enzymes with a lower synthesis cost.

## 6.3 Bacteroides thetaiotaomicron growth during colitis

In the results published in Zhu et al. [2020] it was reported that there was observed growth of *Bacteroides thetaiotaomicron* in community with species from the *Enterobactericiae* family during colitis, where the acquisition of iron was mediated through the utilization of xenosiderophores of *Enterobactericiae* like *Escherichia coli* and *Salmonella Typhimurium*. Moreover, these results represent a mechanism of commensalism between species in gut microbial communities that further reflect the complex interactions that occur in these settings and whose characteristics are the goal of this research.

To that end, it was decided that part of these results would infer on the growth of the samples under different levels of iron and try to observe and capture, to some extent, similarities between the *in vitro* and *in silico* experiments. Moreover, to achieve this, simulations using both **FBA** (Figure 17) and SteadyCom (Tables 9 and 10) were performed using the different samples under the different growth media, therefore allowing the comparison between both the results obtained using the different simulation methods and the observed differences in enzyme-enhanced and non-enhanced samples.

The results obtained from the **Flux Balance Analysis** simulations were rather surprising. On one side, even if the full extent of the mechanism was not meant to be captured using this methodology, the hope would be of it to actually point towards that point. On the other, the expectancy that the enhancement of models would be the solution to the entirety of the problem would be somewhat farfetched. Moreover, since **FBA** does not apply the constraints equally across all members of a commmunity, the fact that organisms like *E. coli* ED1a grow more than *B. thetaiotaomicron* VPI 5482 in a community is within the

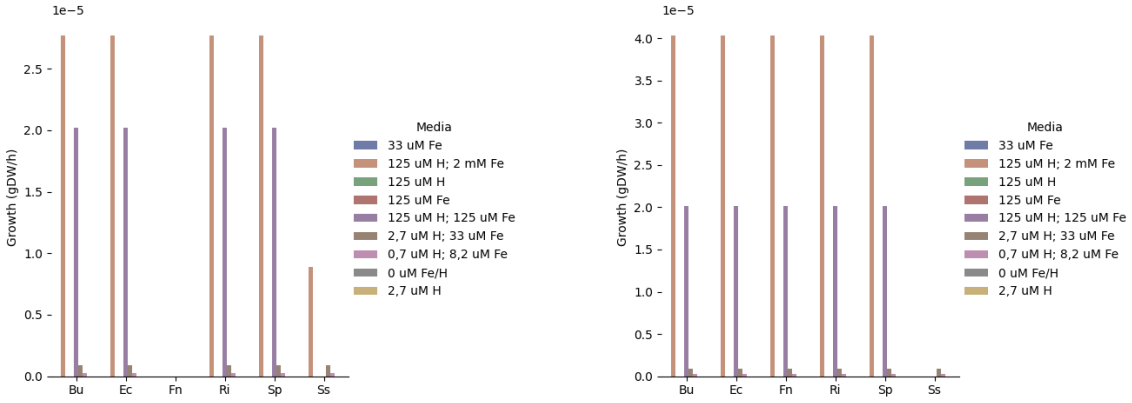



Figure 17: Growth of *B. thetaiotaomicron* in different communities through **Flux Balance Analysis**

expected range of results. One contradictory result to the expected however, happened in sample 6, where the really different uptake profiles observed in Figure 7 between *B. thetaiotaomicron* and *S. salivarius* would have created a sample that would promote a lot of cross-feeding interactions and therefore, growth in situations of low levels of trace metals. It is worthy to note however that, as observed in Figure 13, the similarities in growth suggest that both models do not possess great fitness in terms of adaptability to low levels of iron. More interesting results surged however, when running the simulations with SteadyCom (Tables 9 and 10). Here, a tendency was observed that was constant across enhanced and non-enhanced samples.

When observing the results presented in Table 9, it is possible to observe the occurence of a pattern in the growth rates. While sample 2 (Bt + Ec) was able to maintain a near stable growth rate across all media, the remaining samples were only able to grow in conditions wherre there were no restrictions in terms of iron availability. Moreover, with the exception of samples 1 (Bt + Bu) and 5 (Bt + Sp) all samples presented exorbitant growth rates, that are indicative of their unrestrained ability to grow under media that is not well defined.

It is however, a positive indication of the proof that these results hope to achieve, that despite the near constant growth of sample 2 it was the only one that grew under fluctuating levels of both organic and inorganic iron. Concerning the abundances for each result however, as seen in Table 11 that in sample 2 the abundance for *Bacteroides thetaiotaomicron* was 0 at steady-state across all media, meaning that even though there was community growth, the lack of constraints across all models of each samples made it impossible for a balance in terms of abundances to occur.

Table 9: Growth of non-enhanced samples through SteadyCom

| Constraints | Bt + Bu | Bt + Ec | Bt + Fn | Bt + Ri | Bt + Sp | Bt + Ss |
|---|---|---|---|---|---|---|
| 33 uM Fe | 0 | 153.504883 | 0 | 0 | 0 | 0 |
| 0 uM Fe/H | 0 | 153.504883 | 0 | 0 | 0 | 0 |
| 0,7 uM H; 8,2 uM Fe | 0 | 153.504883 | 0 | 0 | 0 | 0 |
| 2,7 uM H; 33 uM Fe | 0 | 153.504883 | 0 | 0 | 0 | 0 |
| 2,7 uM H | 0 | 153.504883 | 0 | 0 | 0 | 0 |
| 125 uM Fe | 0 | 153.504883 | 0 | 0 | 0 | 0 |
| 125 uM H | 0 | 153.504883 | 0 | 0 | 0 | 0 |
| 125 uM H; 2 mM Fe | 0 | 153.504883 | 0 | 0 | 0 | 0 |
| 125 uM H; 125 uM Fe | 0 | 153.504883 | 0 | 0 | 0 | 0 |
| No constraints | 0.954102 | 153.635742 | 30.661133 | 54.433594 | 1.657227 | 41.800781 |

Table 10: Growth of enzyme-enhanced samples through SteadyCom

| Constraints | Bt + Bu | Bt + Ec | Bt + Fn | Bt + Ri | Bt + Sp | Bt + Ss |
|---|---|---|---|---|---|---|
| 33 uM Fe | 0 | 5.179688 | 0 | 0 | 0 | 0 |
| 0 uM Fe/H | 0 | 5.179688 | 0 | 0 | 0 | 0 |
| 0,7 uM H; 8,2 uM Fe | 0 | 5.179688 | 0 | 0 | 0 | 0 |
| 2,7 uM H; 33 uM Fe | 0 | 5.179688 | 0 | 0 | 0 | 0 |
| 2,7 uM H | 0 | 5.179688 | 0 | 0 | 0 | 0 |
| 125 uM Fe | 0 | 5.179688 | 0 | 0 | 0 | 0 |
| 125 uM H | 0 | 5.179688 | 0 | 0 | 0 | 0 |
| 125 uM H; 2 mM Fe | 0 | 5.179688 | 0 | 0 | 0 | 0 |
| 125 uM H; 125 uM Fe | 0 | 5.179688 | 0 | 0 | 0 | 0 |
| No constraints | 0.477539 | 5.185547 | 0.369141 | 2.761719 | 1.378906 | 0.597656 |

Even though the previously observed pattern in the the non-enhanced samples persisted, there were some changes. In sample 2, a near constant growth rate of approximately 5.2 $gdw^{-1}h^{-1}$ was maintained across all media and similarly positive growth rates were maintained in the remaining samples in conditons where there was no restrainment in terms of iron levels. It was however in the abundances (Table 12) that a shift in the pattern occured. Here, not only has *B. thetaiotaomicron* sustained a high abundace of approximately 31% in sample 3 under no constraints but also, even if prominently low, a positive abundance in sample 2 in 3 different media (M3, M4 and M7) further proving that, even if under extremely low abundances, *Bacteroides thetaiotaomicron* can grow under low transition metal levels in community settings with other microbes belonging to the *Enterobacteriaceae* family like *Escherichia coli*. These results are a extremely positive demonstration that the extension of the models with enzyme constraints is a step towards the ability to capture mechanisms that occur within gut microbial communities. Furthermore, it is possible to conclude that if the a better adjustment of the $k_{cat}$ values was to be made, these results could be further improved.

## 6.4 Limitations

The obtained results were hindered to some extent mainly by both the quality of the models and the inability to better tune these same models, mainly due to the lack of available enzymatic information. Furthermore, even with the continuous surging of simulation methods applied to the simulation of microbial communities, there is still no definitive solution to the challenges encountered when using all of the ones here presented. It is of upmost importance to note, however, that none of these factors were a main cause of impedance to the obtention of relevant results in any stage of this work, and rather a challenge to overcome.

## 6.5 Chapter Summary

In this chapter the results produced by the established framework were presented and analyzed, in accordance with the designed experiments. This process resorted mainly in the comparison of the obtained growths and abundancdes for each set of samples (enzyme-enhanced and not) relatively to the case studies performed, determined by the selected growth media.

The first component of the results was related to the observed changes in the models in terms of similarity before and after the process of enhancement with enzymatic constraints. The results here presented provided strong insights on the impact the enhancement had on the models and set the ground for the results to be presented next.

Next, the differences in uptake profiles of the models shedded lights on the impact that the applied constraints in the models applied on each model's uptake profile and how they correlate with each other. This revealed crucial as the profiling of each model futher allows the comprehension on how constrainnig a model impacts the entire metabolic network. The next set of results comprised of a study on community composition performed that, while not being impactful on the overall work, provided interesting views on how the calibration of each model impacted the final result.

Finally, one of the most important end goals of this work was met as the an atempt to capture an observed mechanism *in vitro* was made. In this section of results, the growth results for the various samples in the selected media was presented as they were simulated through SteadyCom. These results are of most importance to this research and even if lacking some improvement, served as an important basis that further proves the relevance of these types of approaches.

# Chapter 7

# Conclusions

## 7.1 Summary

The objective of this endeavour was to develop a methodology capable of accurately simulate the behaviour of gut microbial communities under fluctuating levels of transition metals. The first step was to review the state-of-the-art, understand the approaches currently being used in this field and try to leverage the combination of some of these methodologies with other and newer ones. This review revealed a lot of interesting approaches and methodologies being employed with the capacity to extend our current knowledge of the interactions that occur within both the microbe-microbe and microbe-host axises. After this review, the desired set of methodologies were selected with the intention of them being applied and extended towards this study case. The selected methodologies included the enhancement of models with enzymatic constraints and the use of various simulation methods like joint **Flux Balance Analysis (FBA)** and SteadyCom, and after the worflow was designed, it was then applied.

The established workflow included the use of *in vitro* results from two main sources: optical density measurements of several microbial strains that reflected their growth under different levels of trace metals provided by the secondary counselor to this dissertation and those published in Zhu et al. [2020] and Tramontano et al. [2018] to base various aspects of the work from, including target mechanisms to capture and abundances for the samples.

Next, the application of the devised approach lead to the obtention of interesting results that were both interesting and encouraging in terms of potential.

Firstly, the produced results demonstrated the importance of the enhancement of models with enzymatic constraints in the stability of a community and further proved their importance in the search for accurate representations of the community space. In another, the results were to some extent set back by the lack of available information that could be crucial in the calibration of each model. Moreover, this hinderance further leaves the possibility for further works to extend and upgrade the work here conducted.

Given the obtained results, it is possible to conclude that the major goal of creating a community modelling method for capturing microbial gut community dynamics regarding fluctuations of transition metals were attained.

## 7.2 Problems and challenges

As previously mentioned in Sections 5.2 and 6.4, even if the complete development of the framework was successful, there were some obstructions both in the development and the obtention of results. The first set of challenges regard the difficulties encountered when dealing with the models and their curation process. This proved itself as rather difficult to overcome as the semi-automated for of geberation from which they were created often leads to problems in the models that in several occasions deficulted or even impeded the progressing of both the enhancement of the models or the simulations. The latter refers mainly to the lack of integration of enzymatic data from either literature or *in vitro* results that promoted a more trial-and-error approach to the calibration process of the models. Even if, to some extent, lacking accuracy, the calibrated models performed within their expected manner and produced results that are worth of notice and upgrade.

## 7.3 Contributions

The presented endeavour introduced several outcomes that have the potential to contribute to the further development of microbial community simulation methods. Those contributions are:

- Development of an automated framework to generate enzyme-enhaced models;
- tmCOM, a framework that possesses many tools and scripts used in this study and can be further extended to other use cases of microbial community simulations;
- Analysis and validation of *in vitro* results regarding the effect of transition metals in gut microbial communities.

## 7.4 Future Work

The developed work presents a foundation for future works through which this can serve as both an extension and an improvement of that employed here, in the search for better and more accurate results. For that effect, future tasks might include:

- Creation of metabolic models from a non-automated approach;
- Better curation of models;
- Integration of proteomics and other enzymatic data in the calibration process.

# Bibliography


Roi Adadi, Benjamin Volkmer, Ron Milo, Matthias Heinemann, and Tomer Shlomi. Prediction of Microbial Growth Rate versus Biomass Yield by a Metabolic Network with Kinetic Parameters. *PLOS Computational Biology*, 8(7):e1002575, May 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002575.

Hiroyuki Akama, Misa Kanemaki, Masato Yoshimura, Tomitake Tsukihara, Tomoe Kashiwagi, Hiroshi Yoneyama, Shin-ichiro Narita, Atsushi Nakagawa, and Taiji Nakae. Crystal structure of the drug discharge outer membrane protein, oprm, of pseudomonas aeruginosa: dual modes of membrane anchoring and occluded cavity end. *Journal of Biological Chemistry*, 279(51):52816–52819, 2004.

Claudia Andreini, Lucia Banci, Ivano Bertini, and Antonio Rosato. Counting the zinc-proteins encoded in the human genome. *Journal of proteome research*, 5(1):196–201, 2006.

Hélène Botella, Pascale Peyron, Florence Levillain, Renaud Poincloux, Yannick Poquet, Irène Brandli, Chuan Wang, Ludovic Tailleux, Sylvain Tilleul, Guillaume M Charrière, et al. Mycobacterial p1-type atpases mediate resistance to zinc poisoning in human macrophages. *Cell host & microbe*, 10(3): 248–259, 2011.

Hélène Botella, Gustavo Stadthagen, Geanncarlo Lugo-Villarino, Chantal de Chastellier, and Olivier Neyrolles. Metallobiology of host–pathogen interactions: an intoxicating new insight. *Trends in microbiology*, 20(3):106–112, 2012.

Elizabeth A Cameron and Vanessa Sperandio. Frenemies: signaling and nutritional integration in pathogen-microbiota-host interactions. *Cell host & microbe*, 18(3):275–284, 2015.

Jorge Carrasco Muriel, Christopher Long, and Nikolaus Sonnenschein. Simultaneous application of enzyme and thermodynamic constraints to metabolic models using an updated Python implementation of GECKO. *Microbiology Spectrum*, 0(0):e01705–23, October 2023. doi: 10.1128/spectrum.01705-23.

Siu Hung Joshua Chan, Margaret N Simons, and Costas D Maranas. Steadycom: predicting microbial abundances while ensuring community stability. *PLoS computational biology*, 13(5):e1005539, 2017.

James R Clark, Stuart J Daines, Timothy M Lenton, Andrew J Watson, and Hywel TP Williams. Individual-based modelling of adaptation in marine microbial populations using genetically defined physiological parameters. *Ecological modelling*, 222(23-24):3823–3837, 2011.

Katharine Z Coyte, Jonas Schluter, and Kevin R Foster. The ecology of the microbiome: networks, competition, and stability. *Science*, 350(6261):663–666, 2015.

Iván Domenzain, Benjamín Sánchez, Mihail Anton, Eduard J Kerkhoven, Aarón Millán-Oropeza, Céline Henry, Verena Siewers, John P Morrissey, Nikolaus Sonnenschein, and Jens Nielsen. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using gecko 2.0. *Nature Communications*, 13(1):3766, 2022.

Matthew J Fagan and Milton H Saier. P-type atpases of eukaryotes and bacteria: sequence analyses and construction of phylogenetic trees. *Journal of molecular evolution*, 38:57–99, 1994.

Yue Fu, Ho-Ching Tiffany Tsui, Kevin E Bruce, Lok-To Sham, Khadine A Higgins, John P Lisher, Krystyna M Kazmierczak, Michael J Maroney, Charles E Dann III, Malcolm E Winkler, et al. A new structural paradigm in copper resistance in streptococcus pneumoniae. *Nature chemical biology*, 9(3):177, 2013.

Ahmed Gaballa and John D Helmann. Bacillus subtilis cpx-type atpases: characterization of cd, zn, co and cu efflux systems. *Biometals*, 16:497–505, 2003.

Sanjoy Ghosh, Chuanbin Dai, Kirsty Brown, Ethendhar Rajendiran, Samantha Makarenko, Jessica Baker, Caixia Ma, Swagata Halder, Marinieve Montero, Vlad-Andrei Ionescu, et al. Colonic microbiota alters host susceptibility to infectious colitis by modulating inflammation, redox status, and ion transporter gene expression. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 301(1):G39–G49, 2011.

Manuel Glöckler, Andreas Dräger, and Reihaneh Mostolizadeh. Ncmw: a python package to analyze metabolic interactions in the nasal microbiome. *Frontiers in Bioinformatics*, 2:11, 2022.

Valentina Gogulancea, Rebeca González-Cabaleiro, Bowen Li, Denis Taniguchi, Pahala Gedara Jayathilake, Jinju Chen, Darren Wilkinson, David Swailes, Andrew Stephen McGough, Paolo Zuliani, et al. Individual based model links thermodynamics, chemical speciation and environmental conditions to microbial growth. *Frontiers in microbiology*, 10:1871, 2019.

Cyril Guilhen, Muhamed-Kheir Taha, and Frédéric J Veyrier. Role of transition metal exporters in virulence: the example of neisseria meningitidis. *Frontiers in Cellular and Infection Microbiology*, 3:102, 2013.

Martín Gutiérrez, Paula Gregorio-Godoy, Guillermo Perez del Pulgar, Luis E Muñoz, Sandra Sáez, and Alfonso Rodríguez-Patón. A new improved and extended version of the multicell bacterial simulator gro. *ACS synthetic biology*, 6(8):1496–1508, 2017.

Christopher J Haney, Gregor Grass, Sylvia Franke, and Christopher Rensing. New developments in the understanding of the cation diffusion facilitator family. *Journal of Industrial Microbiology and Biotechnology*, 32(6):215–226, 2005.

Christopher S. Henry, Matthew DeJongh, Aaron A. Best, Paul M. Frybarger, Ben Linsay, and Rick L. Stevens. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, 28(9):977–982, September 2010. ISSN 1546-1696. doi: 10.1038/nbt.1672.

Eitan Hoch, Wei Lin, Jin Chai, Michal Hershfinkel, Dax Fu, and Israel Sekler. Histidine pairing at the metal transport site of mammalian znt transporters controls zn2+ over cd2+ selectivity. *Proceedings of the National Academy of Sciences*, 109(19):7202–7207, 2012.

Tim A Hoek, Kevin Axelrod, Tommaso Biancalani, Eugene A Yurtsev, Jinghui Liu, and Jeff Gore. Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. *PLoS Biology*, 15(6):e1002606, 2017.

Nobuhiko Kamada, Yun-Gi Kim, Ho Pan Sham, Bruce A Vallance, José L Puente, Eric C Martens, and Gabriel Núñez. Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science*, 336(6086):1325–1329, 2012.

Zachary A. King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A. Lerman, Ali Ebrahim, Bernhard O. Palsson, and Nathan E. Lewis. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44(D1):D515–D522, January 2016. ISSN 0305-1048. doi: 10.1093/nar/gkv1049.

Chunxia Li, Jun Tao, Daqing Mao, and Chaozu He. A novel manganese efflux system, yebn, is required for virulence by xanthomonas oryzae pv. oryzae. *PLoS One*, 6(7):e21983, 2011.

Feiran Li, Le Yuan, Hongzhong Lu, Gang Li, Yu Chen, Martin K. M. Engqvist, Eduard J. Kerkhoven, and Jens Nielsen. Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nature Catalysis*, 5(8):662–672, August 2022. ISSN 2520-1158. doi: 10.1038/s41929-022-00798-z.

H Liesegang, K Lemke, RA Siddiqui, and HG Schlegel. Characterization of the inducible nickel and cobalt resistance determinant cnr from pmol28 of alcaligenes eutrophus ch34. *Journal of Bacteriology*, 175 (3):767–778, 1993.

Elena Litchman, Mark D Ohman, and Thomas Kiørboe. Trait-based approaches to zooplankton communities. *Journal of plankton research*, 35(3):473–484, 2013.

Feng Long, Chih-Chia Su, Hsiang-Ting Lei, Jani Reddy Bolla, Sylvia V Do, and Edward W Yu. Structure and mechanism of the tripartite cuscba heavy-metal efflux complex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1592):1047–1058, 2012.

Christopher A Lopez and Eric P Skaar. The impact of dietary transition metals on host-bacterial interactions. *Cell host & microbe*, 23(6):737–748, 2018.

Min Lu, Jin Chai, and Dax Fu. Structural basis for autoregulation of the zinc transporter yiip. *Nature structural & molecular biology*, 16(10):1063–1067, 2009.

Stefanía Magnúsdóttir, Almut Heinken, Laura Kutt, Dmitry A Ravcheev, Eugen Bauer, Alberto Noronha, Kacy Greenhalgh, Christian Jäger, Joanna Baginska, Paul Wilmes, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature biotechnology*, 35 (1):81–89, 2017.

Satoshi Murakami, Ryosuke Nakashima, Eiki Yamashita, and Akihito Yamaguchi. Crystal structure of bacterial multidrug efflux transporter acrb. *Nature*, 419(6907):587–593, 2002.

Caitlin C Murdoch and Eric P Skaar. Nutritional immunity: the battle for nutrient metals at the host–pathogen interface. *Nature Reviews Microbiology*, pages 1–14, 2022.

LV Nedorezov. The dynamics of the lynx–hare system: an application of the lotka–volterra model. *Biophysics*, 61(1):149–154, 2016.

DIETRICH H Nies and SIMON Silver. Plasmid-determined inducible efflux is responsible for resistance to cadmium, zinc, and cobalt in alcaligenes eutrophus. *Journal of bacteriology*, 171(2):896–900, 1989.

Lauren D Palmer and Eric P Skaar. Transition metals and virulence in bacteria. *Annual review of genetics*, 50:67, 2016.

Francesca Parise, John Lygeros, and Jakob Ruess. Bayesian inference for stochastic individual-based models of ecological systems: a pest control simulation study. *Frontiers in Environmental Science*, 3: 42, 2015.

Ian T Paulsen, Jay H Park, Peter S Choi, and Milton H Saier Jr. A family of gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from gram-negative bacteria. *FEMS microbiology letters*, 156(1):1–8, 1997.

Vítor Pereira, Fernando Cruz, and Miguel Rocha. Mewpy: a computational strain optimization workbench in python. *Bioinformatics*, 37(16):2494–2496, 2021.

Jonathan L Robinson, Pınar Kocabaş, Hao Wang, Pierre-Etienne Cholley, Daniel Cook, Avlant Nilsson, Mihail Anton, Raphael Ferreira, Iván Domenzain, Virinchi Billa, et al. An atlas of human metabolism. *Science signaling*, 13(624):eaaz1482, 2020.

Helmuth Sprinz, Donald W Kundel, Gustave J Dammin, Richard E Horowitz, Herman Schneider, and Samuel B Formal. The response of the germ-free guinea pig to oral bacterial challenge with escherichia coli and shigella flexneri: with special reference to lymphatic tissue and the intestinal tract. *The American journal of pathology*, 39(6):681, 1961.

Frank Nils Stähler, Stefan Odenbreit, Rainer Haas, Julia Wilrich, Arnoud H. M. Van Vliet, Johannes G. Kusters, Manfred Kist, and Stefan Bereswill. The Novel Helicobacter pylori CznABC Metal Efflux Pump Is Required for Cadmium, Zinc, and Nickel Resistance, Urease Modulation, and Gastric Colonization. *Infection and Immunity*, 74(7):3845–3852, July 2006. ISSN 0019-9567. doi: 10.1128/IAI.02025-05.

Reed Taffs, John E Aston, Kristen Brileya, Zackary Jay, Christian G Klatt, Shawn McGlynn, Natasha Mallette, Scott Montross, Robin Gerlach, William P Inskeep, et al. In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC systems biology*, 3:1–16, 2009.

Melanie Tramontano, Sergej Andrejev, Mihaela Pruteanu, Martina Klünemann, Michael Kuhn, Marco Galardini, Paula Jouhten, Aleksej Zelezniak, Georg Zeller, Peer Bork, Athanasios Typas, and Kiran Raosaheb Patil. Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nature Microbiology*, 3(4):514–522, April 2018. ISSN 2058-5276. doi: 10.1038/s41564-018-0123-9.

Naomi Iris van den Berg, Daniel Machado, Sophia Santos, Isabel Rocha, Jeremy Chacón, William Harcombe, Sara Mitri, and Kiran R Patil. Ecological modelling approaches for predicting emergent properties in microbial communities. *Nature Ecology & Evolution*, pages 1–11, 2022.

Naomi Iris van den Berg, Daniel Machado, Sophia Santos, Isabel Rocha, Jeremy Chacón, William Harcombe, Sara Mitri, and Kiran R. Patil. Ecological modelling approaches for predicting emergent properties in microbial communities. *Nature Ecology & Evolution*, 6(7):855–865, July 2022. ISSN 2397-334X. doi: 10.1038/s41559-022-01746-7.

André Van Gossum and Jean Neve. Trace element deficiency and toxicity. *Current Opinion in Clinical Nutrition & Metabolic Care*, 1(6):499–507, 1998.

Frédéric J Veyrier, Ivo G Boneca, Mathieu F Cellier, and Muhamed-Kheir Taha. A novel metal transporter mediating manganese export (mntx) regulates the mn to fe intracellular ratio and neisseria meningitidis virulence. *PLoS pathogens*, 7(9):e1002261, 2011.

Eberhard Otto Voit, Jacob D Davis, Daniel Vigário Olivença, and Sam P Brown. Methods of quantifying interactions among populations using lotka-volterra models. *Frontiers in Systems Biology*, page 35.

Ashley J Worlock and Ronald L Smith. Zntb is a novel zn2+ transporter in salmonella enterica serovar typhimurium. *Journal of bacteriology*, 184(16):4369–4373, 2002.

Jason SL Yu, Clara Correia-Melo, Francisco Zorrilla, Lucia Herrera-Dominguez, Mary Y Wu, Johannes Hartl, Kate Campbell, Sonja Blasche, Marco Kreidl, Anna-Sophia Egger, et al. Microbial communities form rich extracellular metabolomes that foster metabolic interactions and promote drug tolerance. *Nature microbiology*, 7(4):542–555, 2022.

Zuzana Zachar and DWAYNE C Savage. Microbial interference and colonization of the murine gastrointestinal tract by listeria monocytogenes. *Infection and immunity*, 23(1):168–174, 1979.

Aleksej Zelezniak, Sergej Andrejev, Olga Ponomarova, Daniel R Mende, Peer Bork, and Kiran Raosaheb Patil. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences*, 112(20):6449–6454, 2015.

Wenhan Zhu, Maria G. Winter, Luisella Spiga, Elizabeth R. Hughes, Rachael Chanin, Aditi Mulgaonkar, Jenelle Pennington, Michelle Maas, Cassie L. Behrendt, Jiwoong Kim, Xiankai Sun, Daniel P. Beiting, Lora V. Hooper, and Sebastian E. Winter. Xenosiderophore Utilization Promotes Bacteroides thetaiotaomicron Resilience during Colitis. *Cell Host & Microbe*, 27(3):376–388.e8, March 2020. ISSN 1931-3128. doi: 10.1016/j.chom.2020.01.010.

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

Emily M Zygiel and Elizabeth M Nolan. Transition metal sequestration by the host-defense protein calprotectin. *Annual review of biochemistry*, 87:621–643, 2018.

# Part III
# Appendices

# Appendix 8
# Details of results

Table 11: Abundance of B. thetaiotaomicron under SteadyCom in non-enhanced samples

| Constraints | Partner | Abundance |
|---|---|---|
| 33 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| No constraints | M_Bacteroides_uniformis_ATCC_8492 | 1.000000 |
| 125 uM H; 2 mM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 125 uM H | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 125 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 125 uM H; 125 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 0 uM Fe/H | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 2,7 uM H | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 125 uM H; 125 uM Fe | M_Escherichia_coli_ED1a | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Escherichia_coli_ED1a | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Escherichia_coli_ED1a | 0.000000 |
| 2,7 uM H | M_Escherichia_coli_ED1a | 0.000000 |
| 125 uM Fe | M_Escherichia_coli_ED1a | 0.000000 |
| 125 uM H | M_Escherichia_coli_ED1a | 0.000000 |
| 125 uM H; 2 mM Fe | M_Escherichia_coli_ED1a | 0.000000 |
| No constraints | M_Escherichia_coli_ED1a | 0.000000 |
| 33 uM Fe | M_Escherichia_coli_ED1a | 0.000000 |
| 0 uM Fe/H | M_Escherichia_coli_ED1a | 0.000000 |
| No constraints | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.991742 |

| | | |
|---|---|---|
| 125 uM H; 2 mM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 125 uM H | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 125 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 0 uM Fe/H | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 33 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 125 uM H; 125 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 2,7 uM H | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| No constraints | M_Roseburia_intestinalis_L1_82 | 0.017450 |
| 2,7 uM H | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 33 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 0 uM Fe/H | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 125 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 125 uM H | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 125 uM H; 2 mM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 125 uM H; 125 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 33 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 0 uM Fe/H | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| No constraints | M_Streptococcus_parasanguinis_ATCC_15912 | 0.576236 |
| 125 uM H; 125 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 125 uM H; 2 mM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 125 uM H | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 125 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 2,7 uM H | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 125 uM H; 2 mM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 125 uM H | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |

| | | |
|---|---|---|
| 125 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 2,7 uM H | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 33 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 0 uM Fe/H | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 125 uM H; 125 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| No constraints | M_Streptococcus_salivarius_DSM_20560 | 0.022845 |

Table 12: Abundance of B. thetaiotaomicron under SteadyCom in enzyme-enhanced samples

| Constraints | Partner | Abundance |
|---|---|---|
| 33 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| No constraints | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 125 uM H; 2 mM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 125 uM H | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 125 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 125 uM H; 125 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 0 uM Fe/H | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 2,7 uM H | M_Bacteroides_uniformis_ATCC_8492 | 0.000000 |
| 125 uM H; 125 uM Fe | M_Escherichia_coli_ED1a | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Escherichia_coli_ED1a | 0.000231 |
| 2,7 uM H; 33 uM Fe | M_Escherichia_coli_ED1a | 0.000893 |
| 2,7 uM H | M_Escherichia_coli_ED1a | 0.000000 |
| 125 uM Fe | M_Escherichia_coli_ED1a | 0.000000 |
| 125 uM H | M_Escherichia_coli_ED1a | 0.000300 |
| 125 uM H; 2 mM Fe | M_Escherichia_coli_ED1a | 0.000000 |
| No constraints | M_Escherichia_coli_ED1a | 0.000000 |
| 33 uM Fe | M_Escherichia_coli_ED1a | 0.000000 |

| | | |
|---|---|---|
| 0 uM Fe/H | M_Escherichia_coli_ED1a | 0.000000 |
| No constraints | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.314043 |
| 125 uM H; 2 mM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 125 uM H | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 125 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 0 uM Fe/H | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 33 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 125 uM H; 125 uM Fe | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| 2,7 uM H | M_Fusobacterium_nucleatum_subsp_nucleatum_ATCC... | 0.000000 |
| No constraints | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 2,7 uM H | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 33 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 0 uM Fe/H | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 125 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 125 uM H | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 125 uM H; 2 mM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 125 uM H; 125 uM Fe | M_Roseburia_intestinalis_L1_82 | 0.000000 |
| 33 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 0 uM Fe/H | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| No constraints | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 125 uM H; 125 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 125 uM H; 2 mM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 125 uM H | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 125 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 2,7 uM H | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Streptococcus_parasanguinis_ATCC_15912 | 0.000000 |

| | | |
|---|---|---|
| 125 uM H; 2 mM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 125 uM H | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 125 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 2,7 uM H | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 33 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 0,7 uM H; 8,2 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 0 uM Fe/H | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 125 uM H; 125 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| 2,7 uM H; 33 uM Fe | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |
| No constraints | M_Streptococcus_salivarius_DSM_20560 | 0.000000 |