

UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

CARRERA DE BIOINGENIERIA



**OPTIMIZING THE GENOMIC-ENZYMATIC MODEL OF
Yarrowia lipolytica FOR BIOFUEL PRODUCTION: A
CONSTRAINT-BASED RECONSTRUCTION ANALYSIS
APPROACH**

TRABAJO DE INVESTIGACIÓN

Para optar el grado de Ingeniero en Bioingeniería

AUTOR(ES)

Jeremy Rodrigo Guerrero Alejos

ASESOR(ES)

Paul Cardenas

Martin GUTIERREZ

Sabine PERES

Lima - Perú

2023

Dedicatoria:

Dedico este trabajo a toda mi familia. A mi papá y mi mamá, por su infinita paciencia conmigo; a mi hermana Astrid, quien ha estado a mi lado desde que tengo memoria; a mis hermanos menores, Nicolas y Salvador, que me han ayudado a crecer; a mi abuela Elisabeth, quien me enseñó cómo ser una buena persona; y a mi querida Francesca, quien me ha motivado cada vez que sentía que no podía continuar, me impulsó a seguir esforzándome y se quemó el cerebro junto a mí durante todo este proceso.

Agradecimientos:

Agradezco a todos mis profesores de la UTEC. Gracias a sus enseñanzas he logrado llegar hasta donde estoy ahora. Quisiera agradecer especialmente a mi profesora Sabine Peres y a Martin Gutierrez, quienes fueron piezas clave para el desarrollo de este proyecto. Cada reunión y consejo que me brindaron me acercaron cada vez más a la meta final de la investigación. Fue gracias a sus enseñanzas que mi mente y mi curiosidad se orientaron cada vez más hacia el producto final del proyecto. Merci beaucoup, Professeur Sabine. Muchas gracias, Profesor Martin.

TABLA DE CONTENIDO

	Pág.
RESUMEN	1
ABSTRACT	2
INTRODUCTION	3
Presentation of the research topic	3
Description of the problematic situation	3
Problem formulation	5
Investigation objectives	6
Justification	7
Scope and limitations	7
CAPÍTULO I CRITIAL REVIEW OF THE LITERATURE	8
CAPÍTULO II THEORETICAL FRAMEWORK	18
2.1 <i>Y. lipolytica</i> anatomy as unconventional yeast	18
2.2 Central Metabolism Pathways for Carbon	19
2.2.1 Embden-Meyerhof-Parnas Pathway	19
2.2.2 Krebbs Cycle	20
2.2.3 Propionate Pathway	21
2.3 Odd Chain Fatty Acid Metabolism in <i>Y. lipolytica</i>	22
2.3.1 Genomic Characteristics of <i>Y. lipolytica</i>	22
2.3.2 <i>Y. lipolytica</i> Respiration	23

2.3.3 Odd Chain Fatty Acids production optimization	24
2.4 Biotechnological Applications	25
2.4.1 <i>Y. lipolytica</i> as chassis	25
2.4.2 Industrial Applications of C15 and C17	27
2.4.3 Propionate as Carbon Source	28
2.5 Modelling an Metabolic Organism	30
2.5.1 Reconstruction of a Genome-Enzymatic Model	30
2.5.2 iYali2019 Model	31
2.6 Quasistationary Modelling of an Organism	32
2.6.1 LP feasibility problem	32
2.6.2 COBRApy	33
2.6.3 Dynamic Flux Balance Analysis	34
CAPÍTULO III METHODOLOGICAL FRAMEWORK	36
3.1 Description of the model	38
3.1.1 <i>Y. lipolytica</i> model comparison	39
3.1.2 Missing Propionate Metabolic Pathway	40
3.1.3 Missing OCFAs Metabolic Pathway	40
3.2 Model Development	42
3.2.1 Modeling of Metabolic Networks	42
3.2.2 Linear Programming	43
3.2.3 Gene-Protein-Reaction Rule	45
3.2.4 COBRApy	47
3.2.5 COBRApy needed to add metabolites, substrate and reactions	49
3.3 Flux Balance Analysis	50
3.3.1 Stationary Metabolic Network	50
3.3.2 Tools for FBA Analysis with COBRApy	51
3.4 dynamic Flux Balance Analysis	52
3.4.1 Michaelis-Menten Equation	53

3.4.2 Kinetics used for the dFBA model	54
3.4.3 Tools for dFBA Analysis	56
3.5 Performance Indicators	58
3.5.1 Contrasting Laboratory Data vs Modelled Data	58
3.5.2 Modelled Performance Indicators	58
3.6 Independent and Dependent Variables	60
3.7 Ethical Considerations, Limitations and Assumptions	60
CAPÍTULO IV RESULTS	65
4.1 New iYali Model	65
4.2 Analysis of the newYali model and Selection of Genes to be modified	66
4.3 Enhanced Phenotype	67
4.4 Flux Analysis of Created Phenotypes with FBA	69
4.5 Looking for key pathways on the optimization of C15 and C17 production	71
4.6 Predicted Production of of Highly Valuable Metabolites	73
4.7 Real Data vs Model Comparison with dFBA tools	74
4.7.1 Biomass Growth in Propanoate and Glucose Medium	74
4.7.2 Biomass Growth in Propanoate and Acetate Medium	78
4.8 Real Data vs Modelled Production of Valuable Compounds	81
4.9 Real Data vs Modelled Production of Lipids and OCFAs	82
CONCLUSIONES	85
RECOMENDACIONES	87
ANEXOS	103
4.10 Code to add pathways 1	103

ÍNDICE DE TABLAS

Tabla 3.4	List of different COBRApy functions	50
Tabla 3.5	List of different COBRApy functions for Flux Balance Analysis	52
Tabla 3.6	List of used kinetics for dFBA	56
Tabla 3.7	List of different dfba functions	57
Tabla 3.1	Summary of the characteristics of each model. The data was obtained partially from the work of Mishra et al. [83] and Guo et al. [81]. *The data from the iYali21 model was unavailable to access.	62
Tabla 3.2	List of Propionate Pathway Reactions Added into the iYali model . .	63
Tabla 3.3	List of OCFA Pathway Reactions Added into the iYali model	64
Tabla 4.1	Contrast between base model iYali4 and develop model new iYali. .	66
Tabla 4.2	List of Performance Indicators of a Predictive Model	78
Tabla 4.3	List of Performance Indicators of a Predictive Model	80

ÍNDICE DE FIGURAS

Figura 2.1 Morphology of <i>Y. lipolytica</i> . a) Differential interference contrast (DIC) of engineer strain (JMY3501) with increase lipid accumulation. b) <i>Y. lipolytica</i> colony stain of non-polar lipids with Bodipy® for fluorescence [1]	18
Figura 2.2 <i>Y. lipolytica</i> W29 genome in circular representation. The blue ring represents the forward strand, the red the reverse strand and the yellow and purple are the percentage of guanine and cytosine.	23
Figura 2.3 <i>Y. lipolytica</i> industrial characteristics and potential applications. .	27
Figura 3.1 Workflow of the Project. The red bubbles correspond to the Design step, the green to the Analysis step, and the blue to the Design step of the Design-Analysis-Built cycle in science.	37
Figura 3.2 Illustration of the Included Propionate Pathway.	41
Figura 3.3 Illustration of the Included OCFA Pathway.	42
Figura 3.4 Illustration of the abstraction process to construct a GEM from the genomics information of an organism to the build of an stoichiometric matrix. 1) It is necessary to have the information of an annotated genome of the organism. 2) Using this data, it is possible to recognise specific genetic circuits and gene encoding proteins. 3) From this circuits, it is possible to reconstruct the different enzymatic networks of the organism. 4) Finally, it is possible to obtain the stoichiometric matrix representation.	44

Figura 3.5 Illustration of the use of linear programming to compute the optimization of the Genome-scale Metabolic (GEM) model. A) The reconstructed metabolic network. B) This is then used for the construction of the stoichiometric matrix. C) Finally, our stoichiometric matrix can be represented as our unconstrained space that is then limited by our rules and optimization conditions.	45
Figura 3.6 Diverse set of gene-protein-reaction that are possible. A) Reaction A catalyzed by a single protein and a single gene. B) Reaction B that can be catalyzed by two isoenzymes. C) Reaction C produced by a protein complex. D) Redundant protein complex that is needed to carry out the reaction D.	46
Figura 3.7 Illustration of the algorithm structure of COBRApy.	47
Figura 3.8 Graph function of the speed of a metabolic reaction using the Michaelis-Menten equation. The x axis represent the concentration of the substrate, the y axis is the speed of the reaction, Vmax is the maximum speed of the reaction, and the Km is the Michaelis-Menten constant known as the constant affinity of an enzyme for its substrate.	53
Figura 4.1 Illustration of the <i>Y. lipolytica</i> GEM with the inclusion of the new pathways. The included metabolites and pathways are highlighted.	65
Figura 4.2 Graphic that illustrates the different growth rates in the model given each one gene knockout.	67
Figura 4.3 Graphic that illustrates the different growth rates in the model given each one gene knockout.	68
Figura 4.4 Flux production of different key metabolites in the different proposed models	70
Figura 4.5 Flux density of the new iYali model and the enhanced iYali model in 2 different scenarios. (a) Production optimizations of C15. (b) Production optimizations of C17.	71

Figura 4.6 Predicted production of Oleate, Oleonyl-CoA, Palmitoyl-CoA, Stearate, and Stearyl-CoA under different medium conditions. (a) Only Glucose, Glucose-Succinate, and Glucose-Xylose. (b) Glucose-Stearate, Glucose-Isobutanol, and Glucose-Isocitrate. (c) Glucose-Propionate, and Glucose-Pyruvate	75
Figura 4.7 Laboratory data vs Predicted results of the biomass growth of <i>Y. lipolytica</i> in glucose plus propionate gel medium.	76
Figura 4.8 Statistical graphs showing the performance of the model. (a) Scatter plot of Predicted vs. Real Values(b) Error Distribution Plot.	77
Figura 4.9 Laboratory data vs Predicted results of the biomass growth of <i>Y. lipolytica</i> in acetate plus propionate gel medium.	78
Figura 4.10 Statistical graphs showing the performance of the model. (a) Scatter plot of Predicted vs. Real Values (b) Error Distribution Plot.	80
Figura 4.11 Laboratory data vs Predicted results of new iYali and enhanced new iYalimodel for the production of valuable compounds: Acetyl-CoA, Malonyl-CoA, Methyl Malonyl, Propionyl-CoA, and Succinyl-CoA.	81
Figura 4.12 In-Vitro vs Predicted percentage production of Lipids and OCFA. (a) % of total lipids production vs final biomass. (b) % of total OCFA produced vs total lipids production. (c) % of different types of OCFA and ECFA production vs total lipids production	83

RESUMEN

OPTIMIZACIÓN DEL MODELO GENÓMICO-ENZIMÁTICO DE *Yarrowia lipolytica* PARA LA PRODUCCIÓN DE BIOCUMBUSTIBLES: UN ANÁLISIS BASADO EN RESTRICCIONES Y EN ALGORITMOS METAHEURÍSTICOS

Esta tesis tiene como objetivo abordar la necesidad de alternativas sostenibles a los métodos convencionales de producción industrial. La biotecnología industrial presenta una alternativa prometedora para lograr este objetivo, en particular mediante el uso de organismos vivos como base de nuevos procesos de producción. Sin embargo, el alto costo de la experimentación tradicional en laboratorio requiere el desarrollo de modelos computacionales para acelerar el progreso en este campo.

Esta investigación se centra en el modelado de *Yarrowia lipolytica* como chasis para la producción de ácidos grasos de cadena impar (OCFA) a escala industrial. Utilizando Flux Balance Analysis (FBA), exploraremos la red metabólica de esta levadura no convencional y desarrollaremos un algoritmo metaheurístico para optimizar su fenotipo con el objetivo de maximizar la producción de OCFA. El objetivo de este trabajo es avanzar en nuestra comprensión de *Y. lipolytica* y contribuir al desarrollo de procesos biotecnológicos sostenibles para la producción de compuestos de alto valor.

Palabras clave:

Yarrowia lipolytica; Ácidos Grasos de Cadena Impar; Análisis de Balance de Flujo; Biología de sistemas; Algoritmos Metaheurísticos

ABSTRACT

OPTIMIZING THE GENOMIC-ENZYMATIC MODEL OF *Yarrowia lipolytica* FOR BIOFUEL PRODUCTION: A CONSTRAINT-BASED RECONSTRUCTION ANALYSIS AND METAHEURISTIC ALGORITHMS-BASED APPROACH

This thesis aims to address the pressing need for sustainable alternatives to conventional methods of industrial production. Industrial biotechnology presents a promising avenue for achieving this goal, particularly through the use of living organisms in novel production processes. However, the high cost of traditional laboratory-based experimentation necessitates the development of computational models to accelerate progress in this field.

This research focuses on the modeling of *Yarrowia lipolytica* as a platform for producing Odd-Chain Fatty Acids (OCFAs) at an industrial scale. Using Flux Balance Analysis (FBA), we will explore the metabolic network of this non-conventional yeast and develop a new optimized phenotype for OCFA production. The goal of this work is to advance our understanding of *Y. lipolytica* and contribute to the development of sustainable biotechnological processes for the production of high-value compounds.

Keywords:

Yarrowia lipolytica; Odd-Chain Fatty Acids; Flux Balance Analysis; Systems Biology; Metaheuristic Algorithms

INTRODUCTION

Presentation of the research topic

The present thesis seeks to design a computational model for the in-silico model of the *Yarrowia lipolytica*, to optimize the production of two highly valuable Odd Chain Fatty Acids (OCFAs), specifically C15 and C17. These OCFAs are essential in various industries, such as biofuel, biomarkers, fuels, and plasticizers [1]. This thesis proposes the enhancement of the *Y. lipolytica*'s Genome-Enzymatic Model (GEM) through the implementation of a dynamic Flux Balance Analysis (dFBA) [2]. Through the integration of these methodologies, a diverse range of genetic modifications can be simulated, which in turn can result in the development of different phenotypes of *Y. lipolytica*. By doing so, it would be possible to significantly reduce the time and resources required for the identification of better strains in a laboratory environment. For this thesis, we will use concepts of bioinformatics, process of biological data, and systems biology.

Description of the problematic situation

In recent years, there has been a surge in carbon emissions following the end of lockdowns and the rapid economic recovery of countries such as China and the United States. According to Liu et al. [3], there has been an increase of 4.8% in carbon emissions over the last year. This is a concerning trend as the world continues to grapple with climate change resulting from greenhouse gas emissions. As a result, there is an urgent need to find new ways of producing highly valuable compounds in a sustainable manner. Industrial biotechnology, also known as white biotechnology, is a field of research that aims to apply biotechnological knowledge to utilize various living organisms and enzymes to produce new or existing products using renewable resources. This approach could result in biodegradable solutions with fewer pollutants

being released during their production process [4]. While some promising solutions have been proposed in this area, the process of introducing genetic modifications into an organism to optimize the overproduction of biochemical molecules remains slow and expensive. Especially the development of new phenotypes with new and improved characteristics. However, the process of selecting the appropriate combination of genes to knock out is often laborious and can take several months to complete. Developing computational models has emerged as a new research framework with a central role in investigating general biological functions and applications in biotechnology [2]. Genomic-Scale Metabolic Models (GEMs) are mathematical models that provide a comprehensive representation of an organism's metabolic reactions, enzymes and metabolites [5]. This enables the construction of simulations and predictions as changes in the phenotype of the base organism after specific genetic modifications. However, the construction of these models is complex and requires a new methodology based on stoichiometry and metabolic requirements, which can accurately order and represent each enzymatic reaction involved in the organism's growth under specific conditions [6]. While GEMs can analyze the metabolism of an organism under stable conditions, a dynamic Flux Balance Analysis (dFBA) is necessary to simulate and optimize individual reactions on the stoichiometric matrix over a specific time period. By changing the stoichiometric equation into a differential one it is possible to obtain the metabolic flux profile and the concentration of each metabolite at any time, depending on the specification of the model [7]. To optimize the entire system towards a specific flux, such as biomass production, Constraint-Based Reconstruction and Analysis (COBRA) is employed. This method performs a deterministic analysis of the GEM and optimizes its fluxes to obtain the best outcome of metabolite production with linear programming. While FBA or dFBA would typically result in multiple answers, COBRA obtains the most optimal solution through a series of constraints that limit the list of solutions that can satisfy the GEM equation [8]. One of the most commonly used tools to use

together with the FBA is the OptKnock algorithms. This package has gained prominence due to its utilization of mixed integer linear programming (MILP) to formulate a bi-level linear optimization problem. This approach is highly promising in finding the global optimal solution for selecting the combination of genes to knockout [2]. Despite its successes, the OptKnock algorithm has some limitations. One of the most significant drawbacks of this approach is that it does not account for the natural evolution of microbial genomes, leading to discrepancies when translating the results from the model to the laboratory environment. This highlights the need for novel algorithms that can better simulate the pathway optimization process without sacrificing the integrity of the organism. Therefore, it is imperative to continue researching and developing new computational models that can better capture the complex evolutionary dynamics of microbial genomes. With the ongoing advances in computational tools and molecular biology techniques, there is a great opportunity to revolutionize the field of synthetic biology and to contribute to the development of sustainable and efficient bioproduction systems.

Problem formulation

To address the pressing concern of global warming, it is imperative to fundamentally shift the paradigm of industrial production. With the advancements in laboratory techniques, the production of high-value compounds, such as biofuels and additives, has reached a point where it can potentially transition from the laboratory to industrial-scale production. However, the critical challenge in this transformation lies in the construction of optimized bacterial or yeast strains capable of producing the desired biomolecules at industrial scale. Promisingly, newly discovered organisms like *Y. lipolytica* offer potential as a robust chassis for industrial-scale production due to their expansive genomes and diverse metabolic pathways. Nonetheless, the greater

complexity of these organisms' genomes necessitates the development of novel approaches to effectively explore and construct superior strains that can meet the rigorous demands of industrial production.

Investigation objectives

The main objective of this research that will be:

- Optimize a refined phenotype of *Y. lipolytica*, yielding an enhanced production of Odd Chain Fatty Acids (OCFAs), particularly C15 and C17, facilitated by the manipulation of the enzymatic pathways within our model organism.

The specific objectives are:

- Update the iYali v4.1.2 metabolic model of *Y. lipolytica* to include any missing reactions, specially focusing on the propionate consumption and Fatty Acid synthesis pathway.
- Using COBRApy and FBA tools simulate different genetic modifications into the new *Y. lipolytica* model like overexpression and knockout of genes. In order to improve the production of OCFA without impacting the biomass production.
- Validate simulated results of the updated model of iYali through the contrast with in-silico laboratory data.
- Evaluate the capacity for production of biomass and OCFA through the use of alternative carbon sources of *Y. lipolytica*.
- Evaluate the enhanced capacity of the modified new iYali model for the production of highly valuable compounds without compromising the biomass production of the organism.

Justification

The outcome of this thesis will serve as the foundation for simulating the behavior of *Y. lipolytica* under standard laboratory conditions. These refined phenotype will facilitate the creation of multiple enhanced strains capable of achieving superior biomass yield and OCFAs production.

Scope and limitations

This research aims to enhance the current understanding of *Y. lipolytica* by improving the existing iYali v4.1.2 model and the dynamic Flux Balance Analysis (dFBA) algorithm. It should be noted that the focus of this thesis is not on the construction of *Y. lipolytica* strains in laboratory conditions. The ultimate goal is to contribute to the advancement of knowledge in the field of computational biology and provide insights into the dynamic behavior of *Y. lipolytica*.

CAPÍTULO I

CRITICAL REVIEW OF THE LITERATURE

This chapter provides a comprehensive review of prevalent techniques employed in the manipulation and analysis of Genome-Scale Metabolic Models (GEMs) [9] with a particular emphasis on models pertaining to *Y. lipolytica* [10]. Analytical methodologies such as Flux Balance Analysis (FBA) [8] and Dynamic Flux Balance Analysis (dFBA) [11] are examined, emphasizing their utilization in the exploration of novel phenotypic variations in bacteria and yeast for the production of biomass and other bio-compounds. Progressing further into advanced techniques, we delve into OptKnock [12] methodologies and the development of metaheuristic algorithms [13], both promising avenues for unearthing a complex array of solutions via the automated optimization of the stoichiometric matrix. Despite the consistent advancement of these tools, their contributions to the discovery of superior phenotypes remain limited due to their inadequate representation of the intricate processes through which nature evolves new pathways and organism strains.

The methodology employed in the composition of this review was systematic and thorough, involving a wide-ranging search across several scientifically approved databases including PubMed, Google Scholar, ResearchGate, ScienceDirect, Springer, and Elsevier. To effectively navigate these platforms and pinpoint pertinent papers on the topic, specific key terms were utilized, including "*Y. lipolytica*", "Genomic-Enzymatic Model", ".odd-Chain Fatty Acids", "Flux Balance Analysis", and "Dynamic Flux Balance Analysis". Additionally, temporal relevance was maintained by preferring the selection of papers published between the years 2007 and 2023. Studies that did

not feature these terms, were published prior to 2007, or did not directly address the research questions at hand were excluded from the review.

The selection process for relevant literature was meticulously implemented following a two-stage approach. Initially, the title and abstract of each identified paper were screened, which served to exclude studies of apparent irrelevance to the research subject. Subsequently, a comprehensive evaluation of the remaining papers was conducted, spanning from the introduction to the conclusion. This facilitated the extraction of key insights, which were then highlighted and summarized, enabling a comprehensive and in-depth understanding of the current advancements and limitations associated with the optimization of *Y. lipolytica* models for biofuel production. Over the course of one year, I diligently and personally executed this selection process throughout the duration of the thesis development. Ultimately, crucial metadata from each paper was extracted for inclusion in the thesis bibliography. This encompassed pertinent details such as the authors' names, year of publication, publishing entity, and abstract.

In the pivotal decade of the 1990s, pioneering work by Edwards and Bernhard [14] culminated in the first successful reconstruction of an organism's metabolic genome. Employing genomic data of *Haemophilus influenzae* Rd, initially identified by Fleischmann et al. [15], they assembled a comprehensive stoichiometric matrix, encapsulating 488 metabolic reactions and 343 metabolites. In silico experimentation with this novel model facilitated the postulation of six distinct phenotypes capable of growth under a cocktail of alternative carbon sources, such as fructose, glutamate, and xylose. Moreover, they systematically explored the effects of both single and double gene deletions, delineating potential growth rates associated with each configuration.

This groundbreaking study hypothesized that the construction of such mathematical models could pave the way for sophisticated in silico simulators, capable of

encapsulating intricate cellular functions based on GEMs [14]. This notion was explored previously by Trivedi's article of "Modeling the oddities of biology" [16], caused by the rising availability of genomic data through that decade. These formative ideas underpin contemporary algorithms and techniques designed to model comprehensive genomic pathways. In their work, Edwards and Bernhard adopted techniques such as Flux Balance Analysis and gene deletion methodologies which are now essential components of modern modeling.

Nonetheless, the preliminary nature of these early applications is evident, particularly in the limited scope of the model, when juxtaposed with present-day models such as Recon3D [17]. This representation of the human genome encompasses a staggering 16543 reactions and 4140 metabolites, epitomizing the significant strides taken in this field since its inception.

Shifting the focus towards our primary organism of interest, *Y. lipolytica*, the first GEM for this species was introduced in 2012 by Pan and Hua [18]. This model, a comprehensive construct featuring 1142 enzymatic reactions and 843 metabolites, was meticulously assembled using data retrieved from freely accessible databases such as KEGG [19] and BiGG [20] (refer Table 1). The Flux Balance Analysis was employed to validate the viability of the model, ensuring adherence to fundamental requisites such as maintaining an appropriate mass balance. Employing single-gene knockout techniques and modeling under different substrates, they were able to engineer novel phenotypes of *Y. lipolytica* to augment its production capabilities of citric acid and lipids.

The construction methodology of the GEM was notably resourceful, negating the need for any laboratory experimentation and instead leveraging the vast reserves of data available in public, peer-reviewed databases. This approach is increasingly favored as a means of capitalizing on the extensive amount of data, to the extent that automatic pipelines capable of scanning databases and assembling new, up-to-date

GEMs have been created, as demonstrated by Kerkhove et al. [21]. Despite the apparent efficiency of Pan and Hua's work, it must be noted that their approach lacked depth, evident in the incomplete model, which failed to account for certain cellular compartments and metabolic pathways, such as the peroxisome, liposomes compartments, and the carnitine pathway.

Later in 2012, Loira et al. unveiled an alternative GEM for *Y. lipolytica* [22]. Intriguingly, despite targeting the same organism, these models were developed independently and adopted strikingly distinct methodologies. For their model, Loira and colleagues adopted an iterative, comparative approach, utilizing the model of a closely related yeast, *S. cerevisiae*, as an initial scaffold. Given the longer history of study and more exhaustive model of this organism, it provided a robust foundation from which to build. Leveraging gene homology information from Génolevures protein families [23, 24], they proceeded to tailor this scaffold, adapting specific pathways, reactions, and molecular species. This process culminated in a more complete model, encompassing 2002 reactions and 1847 metabolites, encapsulating the roles of 895 identified genes at the time.

The manual curation of a new model, anchored in a peer-reviewed predecessor, certainly has its merits, principally because of its established foundation. However, this approach may inadvertently propagate misconceptions or errors into subsequent models. The main challenge lies in the fact that, while two organisms may belong to the same family, their genomes - and consequently their metabolic pathways - differ. These differences may manifest as incomplete pathway mapping in some areas, while others may exhibit unnecessary or inaccurate pathways. Hence, the customization of an existing model demands meticulous attention to detail and considerable expertise in discerning the necessary modifications to align the model with the target organism.

The model underpinning this thesis, the iYali4, is one of the most comprehensive GEMs of *Y. lipolytica* and was developed by Kerkhoven et al. [21]. This reconstruction

adopted a multidimensional approach, incorporating both open-access data and manual curation, further augmented by a series of in vitro experiments designed to probe the gene expression profiles of *Y. lipolytica* under controlled stress conditions. These trials, conducted in bioreactors, facilitated the extraction of RNAseq data previously explored by Morin et al. [25] in the same organism.

The introduction of transcriptional data was instrumental in elucidating numerous signaling networks implicated in the organism's response to varying nitrogen and carbon sources. While it is established that changes in protein transcription do not always correlate with alterations in metabolic fluxes, the observation of transcriptional shifts provides researchers with valuable insights into the activation of pathways under specific conditions. Consequently, it was feasible to further refine the model, culminating in a construct encompassing 1671 metabolites, 1925 reactions, and 849 distinct genes, with a comprehensive compartmental structure.

The methodology employed in the development of the iYali4 model is evidently superior to the approaches previously described. This model has incorporated pivotal pathways in biofuel production [26] including lipid synthesis in the endoplasmic reticulum and its membrane, alongside degradation in the peroxisome. Furthermore, it elucidated the role of nitrogen absorption during *Y. lipolytica*'s growth phase. However, the largest impediment to the completion of the *Y. lipolytica* GEM arises from the methodology itself. Stressing a strain to elicit specific transcriptional data equates to capturing a snapshot of the organism in a moment of time. Whilst this offers insights into the activation of specific pathways, it obscures the broader view of interactions between different compartments and pathways within the organism. Biological phenomena, being inherently dynamic, resist analysis through static investigations [27]. As we move forward, it is paramount to embrace methodologies that capture the temporal dimensions of metabolic processes.

The concept of Dynamic Flux Balance Analysis (dFBA) was first introduced by Mahadevan, Edwards, and Doyle in 2002 [7]. In their groundbreaking work, they advanced the conventional Flux Balance Analysis, incorporating a temporal dimension to better reflect the inherent dynamism of metabolic pathways. As established in the literature, enzymatic reactions within a metabolic pathway are not instantaneous, but rather occur over time [28]. Moreover, these pathways are interconnected, with metabolites produced in one reaction being simultaneously utilized in parallel reactions [29]. This interplay manifests in novel interactions, which would be impossible to see with traditional FBA due to its neglect of the temporal dimension. The dFBA methodology was first applied to simulate the diauxic growth of *E. coli* under batch culture condition

The advent of dFBA signaled a promising avenue for the design of novel organism phenotypes with greater precision regarding their metabolic networks. This expansion of FBA into the dynamic realm was an innovative means of understanding and predicting cellular behavior, particularly under varying environmental and internal conditions. This method facilitates the monitoring of metabolite consumption and production within specific time frames. Nonetheless, it was evident that further refinement of both the algorithm and GEM was necessary to fully realize the potential of dFBA as a predictive tool.

The utility of dFBA was further demonstrated in a study conducted by da Viega et al., who leveraged the iYali4 model [21] to simulate the growth of *Y. lipolytica*, employing both dFBA and Flux Balance Analysis tools [30]. The primary objective of this research was to ascertain specific culture conditions and gene knockouts that could optimize the citrate pathway.

Their analysis revealed that during the stationary phase of *Y. lipolytica*, an overflow mechanism could direct the acquired carbon source towards the production of citrate and lipid accumulation, as it is no longer required for biomass production. This

finding aligns with existing literature, which mentions that cells reach a saturation point for biomass production during the steady state, at which point they shift towards the production of secondary biomolecules or energy conservation via lipid bodies [31–33].

Accompanying in silico modeling efforts, da Viega et al. conducted parallel in vitro experiments with the wild-type *Y. lipolytica* W29 strain. Through these experiments, they were able to corroborate the hypothesis that inhibition of the AOX gene, encoding for the acyl-CoA oxidase protein, significantly augments the production of citric acid and lipids. This notion was initially proposed by Wang et al. in 1999, who not only proposed the link between AOX gene inhibition and enhanced production of citric and lipid acids, but also mentioned that such inhibition managed to increased production of short and long odd chain fatty acids [34].

Furthermore, during the experimental phase of da Viega's research, the opportunity arose to juxtapose model outputs with empirical data. This strategy permitted researchers to oversee the validation of their modeling hypotheses, a methodology currently in high demand across many scientific fields. However, this approach also restricts the quantity of potential single and double knockouts that the modeling might propose due to the substantial resources required for concurrent development of numerous *Y. lipolytica* phenotypes. This limitation necessitated the research focus on a specific pathway, thereby diminishing the array of potential viable phenotypes produced.

This issue highlights a recurring predicament within computational biology, wherein the integration of experimental and computational research often demands strategic resource allocation. As such, focusing on specific pathways or phenotypes

becomes an opposite approach to extracting meaningful insights from both computational predictions and experimental validations. This approach, however, necessitates a careful balance between exploring diverse possibilities suggested by computational models and the practical constraints of experimental validations.

Shifting towards more advanced methodologies, it is needed to include metaheuristic algorithms which are sophisticated computational tools designed to address complex optimization problems. The efficacy of these metaheuristics tools comes from proposing satisfactory solutions for intricate problems. This can be attributed to their adeptness in exploring a comprehensive search space of potential solutions.

A notable instance of a metaheuristic algorithm that functions in concert with a Genome-Scale Metabolic (GEM) model is the OptKnock algorithm. This algorithm made its academic debut in the 2003 study helmed by Professors Burgard, Pharkya, and Maranas [12]. In their publication, the authors explained the foundational framework of the OptKnock algorithm, which tactically balances the optimization of two primary variables: biomass and the production of a selected metabolite.

This novel methodology, encapsulating a bi-level optimization strategy, was designed to pinpoint genetic modifications that could promote the overproduction of a desired metabolite. While at the same time it ensured the preservation of cellular growth, thereby positioning itself as a crucial tool for metabolic engineering. These dual considerations elegantly underscore the complexity of metabolic networks and the importance of an all-encompassing approach in optimizing their functions.

OptKnock is not without its limitations, a product of its inherent mathematical abstraction programming and its foundation in Flux Balance Analysis (FBA). The algorithm's dependence on resolving the stoichiometric matrix purely as a mathematical equation can sometimes result in the suggestion of unfeasible phenotypes due to proposed gene knockouts. This can subsequently lead to unrealistic predictions of flux distribution, detracting from the model's accuracy. Moreover, the bi-level optimization

approach with a pronounced focus on the inner level can create a conflict between the two levels. This could potentially yield suggestions for gene knockouts that contradict its intended purpose and fail to accurately mirror the *in vivo* effects of the same gene deletion conditions [35].

In an exemplary application of the OptKnock algorithm, Czjaka et al. integrated it into the GEM of *Y. lipolytica* to optimize the bio-production of organic acids [36]. The researchers meticulously collated data concerning the cultivation conditions, genetic engineering strategies, and production details of *Y. lipolytica*, integrating this information into the GEM model. Subsequently, a machine learning model was trained to utilize this production information for various Flux Balance Analysis operations, thereby predicting strain production titers. This knowledge was then juxtaposed with the OptKnock-derived gene knockout predictions, enabling a more rigorous screening of potential phenotypes with enhanced bio-production capacities.

This quite innovative approach to predict and develop new *Y. lipolytica* phenotypes certainly places a great value in the validation of its prediction through different algorithms. However, one must recognize the fact that the accuracy of the predicted fluxes generated by the machine learning algorithm and Optknock is inherently dependent on the quality of the data used to construct the GEM. One manifestation of this limitation was evident when the biomass growth rate for the FBA simulations only achieved 75 % of the predicted optimal value. It's plausible that the implementation of gene knockouts in the modeled strain led to slower growth due to the metabolic burdens stemming from overexpressed pathways [37].

In conclusion, while remarkable strides have been made in the field of metabolic engineering using GEMs, it is evident that the quest for computational precision and biological fidelity is a continuously evolving journey. The future direction of this field

lies in the seamless integration of sophisticated algorithms with high-quality experimental data to elucidate and harness the complexity of metabolic networks, ultimately paving the way for optimized and novel metabolic engineering strategies.

CAPÍTULO II

THEORETICAL FRAMEWORK

2.1 *Y. lipolytica* anatomy as unconventional yeast

Cells, the quintessential units of life, form the foundation of all known living organisms, irrespective of their multi-cellular or unicellular nature. Depending on the taxonomic domain and kingdom the organism can have some unique differences. These different features can appear in the inherent variations in genetic material handling (RNA vs DNA) and energy production methodologies (Phototrophy vs Fermentation) [38]. Examining the *Y. lipolytica*'s cellular anatomy reveals distinguishing characteristics, setting it apart from other organisms and even fellow yeasts. These differences primarily manifest in the yeast's outer membrane, core genetic constituents, and its notably diverse metabolic activities [39]. In the Figure 2.1 we can see an microscopy image of *Y. lipolytica*.

All cells can be analogized to intricate biochemical factories, due to their capacity to simultaneously execute an array of metabolic processes. These include, but

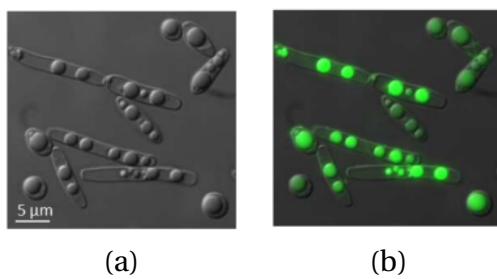


FIGURA 2.1: Morphology of *Y. lipolytica*. a) Differential interference contrast (DIC) of engineer strain (JMY3501) with increase lipid accumulation. b) *Y. lipolytica* colony stain of non-polar lipids with Bodipy® for fluorescence [1]

are not limited to, the biosynthesis of essential molecules such as lipids, ATP energy production, and biomolecular transport throughout the cellular structure [40]. These metabolic activities are indispensable for cellular growth, reproduction, and the preservation of cellular functions. *Y. lipolytica*, in particular, is renowned for its ability to utilize a broad spectrum of compounds as substrates, including alkanes, fats, and oils.

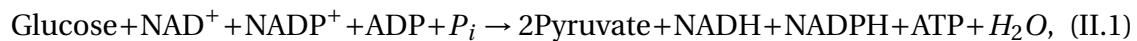
Further, the cell membrane, one of the most complex and crucial components of a cell, serves as a barricade between the cell's interior and the external environment. Its composition varies, consisting of a diverse range of lipids and proteins, contingent on the organism's specific environmental adaptations. *Y. lipolytica*'s cell membrane is characterized by a reduced beta-glucan content and an elevated presence of manno-proteins [41]. In addition, it possesses a cell wall that bestows protection and rigidity. Lastly, metabolic respiration, a vital cellular process, entails the transformation of a sugar source (typically glucose) into adenosine triphosphate (ATP), the primary source of chemical energy for cellular function. As previously stated, *Y. lipolytica* is capable of generating biochemical energy from a wide array of substrates. It is also reported to amass considerable quantities of lipids utilized for energy storage. *Y. lipolytica* performs both aerobic and anaerobic respiration, akin to other yeasts. Under aerobic conditions, it predominantly activates the citric acid pathway, whereas under anaerobic conditions, it produces ethanol, mirroring the metabolic function of *S. cerevisiae* [42].

2.2 Central Metabolism Pathways for Carbon

2.2.1 Embden-Meyerhof-Parnas Pathway

The Embden-Meyerhof (EM) pathway represented in Equation II.1, serves as a pivotal route for glucose assimilation and biochemical energy production in yeast species. Engaging up to ten different enzymes, this pathway facilitates the decomposition and fermentation of glucose into ethanol [43]. Operating under both aerobic and

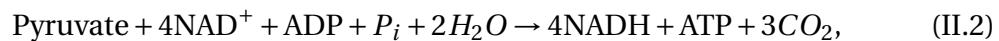
anaerobic conditions, the EM pathway bifurcates into distinct fermentation pathways, yielding two molecules of NADH, two pyruvates, and four ATP molecules, thereby substantiating its role as an energy-dense production mechanism [44].



Beyond the indispensable role in yeast cell sustenance, the EM pathway also demonstrates significant potential within the biotechnological sector. Research conducted by Yang et al. reveals that biodiesel waste can contain up to 80% glycerol, a byproduct with significant potential for repurposing [45]. One such application is its utilization as a carbon source for the production of valuable chemicals such as ethanol, succinate, and propanediol, achieved through the modification of the EM pathway [46]. This underscores the potential of the EM pathway not just for understanding fundamental biological processes, but also for advancing sustainable industrial applications.

2.2.2 Krebs Cycle

The importance of the Krebs cycle can not be underestimated, it is a cornerstone of cellular respiration that is pivotal in the production of both energy and biomass. This process is continually at work in the mitochondria of eukaryotic cells [47].



The Krebs cycle, simplify in Equation II.2, is an elegantly orchestrated series of chemical reactions that operate in a cyclic manner. The catalyst of this cycle is Acetyl-CoA, a by-product of pyruvic acid, which acts as the primary substrate[48]. Through this cycle, the acetyl-coa, combines with a four-carbon molecule called oxaloacetate to produced citrate [49].This citrate molecule then follows a sequence of enzyme-catalyzed transformations, ultimately regenerating oxaloacetate, thereby priming the cycle for another round [50]. This pathway results in the release of CO₂ and culminates in the generation of energy-rich molecules – ATP, NADH, GTP, and FADH₃. The krebs cycle has a central role in all organisms, including yeasts like *Y. lipolytica*. Its importance not only radicates in the production of energy, but also several precursors for many biosynthetic pathways.

2.2.3 Propionate Pathway

The role and potential of propionate as a metabolic intermediate in eukaryotes is very promising. Propionate is a short-chain fatty acid that is mostly found metabolized in the fermentation process of certain bacteria. It has a simple 3-carbon structure and has been observed as an intermediate metabolite in numerous pathways, notably the Krebs cycle [1].

It can be used in the mitochondria in the form of propionyl-coa, in a reaction catalyzed by propionyl-CoA synthetase. Propionyl-CoA is then carboxylated to D-methylmalonyl-CoA by propionyl-CoA carboxylase, a biotin-dependent enzyme. D-methylmalonyl-CoA is then isomerized to L-methylmalonyl-CoA by methylmalonyl-CoA racemase, and finally, L-methylmalonyl-CoA is converted to succinyl-CoA, a TCA cycle intermediate [51].

The implications of propionate metabolism reach further than energy production. Notably, Park et al. highlighted the potential of propionate as a precursor pool

for the biosynthesis of OCFAs in metabolically engineered strains of *Y. lipolytica*. The authors suggest that the odd number of carbon atoms in propionate's structure predisposes it as a suitable substrate for the synthesis of other odd-chain fatty acids, owing to the preferential stacking nature of this process [52].

As with any complex biological system, further research will be needed to elucidate the full extent and implications of propionate metabolism within eukaryotic organisms.

2.3 Odd Chain Fatty Acid Metabolism in *Y. lipolytica*

2.3.1 Genomic Characteristics of *Y. lipolytica*

Y. lipolytica, a eukaryotic unicellular organism belonging to the hemiascomycetes class and dipodascaceae family, possesses a genome comprising 7357 genes and 6472 coded proteins. The organism is characterized by 6 chromosomes, named from Yali A to Yali F, with sizes varying between 2.3 to 4.2 Mb [9, 53]. *Y. lipolytica* is commonly referred to as a "non-conventional yeast" due to the marked differences in its genome compared to the extensively researched *S. cerevisiae*.

The genome represented in the Figure 2.2 is the W29 strain [54]. The blue ring represents the forward strand, the red the reverse strand and the yellow and purple are the percentages of guanine and cytosine. Notably, *Y. lipolytica* is considered a promising candidate for industrial applications due to its ability to tolerate certain by-products and inhibitors [55] and its classification as a non-pathogenic organism by the American Food and Drug Administration (FDA, USA). Moreover, *Y. lipolytica* exhibits the capability to metabolize a diverse array of compounds, utilizing various carbon substrates such as glucose, acetate, propionate, and alcohols, among others [10].

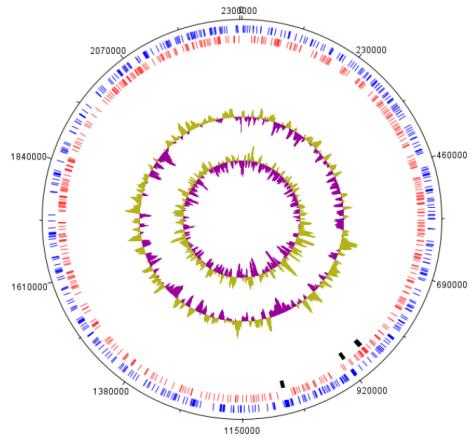


FIGURA 2.2: *Y. lipolytica* W29 genome in circular representation. The blue ring represents the forward strand, the red the reverse strand and the yellow and purple are the percentage of guanine and cytosine.

2.3.2 *Y. lipolytica* Respiration

Y. lipolytica, an eukaryotic yeast, is strictly aerobic and demonstrates respiration patterns strikingly analogous to those found in animal cells. Principally utilizing glucose as its primary carbon source, it employs complex metabolic pathways such as the Embden-Meyerhof (EM) and propionate pathways to catabolize this resource, liberating glucose for further biochemical reactions [56]. This glucose is subsequently converted into pyruvate via the glycolytic process occurring within the cytosol, which, in turn, is translocated into the mitochondria where it is further metabolized in the citric acid cycle.

The electron transport chain within the mitochondria of *Y. lipolytica* resembles that of multicellular organisms, including plants and animals. Comprising four major protein complexes (Complexes I to IV), the respiratory chain operates to facilitate the electron transport processes integral to cellular respiration [57]. Notably, a distinct feature of *Y. lipolytica*'s mitochondrial metabolism lies in the presence of two additional alternative oxidoreductases, specifically, an external NADH dehydrogenase (NDH2e)

and an alternative oxidase (AOX), which together, function as a quasi fifth complex (Complex V) [31].

It is imperative to highlight that the participation of Complex V within the respiratory chain is contingent upon environmental and growth conditions, and has been demonstrated to possess the capacity to interact with the other respiratory complexes. The electrons channeled through the NDH2e have been specifically observed to be directed towards the cytochrome complexes, thereby, potentially augmenting the efficiency and adaptability of *Y. lipolytica*'s metabolic machinery in response to variable growth conditions and external stressors [31].

This contextual dynamism of Complex V within the electron transport chain underscores *Y. lipolytica*'s adaptive metabolic prowess and presents a fascinating area of exploration in the understanding of the eukaryotic mitochondrial respiration and its adaptations. The theoretical and applied implications of these insights may hold profound influence on our capacity to manipulate and harness the bio-energetic processes of *Y. lipolytica* for various biotechnological applications.

2.3.3 Odd Chain Fatty Acids production optimization

The oleaginous yeast *Y. lipolytica* has emerged as a potent platform for the biosynthesis of odd chain fatty acids (OCFAs). Its robust potential for this role is underpinned by two primary factors: its innate biological attributes, and recent strides in strain development.

As an oleaginous organism, *Y. lipolytica* is naturally equipped to produce and store vast quantities of unconventional lipids, an ability analogous to the lipid storage capacities of animal adipose tissue. The biosynthesis of these lipid structures, housed within liposomes, are directly linked to the expression of the AOX2P gene [58]. Previous

studies by Mlickova et al. have demonstrated the feasibility of bolstering liposome production and size via AOX2P overexpression. Moreover, the implementation of strategies that inhibits the beta-oxidation catalytic pathway, a metabolic route located within the mitochondria and peroxisomes of *Y. lipolytica*, has been shown to amplify fatty acid production [59]. By inhibiting this pathway, we ensure the availability of precursors for OCFA synthesis.

Parallel to this, strain development methodologies for *Y. lipolytica* have evolved at a rapid pace. Advancements in synthetic biology tools, including the creation of overexpression cassettes and the application of CRISPR gene knockout techniques, have enabled the development of several *Y. lipolytica* strains that can accumulate high levels of OCFAs [60]. These tools, coupled with novel DNA assembly techniques, the creation of a comprehensive library of DNA parts for expression cassettes, and the advent of innovative computational tools, have substantially enhanced the capacity to manipulate the *Y. lipolytica* genome [61].

All these factors have come together to form an optimal environment for the amplification of biofuel production and other high-value compounds using *Y. lipolytica*. With its inherent biological capabilities and recent progress in strain development, this yeast has emerged as a promising avenue in the exploration and exploitation of OCFA production.

2.4 Biotechnological Applications

2.4.1 *Y. lipolytica* as chassis

The utility of *Y. lipolytica* as a powerful tool in biotechnology has been raising in recent years, with researchers exploring its capacities to produce a broad spectrum of valuable compounds. This non-conventional yeast has emerged as a versatile chassis

for synthetic biology and metabolic engineering endeavors, demonstrating both robustness and metabolic versatility as shown by Larroude et al [61]. The goal of utilizing this organism as a workhorse for industrial production is becoming more and more clear with each passing year.

First, *Y. lipolytica* has been shown to be an effective platform for the synthesis of biopolymers, as demonstrated by Lajus et al. [62]. The inherent metabolic capabilities of this organism, coupled with the relative ease of genetic manipulation, make it an ideal candidate for the biosynthesis of these materials. Biopolymers, which are increasingly seen as a sustainable alternative to petroleum-based polymers, can be produced through the fermentation of renewable feedstocks, thus aligning with goals of environmental sustainability and circular economy.

Secondly, the ability of *Y. lipolytica* to produce biofuels, such as biodiesel, underscores its potential in renewable energy production. Darvishi et al., showcased the feasibility of utilizing *Y. lipolytica* for biofuel production, and this represents just one aspect of its potential in the energy sector [9]. This was demonstrated by the obese strain produced by Park et al., capable of producing several types of OCFAs like pentadecanoic acid (C15), and Heptadecanoic acid (C17) [60]. The development of this strain capable of converting a variety of substrates into biofuels not only provides an alternative to fossil fuels but also presents an opportunity to valorize waste streams, further promoting sustainability.

The points recently mentioned are summarized in Figure 2.3. The employment of *Y. lipolytica* as an industrial workhorse is gaining momentum. By leveraging its unique metabolic characteristics and amenability to genetic manipulation, *Y. lipolytica* has the potential to revolutionize various sectors, from materials science to energy production, thereby redefining what is achievable through biotechnological innovation.

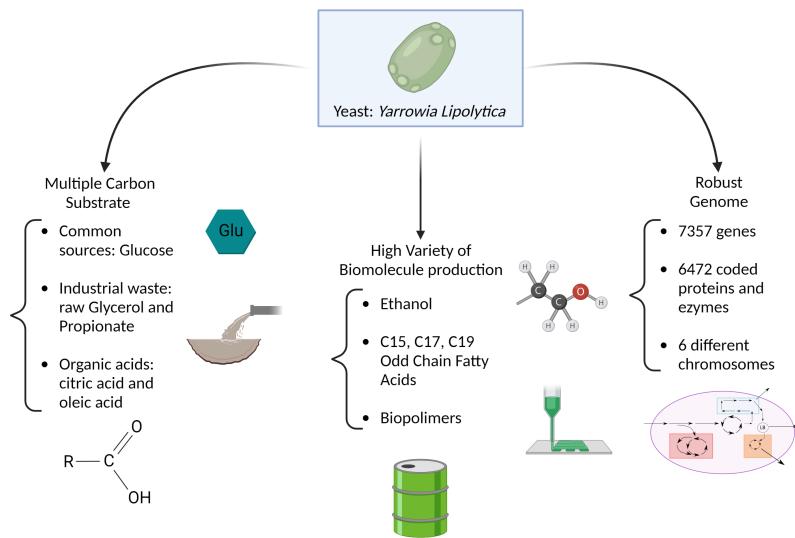


FIGURA 2.3: *Y. lipolytica* industrial characteristics and potential applications.

2.4.2 Industrial Applications of C15 and C17

Drawing on the latest research, the conceptualization of pentadecanoic acid, heptadecanoic acid, and nonadecanoic acid as key components in innovative bio-industrial applications provides a paradigm shift in our understanding of resource utility. These substances, which consist of 15, 17, and 19 carbons respectively, represent the sub-category of odd-chain fatty acids (OCFAs) and play pivotal roles in various biological systems [63–65].

Pentadecanoic acid, and heptadecanoic acid are metabolic byproducts of various organisms, including bacteria, plants, and oleaginous yeasts like *Yarrowia lipolytica* [66]. They are also present in mammalian systems, as elements of blood serum and adipose tissue.

As we dissect the biological and chemical properties of these OCFAs, we begin to appreciate their prospective applications in a multitude of industries. For instance,

the cis-9-heptadecenoic acid exhibits anti-inflammatory properties, suggesting its potential utility in the treatment of allergies and autoimmune diseases [67]. Expanding upon this therapeutic lens, pentadecanoic acid may serve as a bioindicator for food intake, aiding in the management of coronary heart disease and type 2 diabetes mellitus [68].

Heptadecanoic acid, or C17, however, illuminates an alternative pathway to sustainable energy production. Its potential as a third-generation biofuel eliminates the conventional conflicts associated with first- and second-generation biofuels - such as the necessity for arable land and competition with the demand for edible feedstocks [69]. Its primary source, *Yarrowia lipolytica*, facilitates this potential while offering an environmentally friendly and economically viable alternative to fossil fuels.

This theoretical framework not only emphasizes the transformative potential of OCFAs in disparate industries but also underscores the necessity for a multifaceted approach that incorporates biology, chemistry, medicine, and energy production. It reinforces the concept that waste products, once overlooked, may hold the key to innovation in sustainability, health, and bio-industrial applications.

2.4.3 Propionate as Carbon Source

Nowadays, more and more research is beginning to focus on the re-use of waste products as novel resources. This strategy, inspired by the principles of a circular economy, primarily seeks to mitigate the ecological footprint of industrial activities and reduce resource consumption. This line of investigation converges into a domain that can be classified as waste valorization, a process that primarily entails the transformation of waste streams into valuable bioproducts [70].

Among the plethora of waste products, volatile fatty acids (VFAs) are gaining recognition for their potential as precursors in various biotechnological processes. VFAs

are predominantly short-chain fatty acids that primarily originate from anaerobic digestion processes. Their production is largely facilitated through bacterial fermentation of a wide range of biodegradable organic wastes, including food wastes and sludge, positioning them as a low-cost alternative feedstock for several bioprocesses [71].

The potential of VFAs as platform chemicals is increasingly being acknowledged in the realm of industrial biotechnology. Acetic acid, for instance, is a prominent example of a VFA that is widely utilized in food industries as an additive and also in polymer manufacturing [72]. This creates a novel paradigm in biomanufacturing, where waste streams are not merely viewed as disposal liabilities, but as economically and ecologically valuable inputs for the production of diverse bioproducts.

Another VFA that is highly promising is the propionate. Propionate, as described in a previous chapter, is denoted by the molecular formula $C_3H_5O_2$ and serves as a promising volatile fatty acid (VFA) capable of further applications in diverse fields such as biofuels and biotechnology. Within biological systems and at standard biological pH, it primarily exists in its conjugate base form, $C_2H_5COO^-$, as reviewed by Ingram et al. [73].

An intriguing aspect of propionate is its prospective role as a precursor for the synthesis of odd-chain fatty acids (OCFAs). Ingram et al. elucidated the possibility of utilizing exogenous propionate as a primer for OCFA synthesis, presenting an opportunity for the development of biofuel alternatives [73]. In the wake of a growing emphasis on renewable energy sources and sustainable practices, the production of OCFAs from propionate aligns well with these global objectives.

A burgeoning area of investigation has been the exploration of microbial metabolic pathways to enhance propionate utilization and subsequent OCFA production. Notably, attempts have been made to engineer bacterial strains, such as *Escherichia*

coli, and oleaginous yeasts, such as *Y. lipolytica*, to augment OCFAs biosynthesis. A central strategy has been to attenuate propionyl-CoA catabolism through the methyl citrate pathway, thereby preserving the availability of propionate for the formation of OCFAs [73, 74].

While the field is new and the technology in its infancy, the potential environmental advantages offered by leveraging VFAs like propionate to produce valuable bio-based compounds are evident. The exploration of this intersection between waste valorization and advanced biotechnology is anticipated to yield substantial benefits, propelling us towards a more sustainable and resource-efficient future.

2.5 Modelling an Metabolic Organism

2.5.1 Reconstruction of a Genome-Enzymatic Model

The inception of Genome-scale metabolic models (GEMs) marks a pivotal moment in the development of computational biology. Originating from the pioneering work of Edwards et al. in 1999 [14], GEMs have since become an indispensable tool in the exploration and elucidation of an organism's metabolic capabilities. The groundwork laid in the analysis of *Haemophilus influenzae Rd*'s genome has not only served to illuminate the intricate network of metabolic pathways at work within this bacterium but has also set the stage for the development of six distinct optimal metabolic phenotypes. This seminal study thus served to highlight the potential of GEMs as a means of understanding complex biological systems.

In essence, a GEM can be viewed as a compendium of the stoichiometric relationships that underpin each biochemical reaction within an organism [47]. These relationships are organized into a mathematical framework - a stoichiometric matrix - that can encapsulate the instantaneous metabolic state of an organism. GEMs are

subsequently used to study the flux behavior within a metabolic network, providing a means of assessing how changes in the availability of certain metabolites can propagate through the system.

However, despite the utility of GEMs, their limitations must also be acknowledged. They operate under a deterministic framework that disregards the inherent stochasticity of biological systems. As such, GEMs are unable to account for random occurrences or perturbations within the organism or cell. While they provide a robust and systematic representation of metabolic networks, this rigidity might not fully capture the complexity and dynamism inherent in biological systems.

Despite these constraints, GEMs remain a powerful tool for understanding and manipulating biological systems. As we continue to refine these models and incorporate new layers of biological complexity, we can expect them to play an increasingly central role in the development of novel therapeutic strategies and the optimization of biotechnological processes.

2.5.2 iYali2019 Model

The development of the iYali model, a Genome-scale Metabolic Model (GEM), for *Yarrowia lipolytica* W29, a wild-type strain, represents a significant advancement in our understanding of the metabolic capabilities of this yeast. This model was created by Kerkhoven et al. from the Chalmers University of Technology in Sweden, the iYali model is an outcome of an innovative combination of an existing GEM and multilevel omics data [21].

The foundation for the iYali model was derived from the Yeast8 GEM, which describes the metabolic framework of *Saccharomyces cerevisiae* [75]. The choice of Yeast8 as a blueprint was justified by the biological relatedness of *Y. lipolytica* and *S. cerevisiae*, as they both belong to the *Saccharomycetales* order and share numerous metabolic

pathways. However, the iYali model is not a mere adaptation of Yeast8 to *Y. lipolytica*, but a distinct GEM that accurately portrays the unique metabolic capabilities of this organism.

The completion of the iYali model necessitated the integration of multilevel omics data to capture and incorporate the missing metabolic pathways from the Yeast8 model. This integrative approach allowed for the comprehensive description of *Y. lipolytica*'s metabolic network and illustrated the utility of using existing GEMs as a starting point for the construction of novel models.

2.6 Quasistationary Modelling of an Organism

2.6.1 LP feasibility problem

The theoretical underpinning of Flux Balance Analysis (FBA) is rooted in the mathematical framework of linear programming (LP), a problem-solving algorithm conceived in the 1940s by renowned mathematicians such as G. Dantzig, J. von Neuman, and L. V Kantorovich, as a response to the challenges of optimal resource allocation during wartime [76].

In the context of Genome-scale enzymatic models (GEMs), FBA leverages the power of LP to navigate the intricate maze of metabolic reactions. The objective of FBA is to find the optimal distribution of reaction fluxes that maximizes a predefined objective function, such as biomass or ATP production, subject to a set of constraints that encapsulate the biochemical and physicochemical properties of the cell [8].

To conduct FBA, the first step is to compile the stoichiometric matrix, which is a comprehensive collection of all known metabolic reactions in the organism. Each reaction, which is represented as a linear equation, contributes a row to the matrix, with the reactants and products serving as variables. However, owing to the inherent

complexity of metabolic networks, the resulting system is typically underdetermined, with more variables than equations, leading to an infinite space of potential solutions.

To navigate this high-dimensional solution space and identify physiologically relevant flux distributions, FBA introduces additional constraints. These constraints often stem from the assumption of steady-state operation, which posits that the net production of each metabolite is zero, as the rate of production exactly matches the rate of consumption. Additional constraints may be imposed based on experimental measurements or known upper and lower bounds on reaction fluxes, thereby further refining the solution space.

Remarkably, the LP problem can be solved efficiently without any requirement for kinetic parameters, thereby making FBA a computationally tractable method for exploring the metabolic capabilities of an organism. Furthermore, FBA supports the modeling of genetic perturbations, such as gene deletions, by setting the flux of the corresponding reactions to zero.

The FBA methodology provides a robust and scalable framework for interrogating genome-scale metabolic networks. It leverages the mathematical elegance of LP to identify optimal reaction fluxes, thereby opening up a window into the metabolic capabilities and constraints of a cell. As we continue to refine this approach, FBA holds great promise in enhancing our ability to predict cellular phenotypes from genotypes, thereby advancing our understanding of the principles underlying cellular life.

2.6.2 COBRAp

The COBRA (COnstraint-Based Reconstruction and Analysis) toolbox has emerged as a powerful resource for the integrative analysis of genome-scale metabolic models, with a particular focus on quantitative prediction of the stoichiometric fluxes associated with an organism's phenotypic states [77]. Constructed primarily for use with

MATLAB [78], the COBRA toolbox incorporates a plethora of algorithms and software packages, demonstrating its versatility across diverse fields such as systems biology, biomedicine, and bioprocess engineering.

The unique strength of the COBRA toolbox lies in its mathematical approach to analyzing biochemical networks, which offers the opportunity for detailed, quantitative exploration of metabolic behaviors. Such capabilities are invaluable for uncovering the metabolic constraints influencing biological systems and their resultant phenotypes.

Among the various components of the COBRA toolbox, COBRAPy stands out as a collaborative effort from the Open COBRA community to establish a cross-lingual code integration. Specifically designed for compatibility with Python, COBRAPy facilitates the construction and analysis of GEMs within the Python programming environment [79]. This merger not only broadens the accessibility of the toolbox but also enhances its functionality, enabling researchers to leverage the computational power of Python in the realm of metabolic network analysis.

In summary, the COBRA toolbox, particularly its COBRAPy component, represents a significant contribution to the field of systems biology. Its broad application scope and its ability to provide quantitative insights into metabolic networks serve to enhance our understanding of biological systems at a genomic scale. As we continue to develop and refine such tools, we move closer to fully realizing the potential of metabolic modeling in guiding experimental design and interpretation within a diverse range of biological disciplines.

2.6.3 Dynamic Flux Balance Analysis

Dynamic Flux Balance Analysis (dFBA) is an enhancement of the Flux Balance Analysis (FBA) methodology, extending its capabilities to capture dynamic metabolic

behaviors over time. FBA is characterized by its emphasis on analyzing the metabolic flux distribution at a steady state, thereby yielding a snapshot of the system's metabolic activities at a given moment. In contrast, dFBA incorporates time into its framework, offering a continuous panorama of the metabolic landscape [80].

One of the remarkable features of dFBA is its capacity to model the flux of metabolites between different cellular compartments, and between the cell and the external environment. This allows for the analysis of complex systems that are otherwise difficult to assess, extending the scope of the model beyond the intracellular space.

Incorporating a temporal dimension into the optimization framework opens a myriad of new possibilities for modeling intricate metabolic networks spanning multiple organisms simultaneously. In an application of this, Timothy et al. successfully modeled a co-culture of two engineered *E. coli* strains, ALS1008 and ZSC113, each optimized for the uptake of glucose and xylose respectively [11]. The inclusion of time in the model enabled them to predict the dynamic growth rate and productivity over the course of the experiment, offering insights into the synergistic interactions of the two strains.

Furthermore, the capability to interact with an extracellular space allows the modeling of microbial communities within bioreactors. For instance, Radhakrishnan et al. proposed a growth model for a batch of *E. coli* consuming glucose within a bioreactor, successfully reproducing experimental observations qualitatively [7].

In essence, the dFBA methodology offers a robust theoretical framework to simulate, analyze, and predict dynamic metabolic behaviors, marking a significant leap forward in our ability to understand and manipulate complex metabolic networks. This approach, combined with experimental validation, can potentially steer the field towards the realization of optimized microbial communities for diverse biotechnological applications.

CAPÍTULO III

METHODOLOGICAL FRAMEWORK

The main objective of this chapter is to illustrate the precise methodology adopted throughout the entirety of this thesis, along with a thorough delineation of the computational tools employed to realize the initial research objectives. Given to the computational nature of this work, this chapter delves into the complexities involved in creating, analyzing, and optimizing a Genomic-Enzymatic model of *Y. lipolytica*. The ultimate aim of the research is to engineer a refined phenotype of *Y. lipolytica*, yielding an enhanced production of Odd Chain Fatty Acids (OCFAs), particularly C15 and C17, facilitated by the manipulation of the enzymatic pathways within our model organism.

The structure of this chapter will follow the chronological progression of the steps taken previously to construct the novel phenotype. The beginning of the project was rooted in a comprehensive examination of the existing models of *Y. lipolytica*, facilitated by Python and the COBRA library, coupled with the incorporation of essential pathways for OCFA biosynthesis. The subsequent phase was the analysis of the emergent model using Flux Balance Analysis (FBA) and Dynamic Flux Balance Analysis, key in gaining insights into flux density and informing decisions regarding requisite pathway alterations.

A pivotal aspect of this work was the development of a Genetic Algorithm (GA), designed to substantially enhance the optimization process. This evolutionary algorithm served the purpose of discovering optimized versions of the stoichiometric matrix of our model, thereby elucidating new strategies for OCFA production while optimizing biomass production. The integration of these three computational tools formed a cohesive workflow to achieve the research objectives which can be seen in Fig. 3.1.

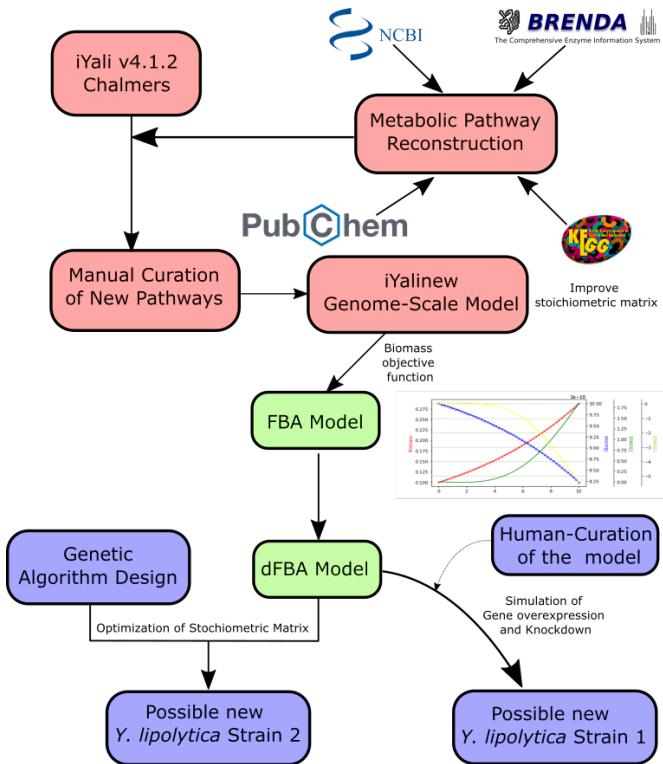


FIGURA 3.1: Workflow of the Project. The red bubbles correspond to the Design step, the green to the Analysis step, and the blue to the Design step of the Design-Analysis-Built cycle in science.

In addition, this chapter will also address the subjects of model validation, research limitations, and other pertinent considerations. It is of significance to note that this study can be classified as applied research due to its specific aim of optimizing OCFA biosynthesis. The intention is to obtain quantifiable results from our in-silico experiments to facilitate a comparison of our model with pre-existing ones, underscoring the progress made in this research.

Ultimately, the intent of this methodology chapter is to not only serve as a guide to our research, but also to provide enough information for the replication and extension of this study.

3.1 Description of the model

The rationale for electing *Y. lipolytica* as the foundation for this investigation has been extensively delineated in preceding chapters. This yeast's untapped potential as a host for bioproduction constitutes a compelling proposition for the sustainable evolution of large-scale biofuel, biopolymer, and other high-value bio-compound production [9, 62].

Y. lipolytica, boasting an elaborate assembly of enzymatic pathways, presents a robust platform for genetic engineering. It comprises an extensive genome, encoding for approximately 7,357 genes and 6,472 proteins [9, 53]. Such genomic complexity allows for substantial manipulation, tuning the organism to meet the specific needs of our study, without overburdening its transcriptional, translational, and replication mechanisms.

Furthermore, the extensive and freely available genomic data for *Y. lipolytica*, along with a diverse set of peer-reviewed models [21, 55], establish a fertile ground for continued research in this field using a variety of tools and algorithms. These resources allow us to build on existing knowledge, facilitating an iterative and collaborative approach to the advancement of sustainable bio-production.

As we advance in our research, the capability of *Y. lipolytica* to withstand extensive genetic manipulation while maintaining its cellular viability forms the crux of our methodology. This characteristic, in concert with the available genomic resources, aligns seamlessly with our ultimate research goal, setting the stage for significant strides in the field of sustainable bio-production.

3.1.1 *Y. lipolytica* model comparison

The historical trajectory of model representations of *Y. lipolytica* W29 wild type traces its lineage through seven distinct iterations. The most recent of these, the iYli21, was devised by Guo et al. in 2022 [81]. The creation of this model employed the iYali4 model, designed by Karkohen et al., as its template. However, it had a glaring discrepancy: the erroneous use of the genomic annotation of an alternate *Y. lipolytica* strain, CLIB122. This incongruity was subsequently rectified in iYli21 through the manual replacement of their corresponding W29 gene homologs and the introduction of absent pathways utilizing BlastKOALA prediction results [82].

The predecessor model iYali4 was constructed using the Yeast8 [75], the consensus genome-scale metabolic model of *S. cerevisiae*, as its fundation. Moving further back, we encounter the iYali647 which was crafted by Mishra et al. in 2018 [83]. This model was derived from the 2015 model of Kavšček et al. [84]. Both these models were subjected to meticulous validation employing FBA tools and empirical data.

The first models in this lineage is represented by iNL895 and iYL619, the genome-wide metabolic network representations of *Y. lipolytica*, designed by Loira et al. [22] and Pan and Hua [18], respectively. The iNL895 was constructed using a consensus model of *S. cerevisiae*, augmented with the lipid synthesis pathway. In contrast, iYL619 was built utilizing freely available genomic data and subsequently validated using in silico tools. This latter genome-scale metabolic model (GEM) was later updated by the same team in 2017 to optimize triacylglycerol production, and renamed as iYL_2.0 [85]

In sum, each iteration of these models presents a refinement in our understanding of *Y. lipolytica*'s metabolic dynamics, establishing a solid foundation for the current research and future advances in this realm.

3.1.2 Missing Propionate Metabolic Pathway

The catabolic process of Propionate, a volatile short chain fatty acid, initiates within the extracellular area. It is through the work of a specific transmembrane protein known as the MFS1 transporter that Propionate is shuttled into the cytoplasm, allowing the organism of its absorption and utilization [86]. Once escorted into the cytoplasmic environment, the enzyme propionate CoA-transferase (Pct) acts upon Propionate, transforming it into propionyl-coenzyme A (propionyl-CoA), the metabolically active form of the compound. This crucial transformation has been detailed in the work of Park et al. [52].

After this, the propionyl-CoA can take two different routes. One possible trajectory is its transport into the mitochondria, where it is harnessed for energy production in the tricarboxylic acid (TCA) cycle, further enhancing cellular metabolic efficacy. On the other hand, it may undergo a series of enzymatic transformations within the cytoplasm, culminating in its conversion into odd-chain acetyl-coenzyme A (acetyl-CoA), a precursor molecule for the synthesis of odd-chain fatty acids (OCFAs) [1]. This intricate biochemical pathway, leading from extracellular Propionate absorption to its intracellular utilization, is illustrated comprehensively in Image X and further elucidated by the series of reactions outlined in Table 3.2.

3.1.3 Missing OCFAs Metabolic Pathway

The synthesis pathway of Odd Chain Fatty Acids (OCFAs) commences via the utilization of distinct carbon sources such as propionate, glucose, or acetate [52]. These substrates undergo a biochemical transformation into acetyl-coenzyme A (acetyl-CoA) mediated by the enzymatic actions of Acetyl-CoA Synthetase 1 (ACS1) and Acetyl-CoA Synthetase 2 (ACS2) [87]. This process sequentially continues with the production of malonyl-CoA variants, characterized by their distinct carbon chain lengths [88].

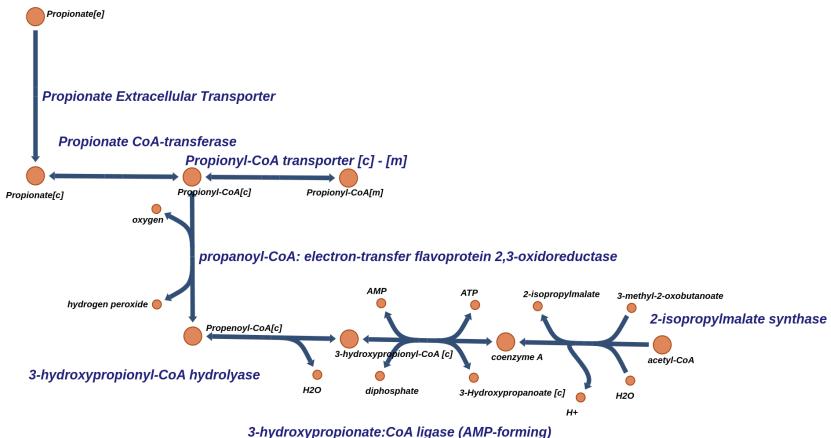


FIGURA 3.2: Illustration of the Included Propionate Pathway.

This malonyl-CoA then serves as a vital substrate, effectively fuelling the Fatty Acid Synthesis (FAS) cycle [88]. Within this metabolic context, two unique routes can be delineated, contingent upon the carbon chain length of the malonyl-CoA precursor.

In instances where the malonyl-CoA chain is of an odd length, the enzymatic process aligns with the Elongation of Very long chain fatty acids (Elovl) pathway, which engages the C14 chain as an intermediate substrate, successively elongating it to C16 and eventually to C18 [89]. On the other hand, the synthesis of odd chain fatty acids follows a distinctive route whereby a C15 chain is synthesized initially, subsequently undergoing elongation to C17 [1].

It is worth noting that this elaborate biochemical cascade occurs within the cytoplasmic compartment of the yeast cells. Importantly, this cell-specific localization lends itself to potential manipulation, potentially offering innovative avenues for metabolic engineering approaches to enhance the production of OCFAs. It is possible to

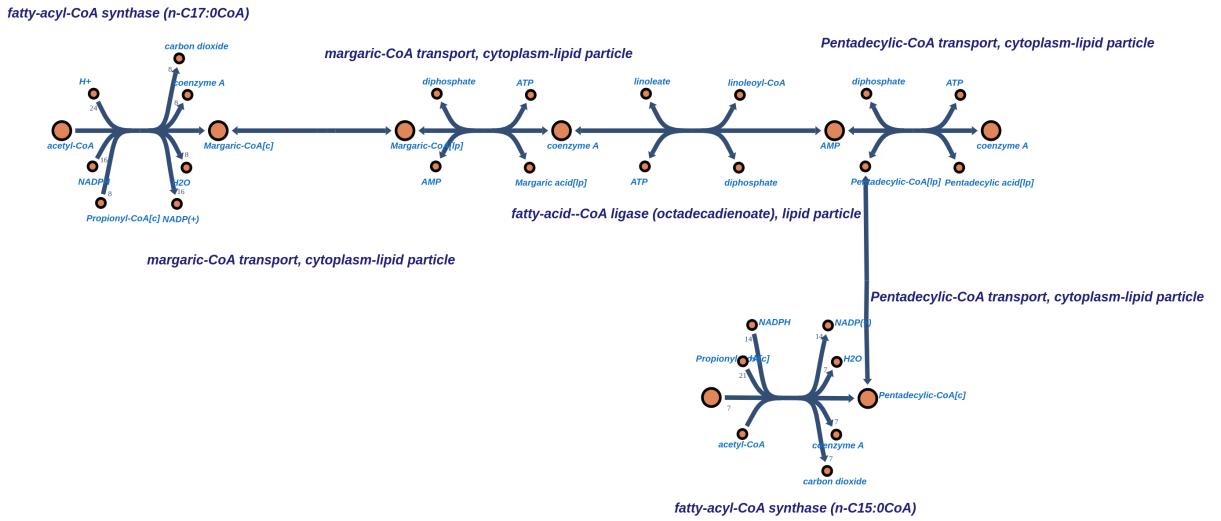


FIGURA 3.3: Illustration of the Included OCFA Pathway.

see the list of reactions added in the Table 3.3 and a representation of the network in the Figure 3.3.

3.2 Model Development

3.2.1 Modeling of Metabolic Networks

Genomic screening of a given organism offers an opportunity to build a comprehensive understanding of its metabolic networks. These networks, composed of substrates, metabolites, enzymes, and biochemical reactions, can be schematically represented through Genome-Scale Models (GEMs) [90]. Translating this biological knowledge into a functional model demands the application of mathematical abstractions

and equations. Notably, we employ abstraction to transcribe complex metabolic reactions into manageable stoichiometric terms, thereby making them conducive for computational modeling [91].

There are primarily two modeling methodologies applied in systems biology. The first, a static approach, posits that the metabolic fluxes within a biological system exist in an equilibrium or quasi-equilibrium state [92]. The second, dynamic modeling, uses experimentally-derived kinetic constants to establish the relationships between metabolic flux expressions and metabolite concentrations. This strategy typically employs ordinary differential equations (ODEs) to resolve the system [93]. Our research leverages both modeling approaches, applying them in Flux Balance Analysis (FBA) and dynamic Flux Balance Analysis (dFBA).

A visual representation of how an organism's metabolic flux can be turned into a stoichiometric matrix can be seen in Figure 3.4. This matrix serves as a mathematical depiction of the intricate network of biochemical transformations that happens within the organism, and is an instrumental component in both static and dynamic metabolic modeling.

3.2.2 Linear Programming

The concept of linear programming (LP) can trace its genesis back to the 1940s, emerging as an innovative solution conceived by G. Dantzig et al. [94]. This concept was borne out of the exigencies of the era, specifically the need to solve intricate problems posed by the global war scenario, characterized by the restricted availability of resources [76]. Linear programming, at its core, aimed at the strategic distribution of these scarce resources across a variety of activities, constrained by the parameters dictated by the specificities of the problem at hand. This methodology, thus, offered a tool

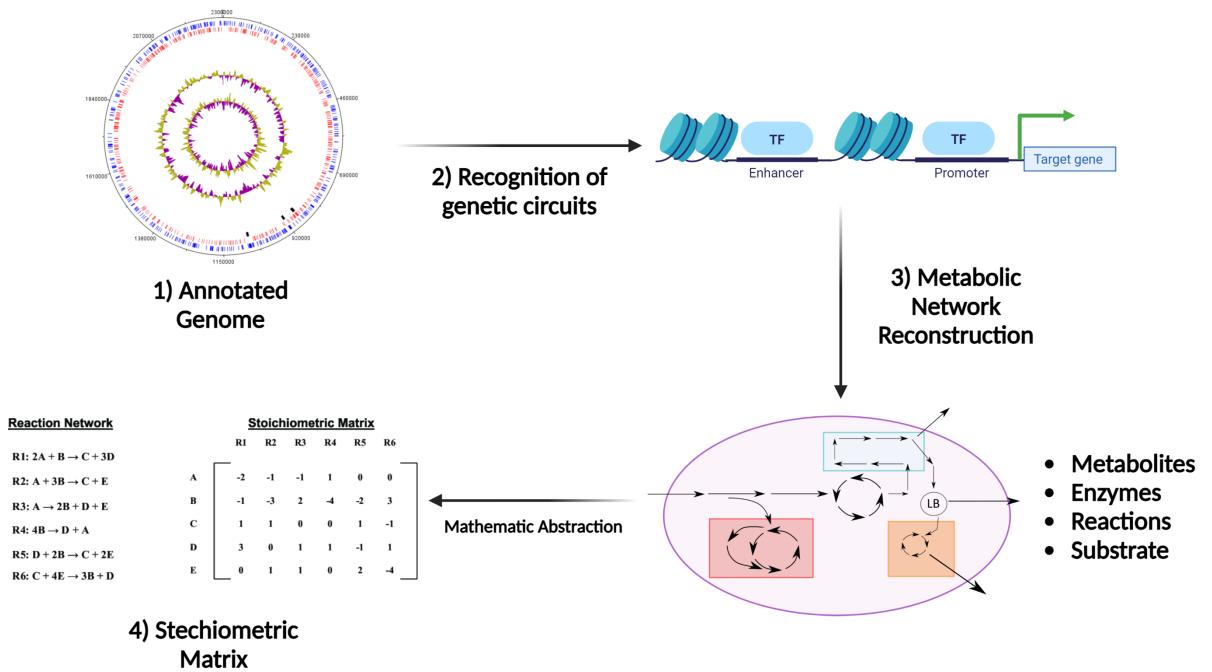


FIGURA 3.4: Illustration of the abstraction process to construct a GEM from the genomic information of an organism to the build of an stoichiometric matrix. 1) It is necessary to have the information of an annotated genome of the organism. 2) Using this data, it is possible to recognise specific genetic circuits and gene encoding proteins. 3) From this circuits, it is possible to reconstruct the different enzymatic networks of the organism. 4) Finally, it is possible to obtain the stoichiometric matrix representation.

for optimizing resource allocation within the binding confines of pre-established rules, thereby signifying a significant leap in computational problem-solving methodologies.

A notable application of linear programming is embodied in its utilization for the optimization of specific reaction fluxes within a Genome-scale Metabolic (GEM) model. This application, colloquially referred to as Flux Balance Analysis (FBA) [8], offers a robust mechanism to holistically understand metabolic flux distribution within a system, thereby enabling precise optimization of desired outcomes. The conceptual representation of this application can be observed in Figure 3.5.

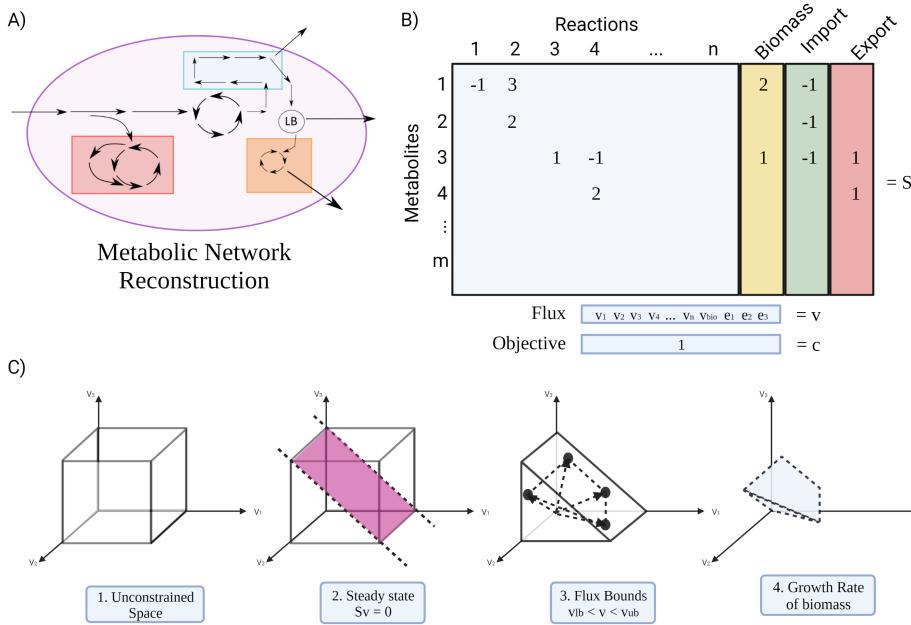


FIGURA 3.5: Illustration of the use of linear programming to compute the optimization of the Genome-scale Metabolic (GEM) model. A) The reconstructed metabolic network. B) This is then used for the construction of the stoichiometric matrix. C) Finally, our stoichiometric matrix can be represented as our unconstrained space that is then limited by our rules and optimization conditions.

3.2.3 Gene-Protein-Reaction Rule

Inherent within the framework of a Genome-scale Metabolic (GEM) model are not only the enzymes and reactions constituting the metabolic network but also encompasses the intricate details of associated genes. This level of granular detail opens a doorway for researchers to examine the phenotypic outcomes arising from genetic alterations, such as gene knockouts.

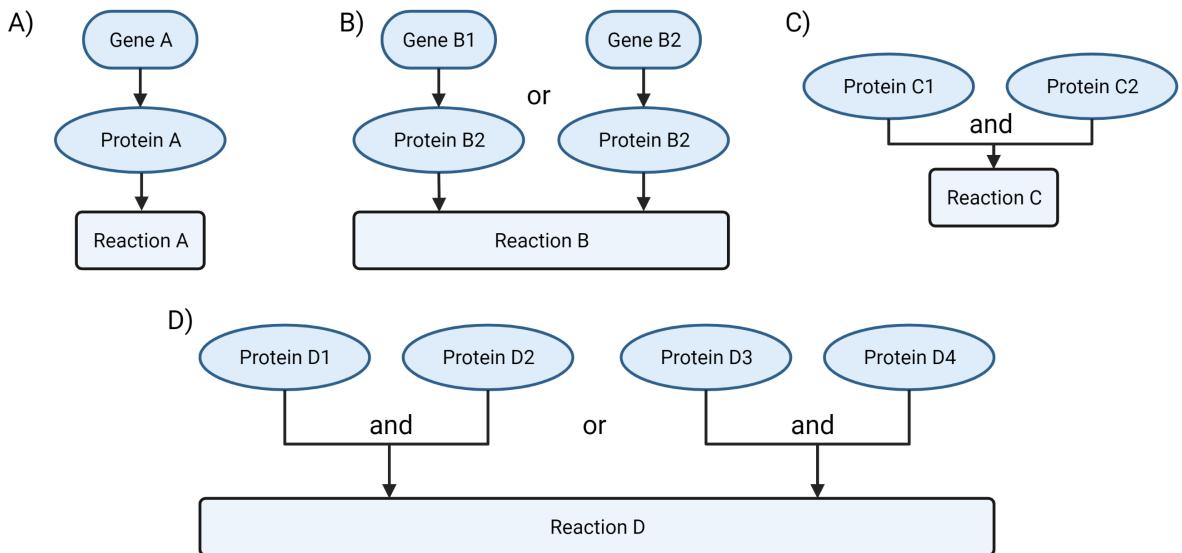


FIGURA 3.6: Diverse set of gene-protein-reaction that are possible. A) Reaction A catalized by a single protein and a single gene. B) Reaction B that can be catalyzed by two isoenzymes. C) Reaction C produced by a protein complex. D) Redundant protein complex that is needed to carry out the reaction D.

One can use these gene knockouts to assess their influence on specific metabolic fluxes or metabolite production, thereby facilitating the understanding of functional genomic elements contributing to cellular metabolism. This association encompassing the gene, enzyme, and the resultant reaction is articulated as Gene-Protein-Reaction associations (GPRs) [95].

As depicted in Figure 3.6, it is necessary to recognize that gene knockouts do not necessarily mean the stop of metabolic flux. This arises due to the common occurrence of enzymatic redundancy within biological systems, where multiple genes often encode for the same enzyme. This phenomenon, whilst conferring robustness to the organism, also provides an opportunity for modulating the rate of a specific reaction, thereby contributing to the flexible and adaptive nature of cellular metabolism.

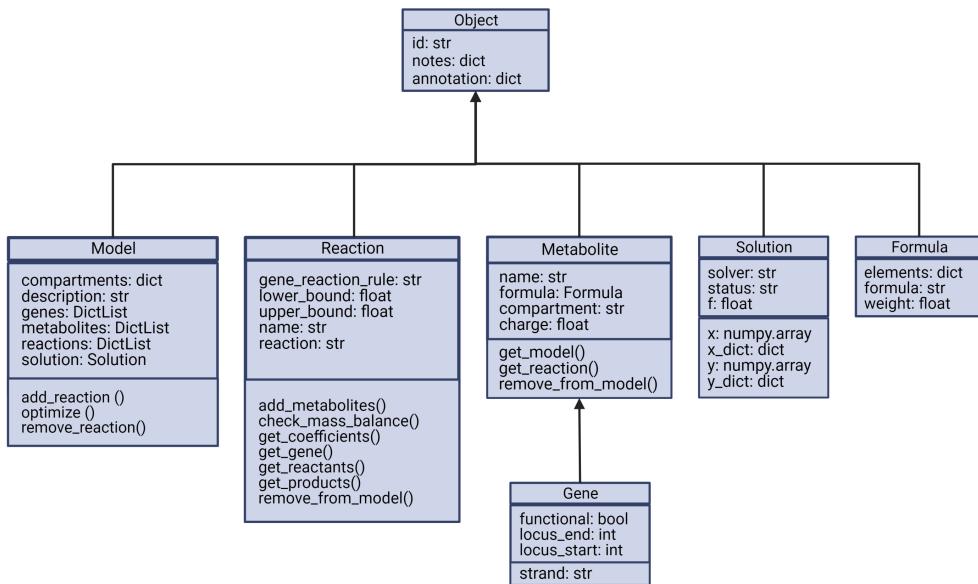


FIGURA 3.7: Illustration of the algorithm structure of COBRApy.

3.2.4 COBRApy

Constraint-Based Reconstruction and Analysis (COBRA) methodology has emerged as an indispensable framework within the realm of metabolic engineering. Primarily, it facilitates genome-scale modeling of metabolic networks encompassing various organisms or even specific tissues [96]. The first implementation of this methodology was achieved through the COBRA Toolbox for MATLAB, as developed by Becker et al. in 2007 [97]. This innovative software facilitated growth prediction of model organisms utilizing a steady-state approach, and enabled the exploration of biomass growth consequences arising from gene deletions. The COBRA Toolbox primarily utilized Systems Biology Markup Language (SBML), an internationally accepted file format for metabolic networks, which streamlined the computations to a few lines of code.

Despite these advances, the COBRA Toolbox initially lacked the framework to effectively integrate multi-omics data. To overcome this shortfall and to enhance the

overall functionality of the software, the openCOBRA community project was conceived [96]. This project, a freely accessible Python library, supports essential COBRA methodologies while also permitting the integration of comprehensive biological data. OpenCOBRA's framework utilizes an object-oriented programming paradigm where each component of the model – be it a gene, reaction, metabolite, or substrate – is treated as an object of a distinct class with specific attributes (as depicted in Figure 3.7). The interactions between these objects within their environment effectively simulate the comprehensive metabolic network of the organism [98]. This holistic approach provides a robust and detailed representation of complex biological processes, bolstering the potential applications of COBRA in metabolic engineering and beyond.

In the context of our Constraint-Based Analysis (CBA) framework, several specific components warrant attention and serve as critical elements of the overall process. Primarily, the Genome-Scale Metabolic Model (GEM) represented by a stoichiometric matrix S is instrumental in managing the metabolic network. The dimensions of this matrix S are defined by the number of metabolites m , represented in the rows, and the number of reactions n , depicted in the columns. This matrix can be interpreted as a system of linear equations where the quantity of equations surpasses the number of variables, thereby rendering it an underdetermined system. To extract particular solutions, it is imperative to incorporate a set of constraints or restrictions into the model.

$$\text{maximize } c^T, \quad (\text{III.1})$$

$$\text{subject to } Sv = 0, \quad (\text{III.2})$$

$$\text{flux vector } V = (v_1, v_2, v_3, v_4, \dots, v_n), \quad (\text{III.3})$$

$$\text{constrain } v_{lb} \leq v \leq v_{up}, \quad (\text{III.4})$$

Subsequently, the fluxes v_i and exchange reactions e_i are encapsulated within a vector

V of length n . To express our steady-state approach in conjunction with the aforementioned elements, we turn to Equation (III.3). In this equation, the metabolite consumption and production within the system are symbolized as $Sv = 0$. Additional constraints are implemented by delineating lower v_{lb} and upper v_{ub} bounds for our fluxes, thereby constricting the solution space.

Lastly, our objective function Z is delineated in Equation Y. It is articulated as a linear combination of our objective fluxes, expressed as $c^t * v$. This equation reflects the interaction of the objective function with each reaction, with the vector of weights c^t signifying the contribution of each reaction to the objective function. In summation, these components establish a comprehensive and rigorous framework for conducting Constraint-Based Analysis in the realm of metabolic engineering.

3.2.5 COBRApy needed to add metabolites, substrate and reactions

As previously mentioned, the objective of the COBRApy framework is to facilitate integrative analysis of genome-scale metabolic models with a distinct emphasis on the quantitative prediction of stoichiometric fluxes that define various phenotypic states of an organism [77]. This toolset, designed to seamlessly integrate with the Python programming environment, equips its users with a plethora of functions for the construction, modification, and analysis of GEMs.

Table 3.4 explains the distinct set of functions offered by COBRApy for various tasks integral to metabolic modeling. This includes the addition of metabolites and reactions, the import of models, etc.

No	Code	Function
1.	model = cobra.io. read_sbml_model('Name.xml')	Imports and reads GEM model that is in a .xml format.
2.	model.compartments	Shows each of the compartments that are specified in the model.
3.	model.metabolites.yyyy	Displays the characteristics and specifications of a described metabolite inside of the model.
4.	model.reactions.xxxx	Displays the characteristics and specifications of a described reaction inside of the model.
5.	reaction = Reaction('xxxx') reaction.name = 'AAAAA' reaction.lowerbound = -1000. reaction.upperbound = 1000.	Creates a specific reaction inside the open model with the code 'xxxx' and the specified name 'AAAAA'. Also describes the lower and upper bound of the reaction's flux, these can range from -1000 to 1000.
6.	a = Metabolite('yyyy', formula='CxHxOx', name='BBBB', compartment='c')	Creates a specific metabolite inside the model that can be used in other reactions. It has the code 'yyyy' and the specified name 'BBBB'. Also describes the chemical formula of the metabolite and the compartment where this metabolite will be included.
7.	reaction.add_metabolites (yyyy_c: -1.0, zzz_c: 1.0)	Adds a certain number of metabolites as the elements that plays a role in a specific reaction. These metabolites can have 2 states: if they are in the -1.0 state then they will be consumed during the reaction, but if they are in the 1.0 state then they will be produced.

TABLA 3.4: List of different COBRApy functions

3.3 Flux Balance Analysis

3.3.1 Stationary Metabolic Network

In the prior chapter, we elucidated the hypothesis of the stationary or quasi-stationary approach as a key component of metabolic network analysis. Illustrated in Figure 3.5, this methodology presumes that both the concentration of extracellular

substrates and intracellular metabolites remain invariant. In this context, the consumption rate of the substrate is meticulously balanced with the production rate of the metabolites [99]. Applying this theoretical framework enables a convenient simplification of Equation (III.5) to Equation (III.6).

The utility of this hypothesis is its allowance for the omission of kinetic constants, such as K_m . Consequently, the velocity of each flux depends only on the stoichiometric of the network. Nonetheless, a computational challenge arises when navigating the solution space of Equation (III.5), as it is undetermined due to the condition $n > m$. Here is where constraint-based modeling in Equations (III.3) and (III.4) becomes instrumental.

The integration of these methodologies enables us to solve the Flux Balance Analysis (FBA) and other analogous approaches. Thus, this conceptual framework provides a pragmatic strategy for investigating the expansive and complex nature of metabolic networks, establishing a foundation for further explorations into the space of feasible solutions.

$$\frac{d(\text{mass})}{d(\text{time})} = S * v(\text{metabolite}, \text{substrate}, \text{product}) = 0, \quad (\text{III.5})$$

$$S * v = 0, \quad (\text{III.6})$$

$$v > 0, \forall i \in \text{irreversible reactions}, \quad (\text{III.7})$$

3.3.2 Tools for FBA Analysis with COBRApy

As it was mentioned, COBRApy offers a extensive list of functions that can help researchers to analyse GEM with the approach of Flux Balance Analysis. In the table 3.5 is possible to see a detail description of each function used during the research:

No	Code	Function
1.	model.objective='xxxx'	It shifts the optimization objective on the model to the specified reaction 'xxxx'.
2.	solution = model.optimize()	Runs an optimization algorithm based on linear programming with the reaction previously specified as an objective. The result would give us the mathematical result of the maximum possible production of the objective reaction.
3.	cobra.flux_analysis.single_reaction_deletion()	Allows for the analysis of the effect of deleting each reaction one at a time
4.	cobra.flux_analysis.single_gene_deletion()	Allows for analysis of the effect of deleting each gene one at a time
5.	cobra.flux_analysis.loopless.loopless_solution()	This function finds a flux distribution that is both optimal and loopless (does not contain internal cycles), which can be more biologically realistic
6.	.find_thermodynamic_discrepancies()	This function checks if a flux distribution is thermodynamically feasible.
7.	cobra.manipulation.delete_model_genes()	This function takes a list of gene IDs to be knocked out and simulates it.
8.	reaction.upper_bound *= 2	This function alters the upper bound of a specific reaction and double its speed in order to simulate a overexpression of that enzyme.

TABLA 3.5: List of different COBRApy functions for Flux Balance Analysis

3.4 dynamic Flux Balance Analysis

Dynamic Flux Balance Analysis is an extension of FBA where dynamic cellular behaviour gets incorporated into the steady-state assumption without contradicting it.

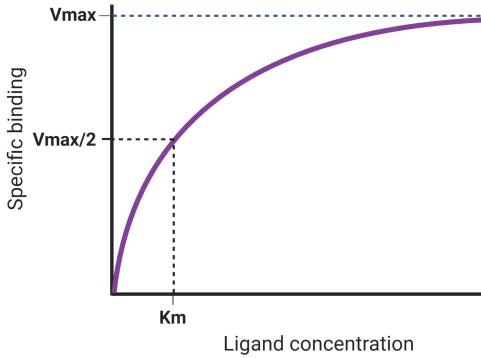


FIGURA 3.8: Graph function of the speed of a metabolic reaction using the Michaelis-Menten equation. The x axis represent the concentration of the substrate, the y axis is the speed of the reaction, V_{max} is the maximum speed of the reaction, and the K_m is the Michaelis-Menten constant known as the constant affinity of an enzyme for its substrate.

The dFBA focuses on representing the transient behaviour of specific metabolic fluxes in response of changing conditions over time [80]. This depends on the established objective of the whole optimization model. The main challenge to implement dFBA is to simulate the time-varying rate of enzymatic reactions.

3.4.1 Michaelis-Menten Equation

One of the approaches for simulating the dynamics of an enzymatic reaction is encapsulated in the Michaelis-Menten equation, a key model in the field of enzyme kinetics.

$$v_i = V_{max} * \frac{m_i}{K_m + m_i}, \quad (\text{III.8})$$

The mathematical expression in Equation (III.8) describes the rate of enzymatic reactions, relating the reaction rate (v) to the concentration of a substrate (S) and the two intrinsic enzyme constants, V_{max} and K_m [100, 101].

When incorporated into the dynamic Flux Balance Analysis (dFBA), the substrate concentration S—representative of an internal metabolite within the model, changes over time. This temporal variation produces changes in the rate of consumption and production across the network until an eventual cessation is reached. Kinetic constants are commonly determined experimentally, although it can be obtained by bibliographic research and data repositories, such as BiGG [102], BioModels [103], and JWS Online [104], provide a valuable resource for their extraction. The integration of this kinetic information contributes a depth of complexity and realism to the metabolic network modeling, rendering the application of the Michaelis-Menten kinetics within dFBA as a potent tool for in-depth and dynamic metabolic investigation.

3.4.2 Kinetics used for the dFBA model

In the application of Dynamic Flux Balance Analysis (dFBA), the integration of the model's stoichiometric matrix with Michaelis-Menten kinetics provides a powerful framework for simulating complex metabolic dynamics. By incorporating substrate uptake kinetics, we can derive a system of equations—namely equations X and Y—which serve as the backbone of our dFBA approach [11].

$$Vg = Vg_{max} * \left(\frac{G}{K_g + G_1} \right) * \left(\frac{1}{1 + \frac{E}{K_i}} \right), \quad (\text{III.9})$$

$$\frac{\partial X_g}{\partial t} = u_g * X_g, \quad (\text{III.10})$$

$$\frac{\partial X_a}{\partial t} = u_a * X_a, \quad (\text{III.11})$$

In Equation (III.9), G represents the extracellular concentration of a specific substrate, Vg_{max} denotes the maximum uptake rates of each substrate, K_g stands for the corresponding saturation constant, and K_i signifies the inhibition constant [7]. The equation thus models the uptake rate of a particular substrate or metabolite, which is an adaptation of the Michaelis-Menten equation with the incorporation of an inhibitory term. This inhibitory term enables the reflection of growth rate suppression effects due to the presence of specific metabolites or by-products.

Conversely, Equation Y encapsulates the cellular growth dynamics where Xg indicates the biomass concentration for specific metabolites or substrates, Veg represents the synthesis velocity of these metabolites, and u_g delineates the individual growth rate.

The particular kinetic constants deployed during this research are listed in Table Y. It is pertinent to note that these constants were selected following a rigorous review of available literature and databases. Furthermore, they were taken from *S. cerevisiae* in-vitro experiments due to been already characterize.

Parameter	Description	<i>S. cerevisiae</i> value	Refs.
Vg_{max} (mmol/g/hr)	max. uptake rate of glucose	22.4	[105]
K_g (g/L)	Saturation constant of glucose	0.8	[105]
Vo_{max} (mmol/g/hr)	max. uptake rate of oxygen	2.5	[105]
K_o (g/L)	Saturation constant of oxygen	0.003	[105]
Vx_{max} (mmol/g/hr)	max. uptake rate of xylose	0.6	[105]
K_x (g/L)	Saturation constant of xylose	0.0165	[106]
$K_{i,g}$ (g/L)	Inhibition constant of glucose	0.5	[106]
$K_{i,e}$ (g/L)	Inhibition constant of ethanol	10	[106]
X_g (g/L)	Molecular weight of glucose	180.1559	[106]
X_x (g/L)	Molecular weight of xylose	150.13	[106]
X_o (g/L)	Molecular weight of oxygen	16.0	[106]
X_e (g/L)	Molecular weight of ethanol	46.0	[106]

TABLA 3.6: List of used kinetics for dFBA

3.4.3 Tools for dFBA Analysis

In order to conduct the Dynamic Flux Balance Analysis (dFBA) for our model, the dfba library specifically designed for Linux was utilized. This algorithm, conceived by Harwood et al. [107], was developed with a primary aim of resolving linear programming issues. Constructed in C++, the algorithm leverages the GNU Linear Programming Kit (GLPK) and the Suite of Nonlinear and Differential/Algebraic Equation Solvers (SUNDIALS), specifically CVODE or IDA. Notably, the dfba library is designed

to harmonize with COBRApy within a Python 3.7 environment, thereby bolstering its utility in the context of our research.

In the table 3.7 is possible to see a detail description of each function used during the research. This tabular resource provides a comprehensive breakdown of each function, offering a detailed description that elucidates its respective application within the context of our research.

No	Code	Function
1.	<code>dfba_model = DfbaModel('xxx')</code>	It grabs a FBA model and initialize a dFBA file
2.	<code>X = KineticVariable("yyy")</code>	'yyy' is a metabolite or substrate that will be analysed in the dfba
3.	<code>dfba_model. add_kinetic_variables([X])</code>	Allows for the analysis of an x number of metabolites at the same time
4.	<code>mu = ExchangeFlux("www")</code>	Allows for analysis of specific fluxes inside of the model
5.	<code>dfba_model. add_exchange_fluxes([mu])</code>	Allows for the analysis of an x number of fluxes at the same time
6.	<code>dfba_model.add_rhs _expression('yy', mu * X)</code>	This function makes a relationship between the flux and the metabolite to analyse
7.	<code>dfba_model.add_exchange_flux_lb</code>	This function adds the Michaelis-Menten kinetic formula using a specific flux and metabolite
8.	<code>dfba_model.add_initial_conditions</code>	This function specify the initial concentration of each studied metabolite
9.	<code>concentrations, trajectories = dfba_model.simulate</code>	This function initialises the dfba with an specific simulated time

TABLA 3.7: List of different dfba functions

3.5 Performance Indicators

3.5.1 Contrasting Laboratory Data vs Modelled Data

The empirical data, harnessed in this study for comparative analyses, was graciously provided by the Laboratory Biologie Intégrative du Métabolisme Lipidique (BIMLip), based in Paris, France. This invaluable dataset, meticulously accumulated throughout the duration of 2020 and 2021, served as the cornerstone for our computational modelling, reinforcing the credibility and robustness of our methodological approach. It is worth emphasizing that this data, at the time of our research, is still in an unpublished status, therefore should be viewed as preliminary and subject to further verification and possible modification upon its formal publication. Despite this, the generous provision of this data by BIMLip has been pivotal for the advancement of this research, substantially informing the ensuing findings and discussions.

3.5.2 Modelled Performance Indicators

In our endeavor to delineate the quantitative performance of our modelled data vis-à-vis the laboratory data, we employed three distinct statistical indicators: the coefficient of determination (R^2), the Mean Absolute Error (MAE), and the Root Mean Squared Error (RMSE). These three measures of fit provide a comprehensive picture of the model's predictive accuracy, each illuminating a unique aspect of the model's performance.

The coefficient of determination, R^2 , is a statistical metric that denotes the proportion of variance in the dependent variable that is predictable from the independent variable(s) in our model. This provides a quantitative measure of how accurately our model encapsulates the inherent variability in the data. An R^2 value that approaches 1 suggests that a significant proportion of the variability in the output variable has been

accounted for by the predictor variables within our model. The precise mathematical definition and derivation of the R² statistic is provided in Equation III.12.

$$R^2 = \sum_{t=1}^T (Y_{real} - Y_{modelled})^2 \quad (\text{III.12})$$

Root Mean Squared Error (RMSE), in contrast, presents an estimate of the standard deviation of the residuals, or prediction errors. It provides a measure of how closely the data are clustered around the line of best fit, giving an estimate of the dispersion of the residuals. A RMSE value of 'X' suggests that, on average, the model's predictions diverge from the actual values by 'X' units. Therefore, a lower RMSE value generally indicates a superior fit of the model. The computational formula for RMSE is detailed in Equation III.13.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}} \quad (\text{III.13})$$

Finally, the Mean Absolute Error (MAE) provides a direct interpretation of the average magnitude of the error in our model's predictions, irrespective of their direction. Unlike RMSE, MAE does not square the errors in the calculation, hence large errors are not overly penalized. Therefore, MAE is a useful metric when the primary interest lies in understanding the average magnitude of the prediction errors, regardless of their direction. The computational process for deriving MAE is detailed in Equation III.14.

$$MSE = \frac{|\sum_{i=1}^N (Predicted_i - Actual_i)|}{N} \quad (\text{III.14})$$

3.6 Independent and Dependent Variables

In light of the methodology delineated heretofore, the primary objective of the current research, as delineated earlier, is to explore avenues for enhancing the biosynthetic capabilities of *Y. lipolytica*. Our investigation seeks to identify an optimized phenotype through in silico simulation of gene overexpressions and knockdowns in the genome-scale model, thereby serving as our independent variables.

The responses to these genetic manipulations - the dependent variables - manifest as alterations in the yield of *Y. lipolytica* biomass and the production of Odd-Chain Fatty Acids (OCFAs), particularly C15 and C17. These are the variables that directly indicate the effectiveness of our proposed genetic modifications.

This encompassing view facilitates a systematic investigation, permitting the discernment of nuanced dependencies and impacts that shape the final outcome of the study. It underscores the breadth of variables manipulated and measured within this research, underscoring the comprehensive nature of our approach to optimizing *Y. lipolytica*.

3.7 Ethical Considerations, Limitations and Assumptions

During this study, its main focus was the computational modeling and optimization techniques through in silico experiments. Therefore, no human subjects or animal experimentation were involved. On the other hand, all data used in this study were publicly available in vitro results or generated through computational simulations, ensuring compliance with ethical standards and intellectual property rights.

Also, it is important to acknowledge the limitations of our methodology. First, the models utilized in this research rely heavily on available past GEM models of *Y. lipolytica* and biochemical data. Any biases, inaccuracies, or incomplete information

present in the underlying data sources may affect the accuracy of the final result. Additionally, the computational simulations performed are based on assumptions and simplifications that may not fully capture the complexity of the biological system under investigation. The predictive power and applicability of the models are limited by the validity of these assumptions.

Furthermore, this research assumes that the enzymatic pathways and kinetic constants implemented in the model accurately reflect the functioning of *Y. lipolytica*. Although efforts were made to gather comprehensive information and integrate the most up-to-date knowledge, it is crucial to recognize that our understanding of biological systems is continually evolving. New discoveries and insights may emerge that could challenge the validity of certain assumptions made in this study.

Despite these limitations, this research provides valuable insights into the potential optimization strategies for Odd Chain Fatty Acid production in *Y. lipolytica*. By acknowledging these limitations and assumptions, we ensure transparency and promote further investigation to refine and enhance the accuracy and applicability of the computational models and optimization techniques employed in this study.

Characteristics	iYL619_PCP	iNL895	IMK735	iYali4	iYL_2.0	iYL647	iYli21*
Year	2012	2012	2015	2016	2017	2018	2022
No. Genes	596	895	735	901	645	647	1058
No. Metabolites	849	1847	1111	1683	1083	1119	1868
No. Reactions	1,142	2,002	1,336	1,985	1,471	1,347	2,285
No. Metabolic Reactions	780	-	895	1162	1000	917	-
Cytosol	650	-	623	577	884	643	-
Mitochondria	130	-	159	181	304	147	-
Peroxisome	-	-	92	104	21	89	-
Endoplasmic reticulum	-	-	5	228	124	3	-
Golgi Body	-	-	7	77	-	6	-
Vacuoles	-	-	3	18	-	3	-
Nuclear	-	-	12	24	-	12	-
Inter-Compartment transport	236	-	236	498	333	306	-
Environmental Exchange Reactions	126	168	125	318	282	123	-

TABLE 3.1: Summary of the characteristics of each model. The data was obtained partially from the work of Mishra et al. [83] and Guo et al. [81]. *The data from the iYali21 model was unavailable to access.

Rx ID	Description	Formula	Lower Bound	Upper Bound
y300090	Propionate Transporter [e] - [c]	s_3717 → s_3718	0.0	1000.0
y300089	Propionate to Propionyl-CoA propanoyl-CoA:electron-transfer	s_3718 → s_3719	0.0	1000.0
y300088	flavoprotein 2,3-oxidoreductase electron transfer flavoprotein	s_3719 → s_1382	0.0	1000.0
y300087	2,3-oxidoreductase	s_1275 + s_3719 → s_0837 + s_3721	0.0	1000.0
y300086	3-hydroxypropionyl-CoA hydrolase	s_3721 → s_0803 + s_3722 s_0423 + s_0633 + s_3722 →	0.0	1000.0
y300085	3-hydroxypropionate:CoA ligase (AMP-forming) 3-hydroxypropanoate:	s_0434 + s_0529 + s_3723 → s_1198 + s_3723 →	0.0	1000.0
y300084	NAD+ oxidoreductase	s_0794 + s_1203 + s_3724	0.0	1000.0
y300083	3-Oxopropanoate:NAD+ oxidoreductase (decarboxylating, CoA-acetylating)	s_0529 + s_1198 + s_3724 → s_0373 + s_0456 + s_0794 + s_1203	0.0	1000.0

TABLA 3.2: List of Propionate Pathway Reactions Added into the iYali model

Rx ID	Description	Formula	Lower Bound	Upper Bound
y300099	fatty-acyl-CoA synthase (n-C17:0CoA) margaric-CoA transport, cytoplasm-lipid particle	$s_0373 + 24.0 s_0794 + 16.0 s_1212 + 8.0 s_3719 \rightarrow$ $8.0 s_0456 + 8.0 s_0529 + 8.0 s_0803 + 16.0 s_1207 + s_3728$	0.0	1000.0
y300100	margaric-CoA transport, margaric-CoA cytoplasm-lipid particle	$s_3728 \leftrightarrow s_3729$	-1000.0	1000.0
y300102	fatty-acyl-CoA synthase (n-C15:0CoA)	$s_0635 + s_2842 + s_3729 \leftrightarrow$ $s_0531 + s_2840 + s_3730$	-1000.0	1000.0
y300104	Pentadecyclic-CoA transport, cytoplasm-lipid particle	$s_0373 + 21.0 s_0794 + 14.0 s_1212 + 7.0 s_3719 \rightarrow$ $7.0 s_0456 + 7.0 s_0529 + 7.0 s_0803 + 14.0 s_1207 + s_3731$	0.0	1000.0
y300105	Pentadecyclic-CoA transport, cytoplasm-lipid particle	$s_3731 \leftrightarrow s_3732$	-1000.0	1000.0
y300106	cytoplasm-lipid particle	$s_0635 + s_2842 + s_3732 \leftrightarrow$ $s_0531 + s_2840 + s_3733$	-1000.0	1000.0

TABLA 3.3: List of OCFA Pathway Reactions Added into the iYali model

CAPÍTULO IV

RESULTS

Puede ser incluido como un capítulo o puede ser una sección aparte dependiendo del enfoque del trabajo.

4.1 New iYali Model

The model construction phase culminated in the integration of 32 metabolites and 16 metabolic networks, thereby enhancing the overall complexity and utility of our metabolic model. The detailed schematic representation of the model, showcasing the interconnected metabolic networks and the novel metabolites, can be found in Figure 4.1. This visual illustration provides a comprehensive perspective on the structural and functional attributes of the new iYali model.

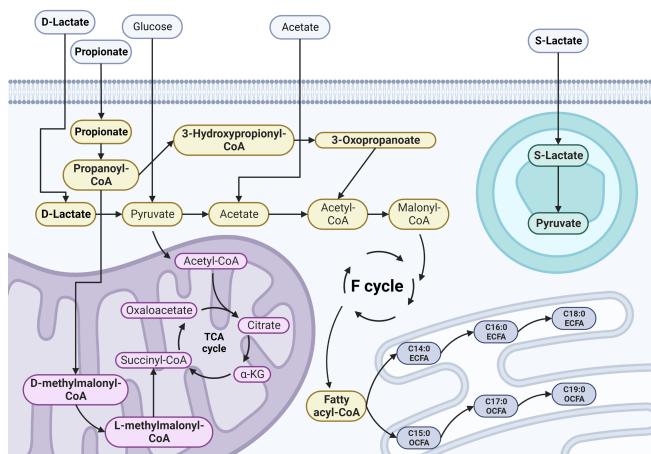


FIGURA 4.1: Illustration of the *Y. lipolytica* GEM with the inclusion of the new pathways. The included metabolites and pathways are highlighted.

It is essential to underscore that Figure 4.1 provides a simplified, illustrative representation and does not encompass the complete array of enzymatic reactions and metabolites incorporated into the revised iYali model. For a more comprehensive comparison, please refer to Table 4.1, which outlines a juxtaposition of the initial iYali4 model and the enhanced iYali version, detailing the final tally of reactions in each model. This side-by-side comparison enables a clearer understanding of the model expansion and its implications for subsequent analyses.

Characteristics	iYali4	new iYali
No. Genes	901	937
No. Metabolites	1,683	1,751
No. Reactions	1,985	2021

TABLA 4.1: Contrast between base model iYali4 and develop model new iYali.

4.2 Analysis of the newYali model and Selection of Genes to be modified

The comprehensive computational model, denoted as iYali and illustrated in Figure 3.7, constructs a network structure that associates each gene incorporated within the model to its corresponding metabolic reaction. This procedure was employed not only for the original genes within the model but also for the newly incorporated enzymatic pathways investigated in the present study.

To dissect the implications and significance of each gene and its associated enzymatic pathway, a specialized computational approach was implemented. This approach permits an exploration into the perturbation effects on the model's growth rate, following a systematic knockout of each individual gene within the model.

The outcomes of this thorough analysis are depicted in Figure 4.2. Each sky-blue line in Figure Y symbolizes the growth rate of the model, specifically under a condition where a distinct gene has been knocked out. The Y-axis denotes the resulting growth

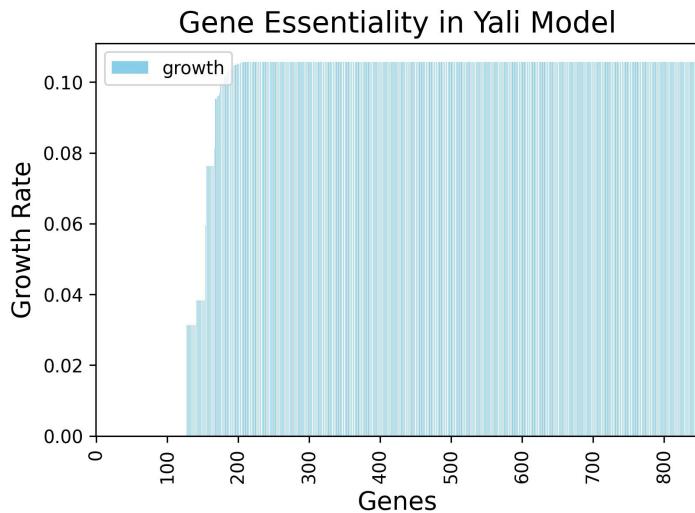


FIGURA 4.2: Graphic that illustrates the different growth rates in the model given each one gene knockout.

rate, quantified in gram dry weight per hour (gDW/h), for each distinct instance of the model. Conversely, the X-axis signifies the label of each gene that has been knocked out in the model. Each label begins with the prefix "YALI0A00." and ends with a unique identifier ranging from 1 to 800. Hence, the label of each gene reflects the knockout that has been performed on the model, allowing for the quantification of the impact of each individual gene on the model's growth rate.

4.3 Enhanced Phenotype

The utility of COBRApy and Genome-Scale Metabolic Models (GEMs) lies in their ability to accurately simulate the impacts of genetic modifications within the model. As depicted in Figure 4.3, we introduced a suite of targeted alterations to create the 'enhanced' version of the new iYali model. During its development, our primary objective was to overexpress key reactions that could potentially unlock an enhanced

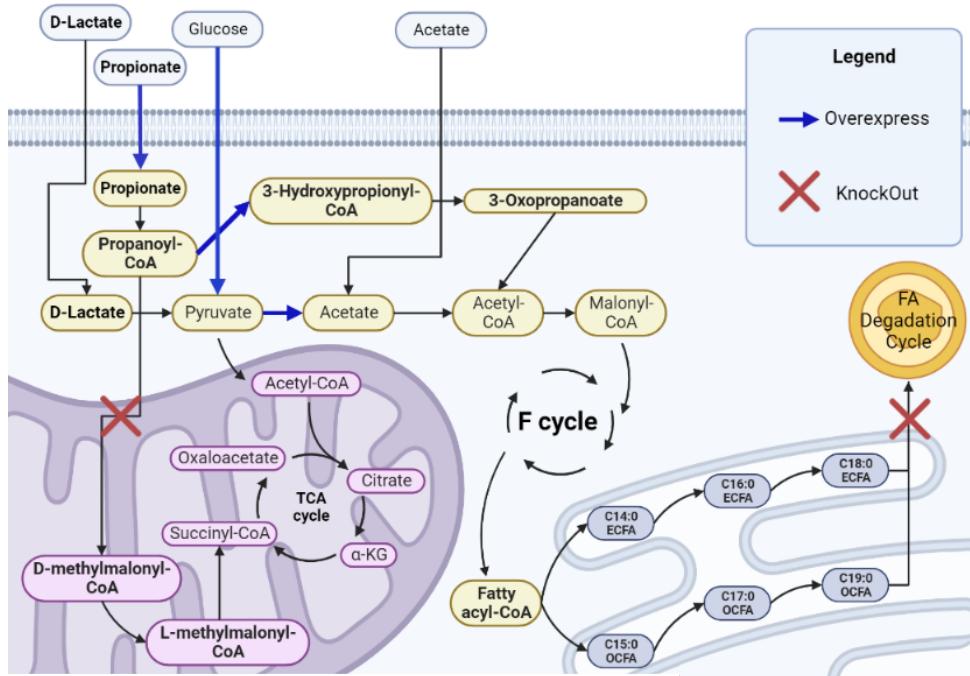


FIGURA 4.3: Graphic that illustrates the different growth rates in the model given each one gene knockout.

production of fatty acids. Consequently, the majority of these genetic modifications were strategically focused on the propionate consumption pathway.

Furthermore, we aimed to channelize the entirety of propionate consumption into the production of acetate. Acetate serves as an intermediary metabolite that feeds into the production of Odd-Chain Fatty Acids (OCFAs) and Even-Chain Fatty Acids (ECFAs). Additionally, we implemented a knockout of the lipid acid transporter leading to the peroxisome. This component functions to recycle these metabolites for further derivation into pyruvate and citrate, subsequently employed for ATP production. By knocking out this pathway, we sought to ensure that the valuable compounds already synthesized within *Y. lipolytica* are not re-routed for internal consumption at a later stage.

These genetic modifications were selected following a rigorous review of the extant literature on *Y. lipolytica* genetic modifications, supplemented by an iterative process of Flux Balance Analysis (FBA) and dynamic FBA (dFBA). Our systematic approach enables us to strike a balance between the theoretical insights drawn from scholarly sources and the practical considerations unearthed through computational modeling.

4.4 Flux Analysis of Created Phenotypes with FBA

As elucidated earlier, the study comprises three distinct models: the original iYali model, the modified iYali inclusive of both the propionate, and the C15 and C17 synthesis pathways, and the further enhanced version of the modified iYali, wherein several gene knockdowns and overexpressions have been simulated to augment the flux density of our key metabolic pathways.

Figure 4.4 offers a comparative assessment of the flux density across a range of critical pathways, encompassing biomass production, amino acid synthesis, propionate absorption, and synthesis of C15 and C17 fatty acids. Positioned to the left of the graph are the codes delineating each reaction integral to these pathways. The names of the models under comparison are furnished at the bottom of the graph. The right side of the graph illustrates the color-coded legend, denoting the flux density, the units for which are designated as millimoles per gram of dry weight per hour ($mmol gDWh^{-1}$). Notably, the flux densities of these pathways provide insights into the metabolic capabilities and efficiencies of each model.

These simulations were conducted utilizing the optimization function embedded within the COBRApy library that was previously mentioned in Table 3.5. In order to emulate the natural metabolic process of the yeast, the optimization objective was set to maximize biomass production, thereby guiding the simulation towards the most biologically relevant state.

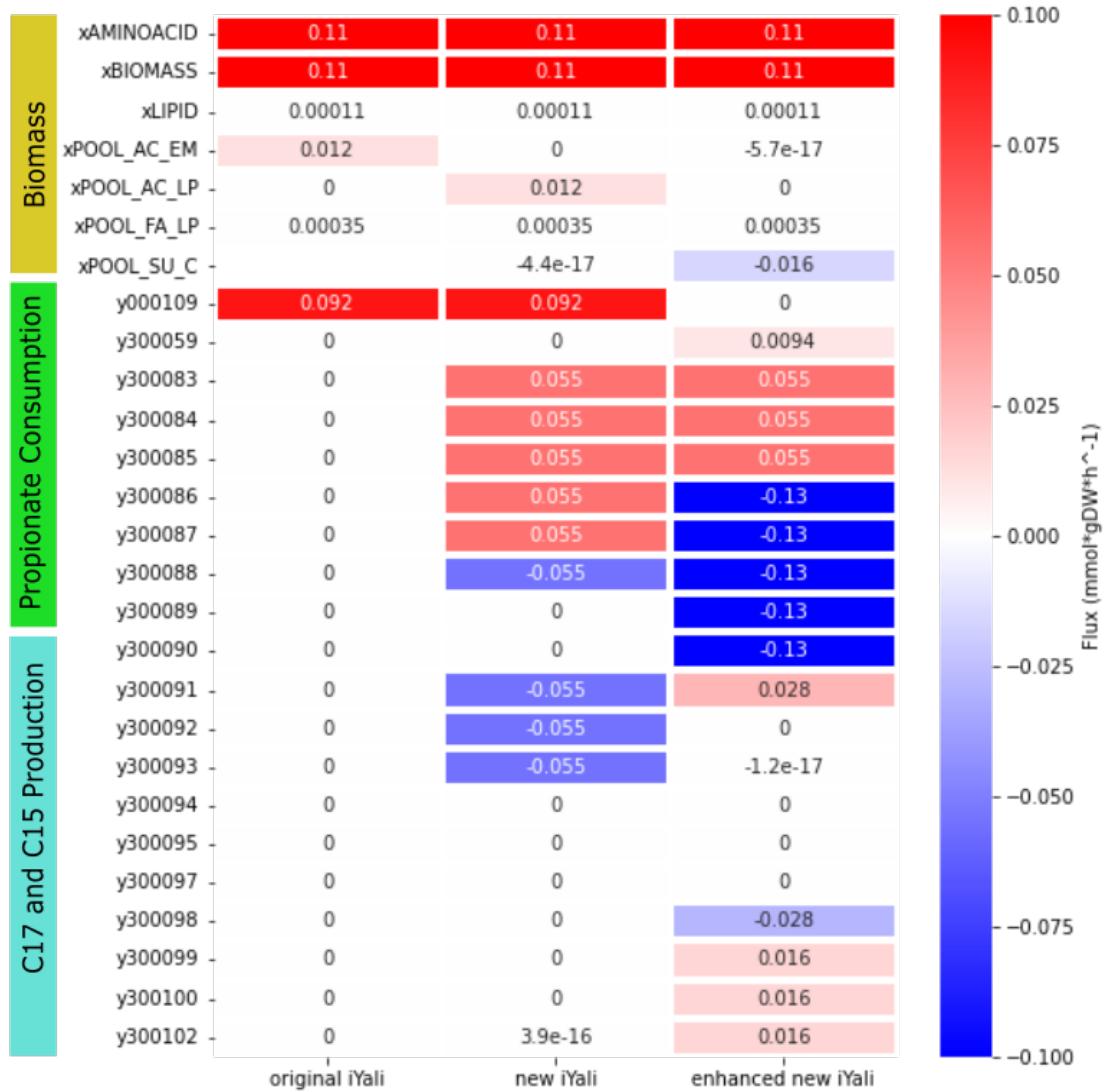


FIGURA 4.4: Flux production of different key metabolites in the different proposed models

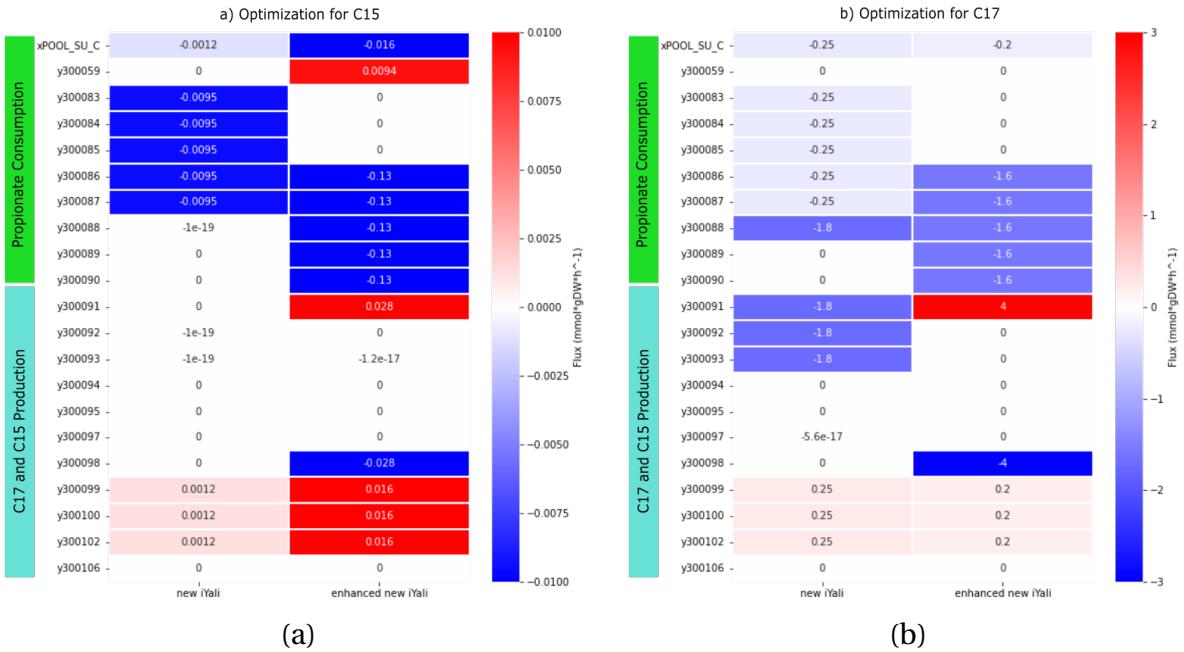


FIGURA 4.5: Flux density of the new iYali model and the enhanced new iYali model in 2 different scenarios. (a) Production optimizations of C15. (b) Production optimizations of C17.

4.5 Looking for key pathways on the optimization of C15 and C17 production

This section juxtaposes two models under investigation, the modified iYali model and the enhanced modified iYali model, under two separate scenarios. It should be noted that despite the appearance of negative and positive fluxes, these values merely signify the direction of specific fluxes and are not our principal concern. Rather, our emphasis lies upon the flux densities, which are essentially the absolute values of the flux.

The subsection denoted as (a) in Figure 4.5 directs our model to prioritize the production of OCFA C15 or pentadecylic acid, represented by the flux of y300106. Interestingly, both the modified and enhanced iYali models yield a null flux for this reaction, a stark contrast to the flux of y300102, corresponding to the synthesis of margaric acid

or C17 OCFA. The comparable flux densities observed in the y30098 to y300100 reactions indicate the model's propensity to produce C17, even when optimized for C15. This behavior could potentially be attributed to the mathematical underpinnings of the *optimize()* function in the COBRApy library as delineated in Table 3.5, wherein it may be more efficient to channel metabolic resources toward the production of Margaric Acid, given the molecular weight of the metabolites involved in both reactions. This hypothesis is corroborated by the observations made in subsection (b) of Figure 4.5.

Distinct differences are observed in the fluxes of the two models in relation to propionate consumption. The new iYali model exhibits higher flux densities from y30083 to y30087, which are instrumental in the transformation of 3-Oxopropanoate to acetyl-CoA on the mitochondria, subsequently contributing to ATP production. Conversely, the enhanced model exhibits superior absorption and synthesis of propionate to propionate-CoA on the cytoplasm, which is then utilized for OCFA production. This can be observed from y300088 to y300091, implying a favorable outcome from the enhanced model, given that the simulated genetic modifications aim to steer the model towards improved propionate absorption.

In subsection (b) of Figure 4.5, the optimization objective is to enhance the production of margaric acid or C17 OCFA, signified by the flux of y300102. A key observation is the increase in the density of the overall reactions compared to the previous scenario, suggesting that C17 production operates more efficiently than C15, potentially hinting at a similar preference in the actual organism. Yet again, a similar trend is observed in the fluxes from y300083 to y300090 in both scenarios, reinforcing the conclusion that the enhanced model is achieving its objective of increased propionate consumption for OCFA synthesis.

The intriguing anomaly in the production of C17 and C15 is that the flux density

for the (R)-lactate to pyruvate conversion (y300098) is substantially higher in the enhanced model. It can be perplexing until we take into account that there is a reaction occurring much later in the metabolic network where pyruvate is converted into acyl-CoA via the y000961 reaction. This acetyl-CoA is subsequently used in the production of C17 via the y300099 reaction. Furthermore, a higher production of C17 is observed in the modified iYali model, which suggests that the glucose absorption pathway, which is downregulated in the enhanced model due to the implemented modifications, also plays a critical role in OCFA production. Thus, glucose absorption should not be overlooked in future optimization strategies.

4.6 Predicted Production of Highly Valuable Metabolites

As depicted in Figure 4.12, we have examined the predicted production of several valuable compounds, namely Oleate, Oleoyl-CoA, Palmitoyl-CoA, Stearate, and Stearyl-CoA. These are Even Chain Fatty Acids (ECFAs) with substantial potential for use as biomaterials or industrial reagents. To assess these compounds' potential, we have utilized the newly modified *Yarrowia lipolytica* W29 strain model, 'new iYali', under various culture conditions.

The range of simulated conditions includes: Glucose alone, Glucose-Isobutanol, Glucose-Isocitrate, Glucose-Stearate, Glucose-Succinate, Glucose-Xylose, Glucose-Propionate, and Glucose-Pyruvate. The rationale for this diverse selection of carbon sources stems from their prevalence in both laboratory and industrial contexts. Specifically, Glucose, Succinate, and Xylose are commonly utilized carbon sources for yeast cultivation. In contrast, Isobutanol, Propionate, Isocitrate, Stearate, and Pyruvate are frequently encountered in wastewater or as by-products from industrial bioprocesses.

Comparative analysis of the ECFA production across all simulated media revealed that the Propionate-Glucose and Glucose-Pyruvate conditions exhibit the most

promising production profiles. This finding underscores the potential for harnessing *Yarrowia lipolytica* as a versatile cellular chassis for the bioconversion of otherwise underutilized or waste compounds into valuable metabolites, thereby promoting more sustainable and efficient biotechnological processes.

4.7 Real Data vs Model Comparison with dFBA tools

This section will be subdivided depending on the different conditions that the model was simulated.

4.7.1 Biomass Growth in Propanoate and Glucose Medium

The employment of dynamic Flux Balance Analysis (dFBA) permitted us to extend our simulation of *Y. lipolytica* growth to a temporal dimension, as denoted in Table 3.7. In particular, we considered the bacteria's consumption of glucose and propionate, two metabolites that, as previously discussed, have been linked to increased lipid accumulation and the production of both even and odd fatty acids.

The contrast between in-vitro and in-silico findings, exhibited in Figure 4.7, elucidates the capability of our newly constructed model, 'new iYali,' in predicting the growth of *Y. lipolytica* in the presence of the specified metabolites. The experiment initialized with propionate and glucose concentrations set at 11.546 mg/L and 5.2875 mg/L respectively. This initial setup precipitated two distinct growth phases: a primary exponential growth phase that persisted until glucose was depleted between 10 and 30 hours, and a secondary, subdued growth phase that ensued between 30 to 40 hours as the culture began propionate consumption. This growth eventually plateaued into a stationary phase that lasted until the end of the experiment at the 50-hour mark, with the final biomass concentration observed at 4.15 gDCW/L.

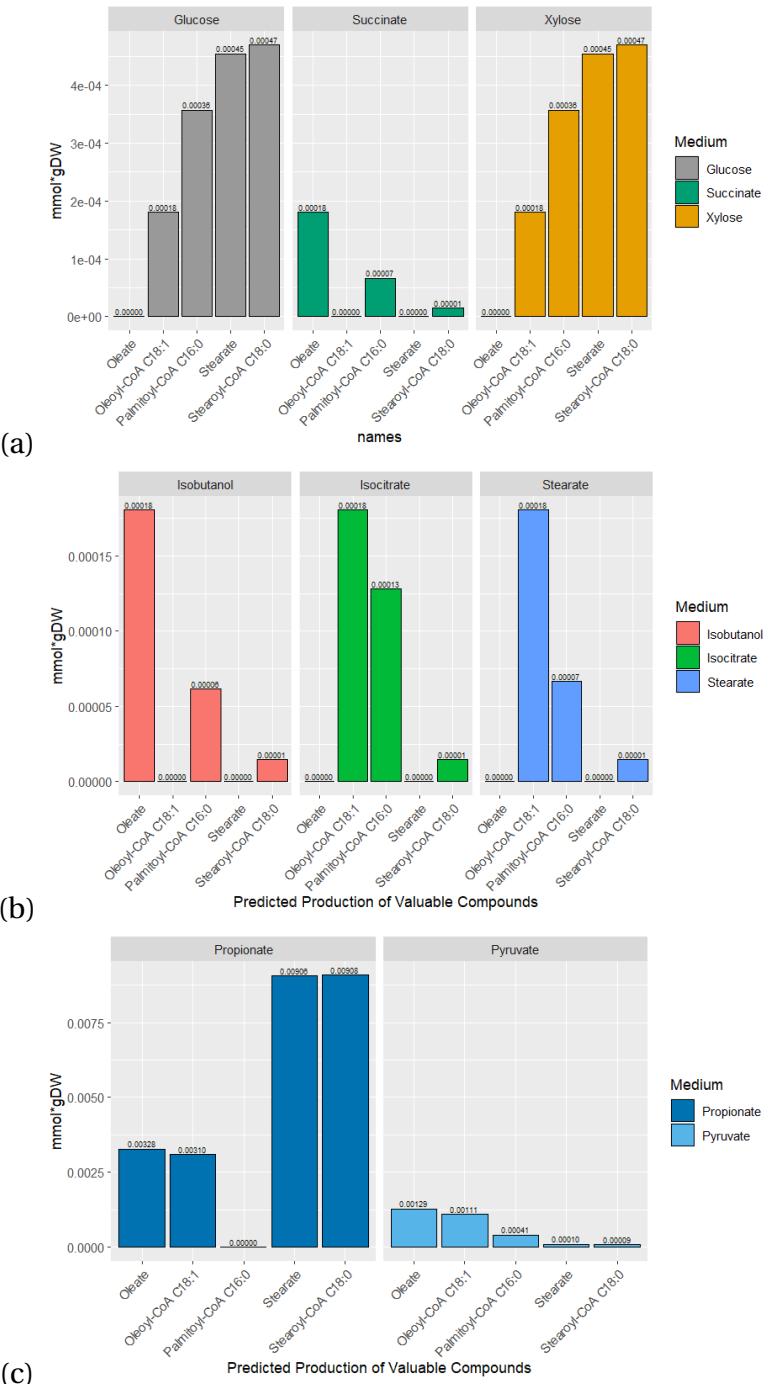


FIGURA 4.6: Predicted production of Oleate, Oleoyl-CoA, Palmitoyl-CoA, Stearate, and Stearyl-CoA under different medium conditions. (a) Only Glucose, Glucose-Succinate, and Glucose-Xylose. (b) Glucose-Stearate, Glucose-Isobutanol, and Glucose-Isocitrate. (c) Glucose-Propionate, and Glucose-Pyruvate

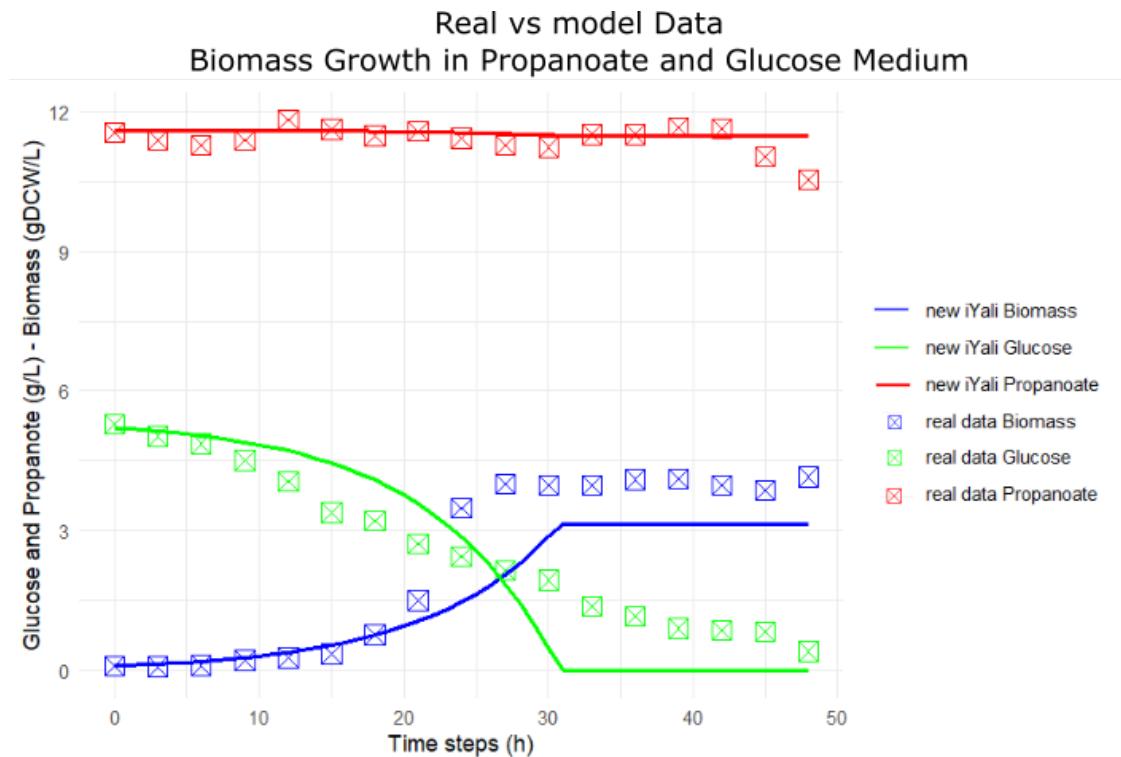


FIGURA 4.7: Laboratory data vs Predicted results of the biomass growth of *Y. lipolytica* in glucose plus propionate gel medium.

Implementing dFBA on the 'new iYali' model using the aforementioned initial concentrations enabled the simulation of these experimental findings. The uptake rates of glucose and propionate, detailed previously in Table 3.6, were minutely adjusted to improve the fit to the empirical data.

The juxtaposition of predicted and actual values for biomass growth in a propanoate and glucose medium, depicted in the scatter plot in Figure 4.8 (a), provides a vital quantitative evaluation of the model's performance. Each data point within the scatter plot corresponds to an individual observation from our dataset, differentiated by color and shape based on the respective variable - biomass, propionate, or glucose. The 45-degree line serves as a reference for ideal correspondence between predicted and actual values, with observations adhering to this line denoting instances where

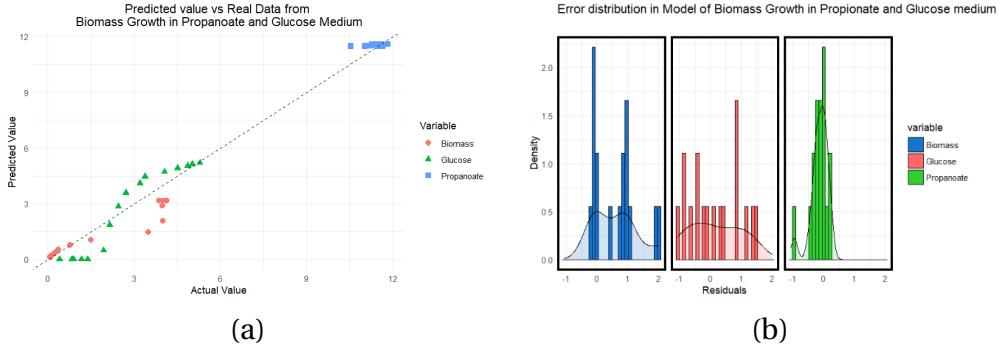


FIGURA 4.8: Statistical graphs showing the performance of the model. (a) Scatter plot of Predicted vs. Real Values(b) Error Distribution Plot.

the model flawlessly replicated empirical results. A close inspection of the scatter plot reveals a substantial alignment of data units along the reference line, particularly for propionate consumption (blue squares). This suggests a strong correlation between the predicted and actual propanoate concentrations in the medium, as opposed to the slightly larger divergence observed for biomass growth and glucose consumption.

On table 4.3 it is possible to see the display of the model's performance metrics. The r-squared values for Biomass and Glucose with figures of 0.919 and 0.918, indicate a high degree of accuracy in the model. This signifies that a large proportion of variability in our data can be accounted for by our model. On the other hand, the R² value for propanoate was relatively low at 0.109, implying that the model's ability to predict propanoate values was less precise.

Finally, the RMSE and MAE, which measure the average magnitude of prediction errors, were calculated for each variable. For biomass, the RMSE and MAE were 0.908 and 0.665, respectively; for glucose, these figures were 0.815 and 0.700; and for propanoate, they were 0.304 and 0.210. While these figures highlight some discrepancies between the model predictions and actual values, they remain relatively small in relation to the range of values observed, suggesting that the model performs fairly well in generating predictions that are reasonably close to the actual observations.

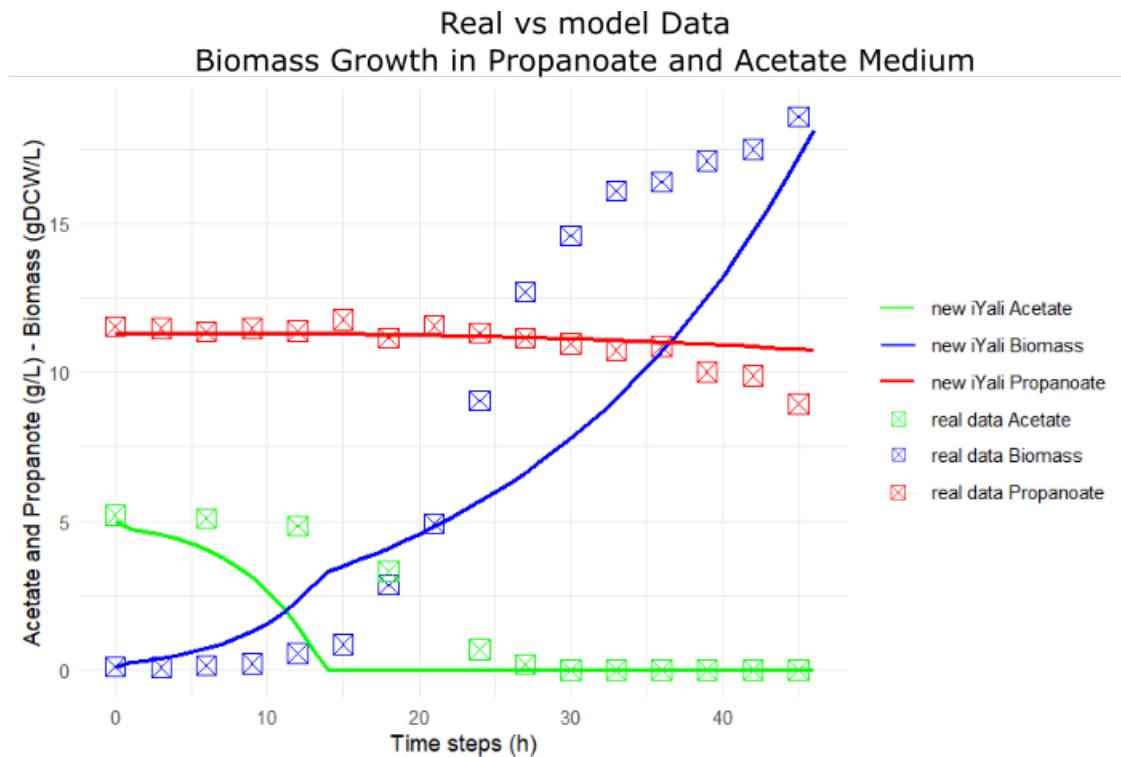


FIGURA 4.9: Laboratory data vs Predicted results of the biomass growth of *Y. lipolytica* in acetate plus propionate gel medium.

Variable	R2	RMSE	MAE
Biomass	0.919380972537853	0.908563069231937	0.665272162044825
Glucose	0.918379765591381	0.8159359969699968	0.700499272430421
Propanoate	0.109170166979195	0.304422757370607	0.210529400938935

TABLA 4.2: List of Performance Indicators of a Predictive Model

4.7.2 Biomass Growth in Propanoate and Acetate Medium

In the subsequent experiment, we turned our focus towards the exploration of *Y. lipolytica*'s growth in an environment that incorporates propionate and acetate. The

gel medium used was specifically chosen to simulate a setting akin to bioprocess waste, thus offering insights into the yeast's potential performance in such a context.

As depicted in Figure 4.9, the modelled and actual experimental data exhibit certain dissimilarities. The initial concentrations of propionate and acetate were noted to be 11.534 mg/L and 5.205 mg/L respectively. These initial quantities were subsequently utilized in configuring the dynamic Flux Balance Analysis (dFBA) algorithm for our model.

Interestingly, the growth curve for this experiment displayed two distinct phases as opposed to the three observed in the preceding experiment. The initial phase was marked by the absorption of acetate from the 5th to the 25th hour, while the latter phase was characterized by a second exponential growth period driven by propionate consumption from the 25th hour until the conclusion of the experiment. Notably, the yeast did not experience a lag phase within the experiment's timeframe.

In contrast, the new iYali model portrayed two growth phases that occurred over differing timeframes. The first phase ceased more rapidly, from the 5th to 13th hour, while the second phase extended from the 13th hour until the experiment's termination at the 45th hour. Despite these discrepancies, it is important to acknowledge that the final biomass growth approximated in both the real and modelled data was around 17-18 gDCW/L, indicating a robust potential for this substrate combination in *Y. lipolytica* medium.

Our scatter plot in Figure 4.10 (a) presents a visual comparison of the real versus predicted values, preserving the trend of higher variance in biomass and acetate consumption compared to propionate, similarly to the previous scenario. Also, in Figure 4.10 (b), it is possible to see that the residual distribution of propanoate and acetate had a normalize distribution.

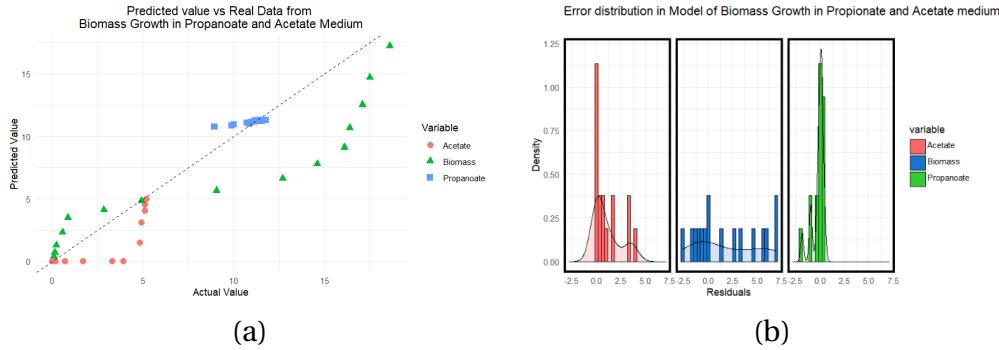


FIGURA 4.10: Statistical graphs showing the performance of the model. (a) Scatter plot of Predicted vs. Real Values (b) Error Distribution Plot.

The R^2 values, as detailed in Table K, suggest a high degree of model accuracy for both biomass and propanoate, with respective values of 0.874 and 0.911. However, a slightly lower value of 0.650 was observed for acetate, indicating a potential limitation in accurately modelling this compound's consumption. Further, the calculated RMSE and MAE values were 3.71 and 2.84 for biomass, around 1.69 and 1.05 for acetate, and approximately 0.61 and 0.39 for propanoate. Despite these figures, the overall interpretation affirms the model's success in effectively predicting the organism's real-life behavior.

In conclusion, our model presents a competent performance in simulating the *Y. lipolytica* growth and substrate consumption under the provided conditions. Future refinements may revolve around enhancing the accuracy of acetate consumption predictions, thereby further bolstering the model's overall effectiveness.

Variable	R^2	RMSE	MAE
Biomass	0.873965765643247	3.7067526145362	2.83569032395332
Acetate	0.64963517576278	1.68735889524694	1.04679040712518
Propanoate	0.910887929965844	0.60641998690826	0.389761221138856

TABLA 4.3: List of Performance Indicators of a Predictive Model

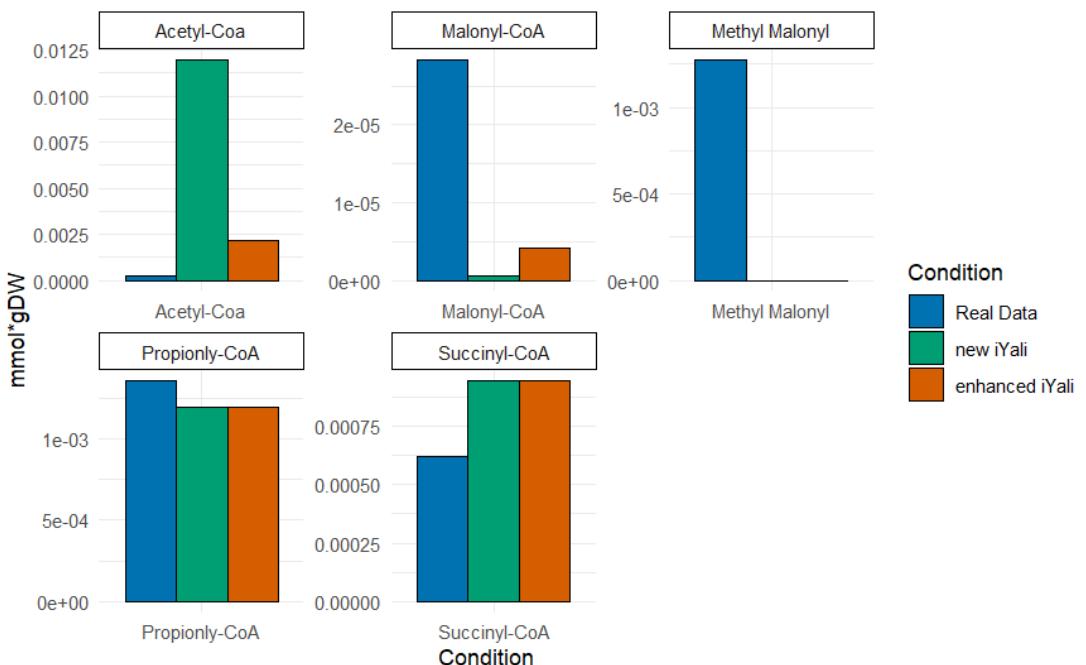


FIGURA 4.11: Laboratory data vs Predicted results of new iYali and enhanced new iYali model for the production of valuable compounds: Acetyl-CoA, Malonyl-CoA, Methyl Malonyl, Propionyl-CoA, and Succinyl-CoA.

4.8 Real Data vs Modelled Production of Valuable Compounds

In order look deeper into the predictive capacity of our new iYali model, we endeavored to analyze the production of crucial metabolites including Acetyl-CoA, Malonyl-CoA, Methyl Malonyl, Propionyl-CoA, and Succinyl-CoA. As presented in Figure 4.11, three distinct bars encapsulate the comparative representation of these metabolites. The dark blue bar delineates the empirical in-vitro production of each metabolite, whereas the green and dark orange bars depict the outputs from the new iYali model and the enhanced iYali model, respectively. The data for the models were procured utilizing the optimize() function of the COBRApy toolkit.

Upon analysing the diagram, we discover that the new iYali model exhibits a credible resemblance in the production of Propionyl-CoA and Succinyl-CoA to the in-vitro

data. However, the simulation fails to adequately capture the production of Acetyl-CoA, Malonyl-CoA, and Methyl Malonyl. The disparity is particularly salient for Malonyl-CoA and Methyl Malonyl, where the in-vitro results markedly surpass the projected outputs from both models.

This discrepancy signals a potential shortfall in the model's capacity to accurately replicate the production network for these metabolites. It implies the necessity for further refinement of the model to better mirror the organism's metabolic behaviour and enhance its predictive capability.

4.9 Real Data vs Modelled Production of Lipids and OCFAs

We have conducted a comparative study focusing on the percentage production of total lipids, Odd-Chain Fatty Acids (OCFAs), and Even-Chain Fatty Acids (ECFAs). It is imperative to underline that while the focus is not placed on the precise quantity of metabolites produced, our aim lies in appraising the proportionate behavior of the model to ascertain its congruence with empirical findings.

Figure 4.12 (a) delineates the relationship between the percentage of total lipids produced and the final biomass concentration. Upon close examination, the empirical data significantly surpasses the simulated output by an approximate 5 percent units, with empirical and modeled lipid production percentages standing at 6.87% and 1% respectively. While the enhanced new iYali model presented a slight improvement, attaining 3% of total lipids produced, it still fell short of the empirical data.

However, a more encouraging outcome can be observed in Figure 4.12 (b), wherein the new iYali model showcased superior performance in predicting OCFA production. With a prediction of 12% OCFAs in total lipid content, it was only off by a relatively slight 2.135 percentage units from the empirical result of 9.865 %. In stark contrast, the

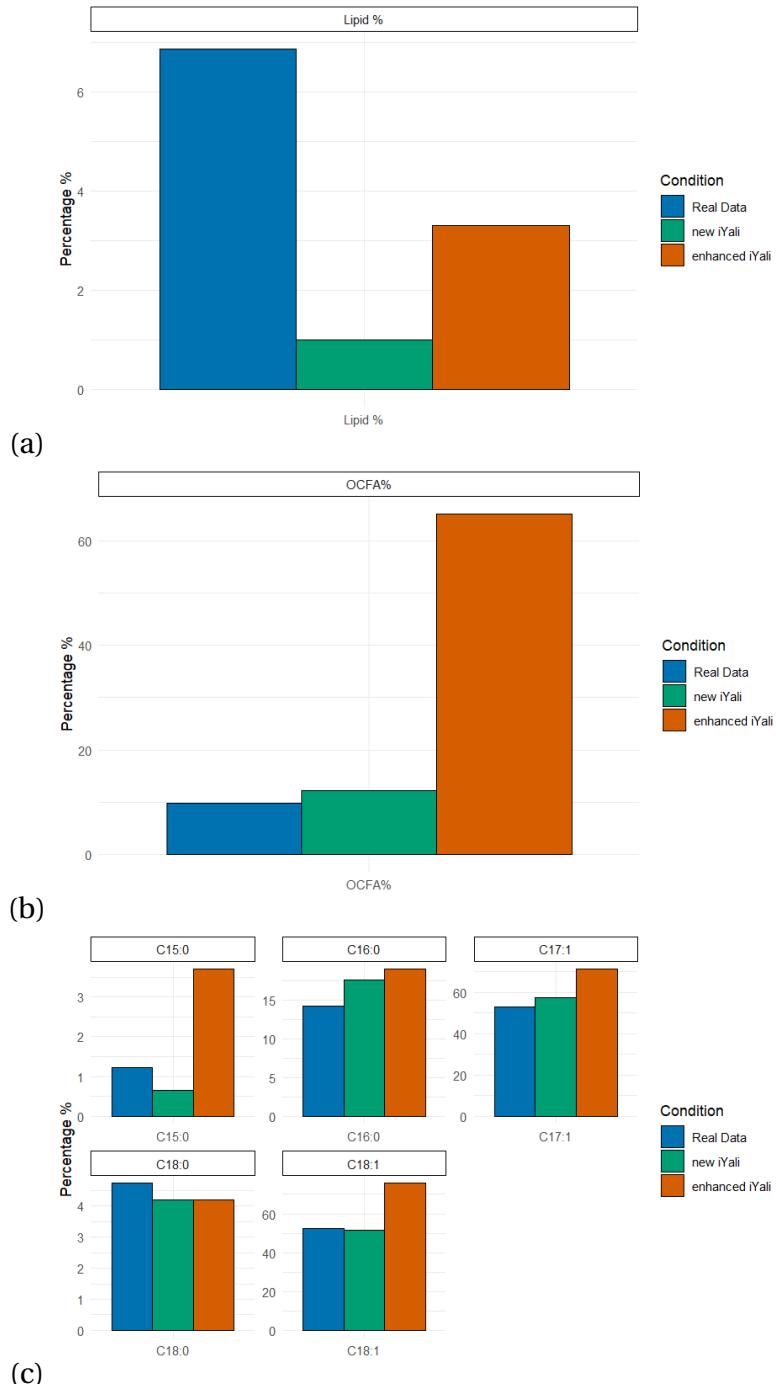


FIGURA 4.12: In-Vitro vs Predicted percentage production of Lipids and OCFA. (a) % of total lipids production vs final biomass. (b) % of total OCFA produced vs total lipids production. (c) % of different types of OCFA and ECFA production vs total lipids production

enhanced new iYali model yielded a highly divergent prediction, estimating a 65 % OC-FA production. Such findings reinforce our conviction that, although our model may not yet attain flawless accuracy, it undoubtedly possesses the capability to offer insightful predictions regarding the production capacity of units *Y. lipolytica* in relation to OCFAs.

In Figure 4.12 (c), we delve into the specific percentages of each OCFA and ECFA against the total lipid content. At first glance, the striking similarity between the new iYali and the empirical data percentages becomes apparent, further solidifying our belief in the model's competency in predicting OCFA and ECFA production. In contrast, the enhanced new iYali model displayed an amplified production of C17:1 OCFA or margaric acid, coupled with a decrease in ECFA and C15:0 OCFA production. These findings suggest that the simulated genetic modifications introduced into the model and the intrinsic enzymatic network of units *Y. lipolytica* provide a conducive environment for enhanced production of C17:1.

CONCLUSIONES

The prevailing exigency for sustainable alternatives to fossil fuels and industrial compounds has ushered the use of *Y. lipolytica* to the forefront as a promising chassis for sustainable biofuel production. However, the translation of this organism's potential into industrial-scale applications has been hampered by the high developmental costs associated with engineering new phenotypes capable of optimized biofuel production. This research thus set out to bridge this gap by harnessing Flux Balance Analysis (FBA) and Dynamic Flux Balance Analysis (dFBA) to optimize a refined phenotype of *Y. lipolytica*, with an eye on enhancing the production of Odd Chain Fatty Acids (OCFAs), specifically C15 and C17.

This thesis marks a significant stride in the ongoing journey to engineer improved phenotypes of *Y. lipolytica*. Our efforts culminated in the completion of the iYali4 model, a comprehensive representation that incorporates 68 metabolites and 36 reactions pivotal to propionate metabolism and OCFA production. This refined model presents a valuable tool for simulating the biosynthesis of highly sought-after compounds, thus making a meaningful contribution to the bioprocess industry.

Utilizing the COBRApY toolkit, alongside FBA and dFBA methodologies, we successfully simulated a variety of genetic modifications within the new iYali model. The enhancements centred on the overexpression of key enzymes facilitating propionate metabolism towards acetate and OCFA production. Concurrently, the knockdown of fatty acid transporters directed towards the peroxisome ensures the effective accumulation of OCFAs and Even Chain Fatty Acids (ECFAs) within the yeast cells.

We subsequently benchmarked the performance of our model against in-vitro

laboratory data. This comparative analysis facilitated the tuning of various model parameters, thereby aligning our in-silico predictions closer to the actual behaviour exhibited by *Y. lipolytica* in varied conditions. Moreover, our model demonstrated a commendable ability to simulate the organism's metabolic responses with a tolerable degree of precision.

The culmination of our research involved an evaluation of the enhanced iYali model's capability to produce OCFAs and other high-value compounds. Our findings suggest that the genetic modifications yielded a 20

In essence, this thesis has delivered an enhanced version of *Y. lipolytica*, genetically fine-tuned to utilize alternative carbon sources for sustainable bioproduction. Moreover, our in-silico phenotype demonstrated optimized production of OCFAs, along with other high-value metabolites such as ECFAs, Methyl Malonyl, and Citrate. These strides represent a promising step towards the realization of sustainable biofuel production, as we continue to fine-tune nature's machinery to address our energy needs.

RECOMENDACIONES

In the light of the research findings elucidated in this thesis, we propose a series of recommendations for future investigations in the realm of biofuel production leveraging *Y. lipolytica*.

Primarily, we propose that the experimentally validate the enhanced new iYali model. A laboratory-based verification of the model's predictions is of paramount importance in ascertaining its efficacy and the authenticity of its simulation outcomes. This will entail growing the organism under conditions identical to those modelled, with subsequent analysis of the metabolic products and their comparison with the model's predictions.

Additionally, we propose a transition from the traditional trial-and-error method of phenotype discovery, to a more sophisticated, algorithmic approach. A metaheuristic algorithm could serve as a robust tool for exploring the genetic modification space, enabling the identification of promising new phenotypes with greater accuracy and efficiency. By employing an algorithmic approach, we stand to capitalize on computational power to streamline and enhance the genetic optimization process, ultimately saving significant time and resources.

Finally, we advocate for the development of an intuitive visualization tool, capable of dynamically representing the simulation data. A graphical user interface that

allows researchers to visually follow the metabolic changes in the simulated *Y. lipolytica* could serve as a powerful tool for data interpretation and analysis. Such a tool could greatly facilitate comprehension of the complex metabolic interplay occurring within *Y. lipolytica* during the production of biomass and high-value compounds.

These recommendations, if enacted, stand to significantly augment our ability to understand and manipulate *Y. lipolytica* for biofuel production, paving the way for new breakthroughs in this burgeoning field of research.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Y. kyoung Park, R. Ledesma-Amaro, and J.-M. Nicaud, “De novo biosynthesis of odd-chain fatty acids in *yarrowia lipolytica* enabled by modular pathway engineering,” *Frontiers in Bioengineering and Biotechnology*, vol. 7, 1 2020.
- [2] W. Alt, A. Deutsch, and L. Preziosi, “Computational cell biology: Second theme issue on “computational biology”,” *Journal of Mathematical Biology*, vol. 58, pp. 1–5, 1 2009.
- [3] Z. Liu, Z. Deng, S. Davis, and P. Ciais, “Monitoring global carbon emissions in 2022,” *Nature Reviews Earth Environment*, vol. 4, pp. 205–206, 3 2023.
- [4] W. Soetaert and E. Vandamme, “The impact of industrial biotechnology,” *Bio-technology Journal*, vol. 1, pp. 756–769, 8 2006.
- [5] J. Schellenberger, R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. Palsson, “Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2.0,” *Nature Protocols*, vol. 6, pp. 1290–1307, 9 2011.
- [6] A. Varma and B. O. Palsson, “Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *escherichia coli* w3110,” *Applied and Environmental Microbiology*, vol. 60, pp. 3724–3731, 10 1994.

- [7] R. Mahadevan, J. S. Edwards, and F. J. Doyle, “Dynamic flux balance analysis of diauxic growth in escherichia coli,” *Biophysical Journal*, vol. 83, pp. 1331–1340, 9 2002.
- [8] J. D. Orth, I. Thiele, and B. Palsson, “What is flux balance analysis?” *Nature Biotechnology*, vol. 28, pp. 245–248, 3 2010.
- [9] F. Darvishi, Z. Fathi, M. Ariana, and H. Moradi, “Yarrowia lipolytica as a workhorse for biofuel production,” *Biochemical Engineering Journal*, vol. 127, pp. 87–96, 11 2017.
- [10] P. Fickers, J. Destain, and P. Thonart, “Improvement of *< i>yarrowia lipolytica</i>* lipase production by fed-batch fermentation,” *Journal of Basic Microbiology*, vol. 49, pp. 212–215, 4 2009.
- [11] T. J. Hanly and M. A. Henson, “Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures,” *Biootechnology and Bioengineering*, vol. 108, pp. 376–385, 2 2011.
- [12] A. P. Burgard, P. Pharkya, and C. D. Maranas, “Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization,” *Biotechnology and Bioengineering*, vol. 84, pp. 647–657, 12 2003.
- [13] M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah, “Chapter 10 - metaheuristic algorithms: A comprehensive review,” in *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*, ser. Intelligent Data-Centric Systems, A. K. Sangaiah, M. Sheng, and Z. Zhang, Eds. Academic Press, 2018, pp. 185–231. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128133149000104>
- [14] J. S. Edwards and B. O. Palsson, “Systems properties of the haemophilus influenzae metabolic genotype,” *Journal of Biological Chemistry*, vol. 274, pp. 17410–17416, 6 1999.

- [15] R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Ker-lavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. F. , C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L. Ing Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghegan, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter, “Whole-genome random sequencing and assembly of *< i>haemophilus influenzae</i>* rd,” *Science*, vol. 269, pp. 496–512, 7 1995.
- [16] B. Trivedi, “Modeling the oddities of biology,” *Nature Biotechnology*, vol. 16, pp. 1316–1317, 12 1998.
- [17] E. Brunk, S. Sahoo, D. C. Zielinski, A. Altunkaya, A. Dräger, N. Mih, F. Gatto, A. Nilsson, G. A. P. Gonzalez, M. K. Aurich, A. Prlić, A. Sastry, A. D. Danielsdot-tir, A. Heinken, A. Noronha, P. W. Rose, S. K. Burley, R. M. T. Fleming, J. Nielsen, I. Thiele, and B. O. Palsson, “Recon3d enables a three-dimensional view of gene variation in human metabolism,” *Nature Biotechnology*, vol. 36, pp. 272–281, 3 2018.
- [18] P. Pan and Q. Hua, “Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, *yarrowia lipolytica*,” *PLoS ONE*, vol. 7, p. e51535, 12 2012.
- [19] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi, “Kegg for linking geno-mes to life and the environment,” *Nucleic Acids Research*, vol. 36, pp. D480–D484, 12 2007.
- [20] J. Schellenberger, J. O. Park, T. M. Conrad, and B. Palsson, “Bigg: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions,” *BMC Bioinformatics*, vol. 11, p. 213, 12 2010.

- [21] E. J. Kerkhoven, K. R. Pomraning, S. E. Baker, and J. Nielsen, “Regulation of amino-acid metabolism controls flux to lipid accumulation in *yarrowia lipolytica*,” *npj Systems Biology and Applications*, vol. 2, p. 16005, 3 2016.
- [22] N. Loira, T. Dulermo, J.-M. Nicaud, and D. J. Sherman, “A genome-scale metabolic model of the lipid-accumulating yeast *yarrowia lipolytica*,” *BMC Systems Biology*, vol. 6, p. 35, 12 2012.
- [23] D. J. Sherman, T. Martin, M. Nikolski, C. Cayla, J.-L. Souciet, and P. Durrens, “Génolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes,” *Nucleic Acids Research*, vol. 37, pp. D550–D554, 1 2009.
- [24] M. Nikolski and D. J. Sherman, “Family relationships: should consensus reign?—consensus clustering for protein families,” *Bioinformatics*, vol. 23, pp. e71–e76, 1 2007.
- [25] N. Morin, J. Cescut, A. Beopoulos, G. Lelandais, V. L. Berre, J.-L. Uribelarrea, C. Molina-Jouve, and J.-M. Nicaud, “Transcriptomic analyses during the transition from biomass production to lipid accumulation in the oleaginous yeast *yarrowia lipolytica*,” *PLoS ONE*, vol. 6, p. e27966, 11 2011.
- [26] K. Qiao, S. H. I. Abidi, H. Liu, H. Zhang, S. Chakraborty, N. Watson, P. K. Ajikumar, and G. Stephanopoulos, “Engineering lipid overproduction in the oleaginous yeast *yarrowia lipolytica*,” *Metabolic Engineering*, vol. 29, pp. 56–65, 5 2015.
- [27] M. Abedi, R. Fatehi, K. Moradzadeh, and Y. Gheisari, “Big data to knowledge: common pitfalls in transcriptomics data analysis and representation,” *RNA Biology*, vol. 16, pp. 1531–1533, 11 2019.
- [28] G. Betts and D. Srivastava, “The rationalization of high enzyme concentration in metabolic pathways such as glycolysis,” *Journal of Theoretical Biology*, vol. 151, pp. 155–167, 7 1991.

- [29] “Kegg pathway: Metabolic pathways - yarrowia lipolytica,” <https://www.genome.jp/pathway/yli01100+M00916>, accessed: 2023-05-29.
- [30] J. da Veiga Moreira, M. Jolicœur, L. Schwartz, and S. Peres, “Fine-tuning mitochondrial activity in yarrowia lipolytica for citrate overproduction,” *Scientific Reports*, vol. 11, p. 878, 1 2021.
- [31] S. Guerrero-Castillo, M. Vázquez-Acevedo, D. González-Halphen, and S. Uribe-Carvajal, “In yarrowia lipolytica mitochondria, the alternative nadh dehydrogenase interacts specifically with the cytochrome complexes of the classic respiratory pathway,” *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1787, pp. 75–85, 2 2009.
- [32] K. M. Shaikh and A. A. Odaneth, “Metabolic engineering of yarrowia lipolytica for the production of isoprene,” *Biotechnology Progress*, vol. 37, no. 6, p. e3201, 2021. [Online]. Available: <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/btpr.3201>
- [33] E. Carsamba, S. Papanikolaou, P. Fickers, and H. Erten, “Lipids by yarrowia lipolytica strains cultivated on glucose in batch cultures,” *Microorganisms*, vol. 8, p. 1054, 7 2020.
- [34] H. J. Wang, M.-T. L. Dall, Y. Wache, C. Laroche, J.-M. Belin, C. Gaillardin, and J.-M. Nicaud, “Evaluation of acyl coenzyme a oxidase (aox) isozyme function in the *n*-alkane-assimilating yeast *yarrowia lipolytica*,” *Journal of Bacteriology*, vol. 181, pp. 5140–5148, 9 1999.
- [35] M. Apaydin, B. Zeng, and X. Qian, “A reliable alternative of optknock for desirable mutant microbial strains.” IEEE, 2 2016, pp. 573–576.
- [36] J. J. Czajka, T. Oyetunde, and Y. J. Tang, “Integrated knowledge mining, genome-scale modeling, and machine learning for predicting yarrowia lipolytica bioproduction,” *Metabolic Engineering*, vol. 67, pp. 227–236, 9 2021.

- [37] J. J. Czajka, B. Okumuş, M. A. Koffas, M. Blenner, and Y. J. Tang, “Mitigation of host cell mutations and regime shift during microbial fermentation: a perspective from flux memory,” *Current Opinion in Biotechnology*, vol. 66, pp. 227–235, 12 2020.
- [38] X. Guo, J. Sun, D. Li, and W. Lu, “Heterologous biosynthesis of (+)-nootkatone in unconventional yeast *yarrowia lipolytica*,” *Biochemical Engineering Journal*, vol. 137, pp. 125–131, 9 2018.
- [39] K. Papouskova and H. Sychrova, “*yarrowia lipolytica* possesses two plasma membrane alkali metal cation/h ⁺ antiporters with different functions in cell physiology,” *FEBS Letters*, vol. 580, pp. 1971–1976, 4 2006.
- [40] P. R. Ellis, M. J. Hayes, N. Macleod, S. J. Schuyten, C. L. Tway, and C. M. Zalitis, “Carbon conversion: opportunities in chemical productions,” pp. 479–524, 2023.
- [41] J. Ruiz-Herrera, P. García-Maceira, L. C. Castillo-Barahona, E. Valentín, and R. Sentandreu, “Cell wall composition and structure of *yarrowia lipolytica* transposon mutants affected in calcofluor sensitivity,” *Antonie van Leeuwenhoek*, vol. 84, pp. 229–238, 2003.
- [42] H. Pereira, F. Azevedo, L. Domingues, and B. Johansson, “Expression of *yarrowia lipolytica* acetyl-coa carboxylase in *saccharomyces cerevisiae* and its effect on in-vivo accumulation of malonyl-coa,” *Computational and Structural Biotechnology Journal*, vol. 20, pp. 779–787, 2022.
- [43] M. Papagianni, “Metabolic engineering of lactic acid bacteria for the production of industrially important compounds,” *Computational and Structural Biotechnology Journal*, vol. 3, p. e201210003, 10 2012.
- [44] J. Peretó, “Embden-meyerhof-parnas pathway,” pp. 485–485, 2011.

- [45] L.-B. Yang, X.-B. Zhan, Z.-Y. Zheng, J.-R. Wu, M.-J. Gao, and C.-C. Lin, “A novel osmotic pressure control fed-batch fermentation strategy for improvement of erythritol production by *yarrowia lipolytica* from glycerol,” *Bioresource Technology*, vol. 151, pp. 120–127, 1 2014.
- [46] G. P. da Silva, M. Mack, and J. Contiero, “Glycerol: A promising and abundant carbon source for industrial microbiology,” *Biotechnology Advances*, vol. 27, pp. 30–39, 1 2009.
- [47] Y. Seif and B. Ørn Palsson, “Path to improving the life cycle and quality of genome-scale models of metabolism,” *Cell Systems*, vol. 12, pp. 842–859, 9 2021.
- [48] J. J. Zimmerman, A. von Saint André-von Arnim, and J. McLaughlin, “Cellular respiration,” pp. 1058–1072, 2011.
- [49] G. Napolitano, G. Fasciolo, S. D. Meo, and P. Venditti, “Mitochondrial redox biology: Reactive species production and antioxidant defenses,” pp. 105–125, 2021.
- [50] J. D. Enderle, “Biochemical reactions and enzyme kinetics,” pp. 447–508, 2012.
- [51] R. Gonzalez-Garcia, T. McCubbin, L. Navone, C. Stowers, L. Nielsen, and E. Marcellin, “Microbial propionic acid production,” *Fermentation*, vol. 3, p. 21, 5 2017.
- [52] Y.-K. Park, F. Bordes, F. Letisse, and J.-M. Nicaud, “Engineering precursor pools for increasing production of odd-chain fatty acids in *yarrowia lipolytica*,” *Metabolic Engineering Communications*, vol. 12, p. e00158, 6 2021.
- [53] C. P. Kurtzman, J. W. Fell, and T. Boekhout, *The Yeasts*, 5th ed. Elsevier, 2010.
- [54] B. Dujon, D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. de Montigny, C. Marck, C. Neuvéglise, E. Talla, N. Goffard, L. Frangeul, M. Aigle, V. Anthouard, A. Babour, V. Barbe, S. Barnay, S. Blanchin, J.-M. Beckerich, E. Beyne, C. Bleykasten, A. Boisramé, J. Boyer, L. Cattolico, F. Confaniolieri, A. de Daruvar, L. Despons, E. Fabre, C. Fairhead, H. Ferry-Dumazet, A. Groppi, F. Hantraye,

C. Hennequin, N. Jauniaux, P. Joyet, R. Kachouri, A. Kerrest, R. Koszul, M. Lemaire, I. Lesur, L. Ma, H. Muller, J.-M. Nicaud, M. Nikolski, S. Oztas, O. Ozier-Kalogeropoulos, S. Pellenz, S. Potier, G.-F. Richard, M.-L. Straub, A. Suleau, D. Swennen, F. Tekaia, M. Wésolowski-Louvel, E. Westhof, B. Wirth, M. Zeniou-Meyer, I. Zivanovic, M. Bolotin-Fukuhara, A. Thierry, C. Bouchier, B. Caudron, C. Scarpelli, C. Gaillardin, J. Weissenbach, P. Wincker, and J.-L. Souciet, “Genome evolution in yeasts,” *Nature*, vol. 430, pp. 35–44, 7 2004.

- [55] C. Navarrete and J. L. Martínez, “Non-conventional yeasts as superior production platforms for sustainable fermentation based bio-manufacturing processes,” *AIMS Bioengineering*, vol. 7, pp. 289–305, 2020.
- [56] J.-M. Nicaud, “*< i>yarrowia lipolytica</i>*,” *Yeast*, vol. 29, pp. 409–418, 10 2012.
- [57] G. Benard, N. Bellance, C. Jose, and R. Rossignol, “Relationships between mitochondrial dynamics and bioenergetics,” pp. 47–68, 2011.
- [58] K. Mlickova, E. Roux, K. Athenstaedt, S. d’Andrea, G. Daum, T. Chardot, and J.-M. Nicaud, “Lipid accumulation, lipid body formation, and acyl coenzyme a oxidases of the yeast *< i>yarrowia lipolytica</i>*,” *Applied and Environmental Microbiology*, vol. 70, pp. 3918–3924, 7 2004.
- [59] K. MLIKOVA, “Acyl-coa oxidase, a key step for lipid accumulation in the yeast *yarrowia lipolytica*,” *Journal of Molecular Catalysis B: Enzymatic*, vol. 28, pp. 81–85, 5 2004.
- [60] Y.-K. Park, T. Dulermo, R. Ledesma-Amaro, and J.-M. Nicaud, “Optimization of odd chain fatty acid production by *yarrowia lipolytica*,” *Biotechnology for Biofuels*, vol. 11, p. 158, 12 2018.
- [61] M. Larroude, T. Rossignol, J.-M. Nicaud, and R. Ledesma-Amaro, “Synthetic biology tools for engineering *yarrowia lipolytica*,” *Biotechnology Advances*, vol. 36, pp. 2150–2164, 12 2018.

- [62] S. Lajus, S. Dusséaux, J. Verbeke, C. Rigouin, Z. Guo, M. Fatarova, F. Bellvert, V. Borsenberger, M. Bressy, J.-M. Nicaud, A. Marty, and F. Bordes, “Engineering the yeast *yarrowia lipolytica* for production of polylactic acid homopolymer,” *Frontiers in Bioengineering and Biotechnology*, vol. 8, 10 2020.
- [63] N. C. for Biotechnology Information, “Pentadecanoic acid,” PubChem Compound Database, CID=8181, USA, 2023, [Online]. Available: <https://pubchem.ncbi.nlm.nih.gov/compound/Pentadecanoic-acid> [Accessed: May. 17, 2023].
- [64] ——, “Heptadecanoic acid,” PubChem Compound Database, CID=8188, USA, 2023, [Online]. Available: <https://pubchem.ncbi.nlm.nih.gov/compound/Heptadecanoic-acid> [Accessed: May. 17, 2023].
- [65] ——, “Nonadecanoic acid,” PubChem Compound Database, CID=8210, USA, 2023, [Online]. Available: <https://pubchem.ncbi.nlm.nih.gov/compound/Nonadecanoic-acid> [Accessed: May. 17, 2023].
- [66] L. I. Hellgren and P. Nordby, “Bioactive lipids in dairy fat,” pp. 233–237, 2017.
- [67] W. A. Sahyouni, S. E. Kantar, A. Khelfa, Y.-K. Park, J.-M. Nicaud, N. Louka, and M. Koubaa, “Optimization of cis-9-heptadecenoic acid production from the oleaginous yeast *yarrowia lipolytica*,” *Fermentation*, vol. 8, p. 245, 5 2022.
- [68] H. K. Pedersen, V. Gudmundsdottir, H. B. Nielsen, T. Hyotylainen, T. Nielsen, B. A. H. Jensen, K. Forslund, F. Hildebrand, E. Prifti, G. Falony, E. L. Chatelier, F. Levenez, J. Doré, I. Mattila, D. R. Plichta, P. Pöhö, L. I. Hellgren, M. Arumugam, S. Sunagawa, S. Vieira-Silva, T. Jørgensen, J. B. Holm, K. Trošt, M. Consortium, K. Kristiansen, S. Brix, J. Raes, J. Wang, T. Hansen, P. Bork, S. Brunak, M. Oresic, S. D. Ehrlich, and O. Pedersen, “Human gut microbes impact host serum metabolome and insulin sensitivity,” *Nature*, vol. 535, pp. 376–381, 7 2016.

- [69] Z. Chen, L. Wang, S. Qiu, and S. Ge, “Determination of microalgal lipid content and fatty acid for biofuel production,” *BioMed Research International*, vol. 2018, pp. 1–17, 2018.
- [70] A. Beopoulos, J. Cescut, R. Haddouche, J.-L. Uribelarrea, C. Molina-Jouve, and J.-M. Nicaud, “Yarrowia lipolytica as a model for bio-oil production,” *Progress in Lipid Research*, vol. 48, pp. 375–387, 11 2009.
- [71] Q. Fei, H. N. Chang, L. Shang, J. dal-rae Choi, N. Kim, and J. Kang, “The effect of volatile fatty acids as a sole carbon source on lipid accumulation by *Cryptococcus albidus* for biodiesel production,” *Bioresource Technology*, vol. 102, pp. 2695–2701, 2 2011.
- [72] K. Kolter, A. Dashevsky, M. Irfan, and R. Bodmeier, “Polyvinyl acetate-based film coatings,” *International Journal of Pharmaceutics*, vol. 457, pp. 470–479, 12 2013.
- [73] L. O. Ingram, L. S. Chevalier, E. J. Gabba, K. D. Ley, and K. Winters, “Propionate-induced synthesis of odd-chain-length fatty acids by *Escherichia coli*,” *Journal of Bacteriology*, vol. 131, pp. 1023–1025, 9 1977.
- [74] P. Fontanille, V. Kumar, G. Christophe, R. Nouaille, and C. Larroche, “Bioconversion of volatile fatty acids into lipids by the oleaginous yeast *yarrowia lipolytica*,” *Bioresource Technology*, vol. 114, pp. 443–449, 6 2012.
- [75] H. Lu, F. Li, B. J. Sánchez, Z. Zhu, G. Li, I. Domenzain, S. Marciauskas, P. M. Anton, D. Lappa, C. Lieven, M. E. Beber, N. Sonnenschein, E. J. Kerkhoven, and J. Nielsen, “A consensus *S. cerevisiae* metabolic model yeast8 and its ecosystem for comprehensively probing cellular metabolism,” *Nature Communications*, vol. 10, p. 3586, 8 2019.
- [76] K. Murota, “Linear programming,” pp. 1–7, 2020.

- [77] L. Heirendt, S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdóttir, J. Wachowiak, S. M. Keating, V. Vlasov, S. Magnusdóttir, C. Y. Ng, G. Preciat, A. Žagare, S. H. J. Chan, M. K. Aurich, C. M. Clancy, J. Modamio, J. T. Sauls, A. Noronha, A. Bordbar, B. Cousins, D. C. E. Assal, L. V. Valcarcel, I. Apaola-Zaza, S. Ghaderi, M. Ahookhosh, M. B. Guebila, A. Kostromins, N. Sompairac, H. M. Le, D. Ma, Y. Sun, L. Wang, J. T. Yurkovich, M. A. P. Oliveira, P. T. Vuong, L. P. E. Assal, I. Kuperstein, A. Zinovyev, H. S. Hinton, W. A. Bryant, F. J. A. Artacho, F. J. Planes, E. Stalidzans, A. Maass, S. Vempala, M. Hucka, M. A. Saunders, C. D. Maranas, N. E. Lewis, T. Sauter, B. Palsson, I. Thiele, and R. M. T. Fleming, “Creation and analysis of biochemical constraint-based models using the cobra toolbox v.3.0,” *Nature Protocols*, vol. 14, pp. 639–702, 3 2019.
- [78] MATLAB, *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010.
- [79] N. González-Arrué, I. Inostroza, R. Conejeros, and M. Rivas-Astroza, “Phenotype-specific estimation of metabolic fluxes using gene expression data,” *iScience*, vol. 26, p. 106201, 3 2023.
- [80] J. A. Gomez, K. Höffner, and P. I. Barton, “Dfbalab: a fast and reliable matlab code for dynamic flux balance analysis,” *BMC Bioinformatics*, vol. 15, p. 409, 12 2014.
- [81] Y. Guo, L. Su, Q. Liu, Y. Zhu, Z. Dai, and Q. Wang, “Dissecting carbon metabolism of yarrowia lipolytica type strain w29 using genome-scale metabolic modelling,” *Computational and Structural Biotechnology Journal*, vol. 20, pp. 2503–2511, 2022.
- [82] M. Kanehisa, Y. Sato, and K. Morishima, “Blastkoala and ghostkoala: Kegg tools for functional characterization of genome and metagenome sequences,” *Journal of Molecular Biology*, vol. 428, pp. 726–731, 2 2016.

- [83] P. Mishra, N.-R. Lee, M. Lakshmanan, M. Kim, B.-G. Kim, and D.-Y. Lee, “Genome-scale model-driven strain design for dicarboxylic acid production in *yarrowia lipolytica*,” *BMC Systems Biology*, vol. 12, p. 12, 3 2018.
- [84] M. Kavšček, G. Bhutada, T. Madl, and K. Natter, “Optimization of lipid production with a genome-scale model of *yarrowia lipolytica*,” *BMC Systems Biology*, vol. 9, p. 72, 12 2015.
- [85] S. Wei, X. Jian, J. Chen, C. Zhang, and Q. Hua, “Reconstruction of genome-scale metabolic model of *yarrowia lipolytica* and its application in overproduction of triacylglycerol,” *Bioresources and Bioprocessing*, vol. 4, p. 51, 12 2017.
- [86] C. Ma, Q. Mu, Y. Xue, Y. Xue, B. Yu, and Y. Ma, “One major facilitator superfamily transporter is responsible for propionic acid tolerance in *< i>pseudomonas putida</i> kt2440*,” *Microbial Biotechnology*, vol. 14, pp. 386–391, 3 2021.
- [87] J. Zhou, X. Yin, C. Madzak, G. Du, and J. Chen, “Enhanced -ketoglutarate production in *yarrowia lipolytica* wsh-z06 by alteration of the acetyl-coa metabolism,” *Journal of Biotechnology*, vol. 161, pp. 257–264, 10 2012.
- [88] S. Clarke and M. Nakamura, “Fatty acid structure and synthesis,” pp. 285–289, 2013.
- [89] S. Abreu, Y.-K. Park, C. P. de Souza, L. Vidal, P. Chaminade, and J.-M. Nicaud, “Lipid readjustment in *yarrowia lipolytica* odd-chain fatty acids producing strains,” *Biomolecules*, vol. 12, p. 1026, 7 2022.
- [90] B. Christensen and J. Nielsen, “Metabolic network analysis,” pp. 209–231, 1999.
- [91] D. Machado, M. J. Herrgård, and I. Rocha, “Stoichiometric representation of gene–protein–reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction,” *PLOS Computational Biology*, vol. 12, p. e1005140, 10 2016.

- [92] D. Machado, R. S. Costa, E. C. Ferreira, I. Rocha, and B. Tidor, “Exploring the gap between dynamic and constraint-based models of metabolism,” *Metabolic Engineering*, vol. 14, pp. 112–119, 3 2012.
- [93] J. Almquist, M. Cvijovic, V. Hatzimanikatis, J. Nielsen, and M. Jirstrand, “Kinetic models in industrial biotechnology – improving cell factory performance,” *Metabolic Engineering*, vol. 24, pp. 38–60, 7 2014.
- [94] G. B. Dantzig, “Linear programming,” pp. 1–6, 2008.
- [95] M. D. Filippo, C. Damiani, and D. Pescini, “Gpruler: Metabolic gene-protein-reaction rules automatic reconstruction,” *PLOS Computational Biology*, vol. 17, p. e1009550, 11 2021.
- [96] A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke, “Cobrapy: Constraints-based reconstruction and analysis for python,” *BMC Systems Biology*, vol. 7, p. 74, 12 2013.
- [97] S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Palsson, and M. J. Herrgard, “Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox,” *Nature Protocols*, vol. 2, pp. 727–738, 3 2007.
- [98] R. H. Ng, J. W. Lee, P. Baloni, C. Diener, J. R. Heath, and Y. Su, “Constraint-based reconstruction and analyses of metabolic models: Open-source python tools and applications to cancer,” *Frontiers in Oncology*, vol. 12, 7 2022.
- [99] S. Schuster, C. Hilgetag, J. Woods, and D. Fell, “Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism,” *Journal of Mathematical Biology*, vol. 45, pp. 153–181, 8 2002.
- [100] A. Cornish-Bowden, “The origins of enzyme kinetics,” *FEBS Letters*, vol. 587, pp. 2725–2730, 9 2013.

- [101] A. Cornish-Bowden, J.-P. Mazat, and S. Nicolas, “Victor henri: 111 years of his equation,” *Biochimie*, vol. 107, pp. 161–166, 12 2014.
- [102] Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. O. Palsson, and N. E. Lewis, “Bigg models: A platform for integrating, standardizing and sharing genome-scale models,” *Nucleic Acids Research*, vol. 44, pp. D515–D522, 1 2016.
- [103] C. Li, M. Donizelli, N. Rodriguez, H. Dharuri, L. Endler, V. Chelliah, L. Li, E. He, A. Henry, M. I. Stefan, J. L. Snoep, M. Hucka, N. L. Novère, and C. Laibe, “Bio- models database: An enhanced, curated and annotated resource for published quantitative kinetic models,” *BMC Systems Biology*, vol. 4, p. 92, 12 2010.
- [104] M. Peters, J. J. Eicher, D. D. van Niekerk, D. Waltemath, and J. L. Snoep, “The jws online simulation database,” *Bioinformatics*, vol. 33, pp. 1589–1590, 5 2017.
- [105] S. Wegerhoff and S. Engell, “Control of the production of *saccharomyces cerevisiae* on the basis of a reduced metabolic model * *this research is funded by the german ministry of research an technology (bmbf) in the context of the yeas- tsent project (031a301a): “volatile metabolites as proxy for metabolic network operation of *saccharomyces cerevisiae”*,” *IFAC-PapersOnLine*, vol. 49, pp. 201–206, 2016.
- [106] J. L. Hjersted, M. A. Henson, and R. Mahadevan, “Genome-scale analysis of *sac- charomyces cerevisiae* metabolism and ethanol production in fed-batch cultu- re,” *Biotechnology and Bioengineering*, vol. 97, pp. 1190–1204, 8 2007.
- [107] S. M. Harwood, K. Höffner, and P. I. Barton, “Efficient solution of ordinary dif- ferential equations with a parametric lexicographic linear program embedded,” *Numerische Mathematik*, vol. 133, pp. 623–653, 8 2016.

ANEXOS

4.10 Code to add pathways 1

```
#Importing packages
import cobra
from cobra import Model, Reaction, Metabolite
from cobra.io import read_sbml_model
import pandas as pd

#Importing model iYali
#yali = cobra.io.read_sbml_model('iYali-tidy W29 strain.xml')
yali = cobra.io.read_sbml_model('iYali-tidy Obese strain.xml')
#yali = cobra.io.read_sbml_model('yalinewpathways.xml')
#yali = cobra.io.read_sbml_model('iYali.xml')

#Creation of all the metabolites
propio_extr = Metabolite(
    's_3717',
    formula='C3H5O2',
    name='Propionate[e]',
    compartment='e')
propio_c = Metabolite(
```

```

's_3718',
formula='C3H5O2',
name='Propionate[c]',
compartment='c')

propiocoa_m = Metabolite(
's_1382')

propiocoa_c = Metabolite(
's_3719',
formula='C24H36N7O17P3S',
name='Propionyl-CoA[c]',
compartment='c')

propecoa_c = Metabolite(
's_3721',
formula='C24H38N7O17P3S',
name='Propenoyl-CoA[c]',
compartment='c')

oxigen_c = Metabolite(
's_1275')

hyprox_c = Metabolite(
's_0837')

hydroxcoa_c = Metabolite(
's_3722',
formula='C24H40N7O18P3S',
name='3-hydroxypropionyl-CoA [c]',
compartment='c')

water_c = Metabolite(
's_0803')

hydroxypropanoate_c = Metabolite(

```

```
's_3723',
formula='C3H5O3',
name='3-Hydroxypropanoate [c]',
compartment='c')

difos_c = Metabolite(
's_0633')

amp_c = Metabolite(
's_0423')

coe_c = Metabolite(
's_0529')

atp_c = Metabolite(
's_0434')

oxopro_c = Metabolite(
's_3724',
formula='C3H4O3',
name='3-oxopropanoate [c]',
compartment='c')

NAD_c = Metabolite(
's_1198',
compartment='c')

NADH_c = Metabolite(
's_1203')

h_c = Metabolite(
's_0794')

acet_c = Metabolite(
's_0373')

car_c = Metabolite(
's_0456')
```

```

rlac_m = Metabolite(
's_0027')

pyr_m = Metabolite(
's_1401')

NADH_m = Metabolite(
's_1205')

h_m = Metabolite(
's_0799')

NAD_m = Metabolite(
's_1200')

slac_c = Metabolite(
's_0063')

slac_p = Metabolite(
's_3725',
formula='C3H5O3',
name='(S)-lactate [p]',
compartment='p')

oxi_p = Metabolite(
's_1279')

pyru_p = Metabolite(
's_0206')

perox_p = Metabolite(
's_0840')

atp_m = Metabolite(
's_0437',
formula='C10H12N5O13P3',
name='ATP [m]',
compartment='m')

```

```

adp_m = Metabolite(
's_0397',
formula='C10H12N5O10P2',
name='ADP[m]',
compartment='m')

smethylma_m = Metabolite(
's_3726',
formula='C25H40N7O19P3S',
name='(S)-Methylmalonyl-CoA[m]',
compartment='m')

rmethylma_m = Metabolite(
's_3727',
formula='C25H40N7O19P3S',
name='(R)-Methylmalonyl-CoA[m]',
compartment='m')

succi_m = Metabolite(
's_1464')

acetyl_c = Metabolite(
's_0373')

h_c = Metabolite(
's_0794')

NADPH_c = Metabolite(
's_1212')

coenzymeA_c = Metabolite(
's_0529')

h2o_c = Metabolite(
's_0803')

NADP_c = Metabolite(

```

```
's_1207')

margariccoa_c = Metabolite(
's_3745',
formula='C38H68N7O17P3S',
name='Margaric-CoA[c]',
compartment='c')

margariccoa_lp = Metabolite(
's_3729',
formula='C38H68N7O17P3S',
name='Margaric-CoA[lp]',
compartment='lp')

margaric_lp = Metabolite(
's_3730',
formula='C17H34O2',
name='Margaric acid[lp]',
compartment='lp')

amp_lp = Metabolite(
's_2842')

disfos_lp = Metabolite(
's_0635')

atp_lp = Metabolite(
's_2840')

coenzymeA_lp = Metabolite(
's_0531')

coenzymeA_lp = Metabolite(
's_0531')

pentidecyliccoa_c = Metabolite(
's_3731',
```

```

formula='C15H30O2',
name='Pentadecylic-CoA[ c ]',
compartment='c')
pentadecyliccoa_lp = Metabolite(
's_3732',
formula='C15H30O2',
name='Pentadecylic-CoA[ lp ]',
compartment='lp')
pentadecylic_lp = Metabolite(
's_3733',
formula='C17H34O2',
name='Pentadecylic acid[ lp ]',
compartment='lp')
coenzymeA_lp = Metabolite(
's_0531')

reaction1 = Reaction('y300083')
reaction1.name = '3-Oxopropanoate:NAD+ oxidoreductase (decarboxylating, CoA-acetylating)'
reaction1.subsystem = 'Cell Propionate Biosynthesis'
reaction1.lower_bound = 0. # This is the default
reaction1.upper_bound = 1000. # This is the default
reaction1.add_metabolites({
oxopro_c: -1.0,
NAD_c: -1.0,
coe_c: -1.0,
NADH_c: 1.0,
h_c: 1.0,
acet_c: 1.0,
car_c: 1.0,
})

```

```

})

reaction1.reaction
yali.add_reactions([reaction1])
yali.reactions.y300083

reaction2 = Reaction('y300084')
reaction2.name = '3-hydroxypropanoate:NAD+ oxidoreductase'
reaction2.subsystem = 'Cell Propionate Biosynthesis'
reaction2.lower_bound = 0. # This is the default
reaction2.upper_bound = 1000. # This is the default
reaction2.add_metabolites({
    hydroxypropanoate_c: -1.0,
    NAD_c: -1.0,
    NADH_c: 1.0,
    h_c: 1.0,
    oxopro_c: 1.0,
})
reaction2.reaction
yali.add_reactions([reaction2])
yali.reactions.y300084

reaction3 = Reaction('y300085')
reaction3.name = '3-hydroxypropionate:CoA ligase (AMP-forming)'
reaction3.subsystem = 'Cell Propionate Biosynthesis'
reaction3.lower_bound = 0. # This is the default
reaction3.upper_bound = 1000. # This is the default
reaction3.add_metabolites({
    hydroxcoa_c: -1.0,
    difos_c: -1.0,
    amp_c: -1.0,
})

```

```

coe_c: 1.0,
atp_c: 1.0,
hydroxypropanoate_c: 1.0,
})

reaction3.reaction
yali.add_reactions([reaction3])
yali.reactions.y300085

reaction4 = Reaction('y300086')
reaction4.name = '3-hydroxypropionyl-CoA hydrolyase'
reaction4.subsystem = 'Cell Propionate Biosynthesis'
reaction4.lower_bound = 0. # This is the default
reaction4.upper_bound = 1000. # This is the default
reaction4.add_metabolites({
    propecoa_c: -1.0,
    hydroxcoa_c: 1.0,
    water_c: 1.0,
})
reaction4.reaction
yali.add_reactions([reaction4])
yali.reactions.y300086

reaction5 = Reaction('y300087')
reaction5.name = 'propanoyl-CoA: electron-transfer flavoprotein 2,3-oxidoreductase'
reaction5.subsystem = 'Cell Propionate Biosynthesis'
reaction5.lower_bound = 0. # This is the default
reaction5.upper_bound = 1000. # This is the default
reaction5.add_metabolites({
    propiocoa_c: -1.0,
    oxigen_c: -1.0,
})

```

```

hyprox_c: 1.0,
propecoa_c: 1.0,
})

reaction5.reaction
yali.add_reactions([reaction5])
yali.reactions.y300087

reaction6 = Reaction('y300088')
reaction6.name = 'Propionyl-CoA transporter [c] - [m]'
reaction6.subsystem = 'Cell Propionate Biosynthesis'
reaction6.lower_bound = 0. # This is the default
reaction6.upper_bound = 1000. # This is the default
reaction6.add_metabolites({
    propiocoa_c: -1.0,
    propiocoa_m: 1.0,
})

reaction6.reaction
yali.add_reactions([reaction6])
yali.reactions.y300088

reaction7 = Reaction('y300089')
reaction7.name = 'Propionate CoA-transferase'
reaction7.subsystem = 'Cell Propionate Biosynthesis'
reaction7.lower_bound = 0. # This is the default
reaction7.upper_bound = 1000. # This is the default
reaction7.add_metabolites({
    propio_c: -1.0,
    propiocoa_c: 1.0,
})

reaction7.reaction

```

```

yali.add_reactions([reaction7])
yali.reactions.y300089
reaction8 = Reaction('y300090')
reaction8.name = 'Propionate Extracellular Transporter'
reaction8.subsystem = 'Cell Propionate Biosynthesis'
reaction8.lower_bound = 0. # This is the default
reaction8.upper_bound = 1000. # This is the default
reaction8.add_metabolites({
    propio_extr: -1.0,
    propio_c: 1.0,
})
reaction8.reaction
yali.add_reactions([reaction8])
yali.reactions.y300090
reaction9 = Reaction('y300091')
reaction9.name = '(R)-Methylmalonyl-CoA CoA-carbonylmutase'
reaction9.subsystem = 'Cell Methylmalonyl Pathway'
reaction9.lower_bound = 0. # This is the default
reaction9.upper_bound = 1000. # This is the default
reaction9.add_metabolites({
    rmethylma_m: -1.0,
    succi_m: 1.0,
})
reaction9.reaction
yali.add_reactions([reaction9])
yali.reactions.y300091
reaction10 = Reaction('y300092')
reaction10.name = 'Methylmalonyl-CoA epimerase'

```

```

reaction10.subsystem = 'Cell Methylmalonyl Pathway'
reaction10.lower_bound = 0. # This is the default
reaction10.upper_bound = 1000. # This is the default
reaction10.add_metabolites({
    smethylma_m: -1.0,
    rmethylma_m: 1.0,
})
reaction10.reaction
yali.add_reactions([reaction10])
yali.reactions.y300092
reaction11 = Reaction('y300093')
reaction11.name = 'Propionyl-CoA carboxylase'
reaction11.subsystem = 'Cell Methylmalonyl Pathway'
reaction11.lower_bound = 0. # This is the default
reaction11.upper_bound = 1000. # This is the default
reaction11.add_metabolites({
    propiocoa_m: -1.0,
    atp_m: -1.0,
    adp_m: 1.0,
    smethylma_m: 1.0,
})
reaction11.reaction
yali.add_reactions([reaction11])
yali.reactions.y300093
reaction12 = Reaction('y300094')
reaction12.name = '(R)-Methylmalonyl-CoA CoA-carboxylmutase'
reaction12.subsystem = 'Cell Methylmalonyl Pathway'
reaction12.lower_bound = -1000. # This is the default

```

```

reaction12.upper_bound = 1000. # This is the default
reaction12.add_metabolites({
rmethylma_m: -1.0,
succ_i_m: 1.0,
})
reaction12.reaction
yali.add_reactions([reaction12])
yali.reactions.y300094
reaction13 = Reaction('y300095')
reaction13.name = 'Methylmalonyl-CoA epimerase'
reaction13.subsystem = 'Cell Methylmalonyl Pathway'
reaction13.lower_bound = -1000. # This is the default
reaction13.upper_bound = 1000. # This is the default
reaction13.add_metabolites({
smethylma_m: -1.0,
rmethylma_m: 1.0,
})
reaction13.reaction
yali.add_reactions([reaction13])
yali.reactions.y300095
reaction14 = Reaction('y300096')
reaction14.name = 'Peroxisomal lactate dehydrogenase'
reaction14.subsystem = 'Cell S-lactate Degradation'
reaction14.lower_bound = -1000. # This is the default
reaction14.upper_bound = 1000. # This is the default
reaction14.add_metabolites({
slac_p: -1.0,
oxi_p: -1.0,
})

```

```

pyru_p: 1.0,
perox_p: 1.0,
})

reaction14.reaction
yali.add_reactions([reaction14])
reaction15 = Reaction('y300097')
reaction15.name = '(S)-lactate transporter [c] - [p]'
reaction15.subsystem = 'Cell S-lactate Degradation'
reaction15.lower_bound = -1000. # This is the default
reaction15.upper_bound = 1000. # This is the default
reaction15.add_metabolites({
    slac_c: -1.0,
    slac_p: 1.0,
})
reaction15.reaction
yali.add_reactions([reaction15])
yali.reactions.y300097
reaction16 = Reaction('y300098')
reaction16.name = '(R)-Lactate:NAD+ oxidoreductase'
reaction16.subsystem = 'Cell D-lactate Degradation'
reaction16.lower_bound = -1000. # This is the default
reaction16.upper_bound = 1000. # This is the default
reaction16.add_metabolites({
    rlac_m: -1.0,
    NAD_m: -1.0,
    NADH_m: 1.0,
    h_m: 1.0,
    pyr_m: 1.0,
})

```

```

})

reaction16.reaction
yali.add_reactions([reaction16])
yali.reactions.y300098

reaction17 = Reaction('y300099')
reaction17.name = 'fatty-acyl-CoA synthase (n-C17:0CoA)'
reaction17.subsystem = 'Odd Chain Fatty Acid Synthesis'
reaction17.lower_bound = 0. # This is the default
reaction17.upper_bound = 1000. # This is the default
reaction17.add_metabolites({
    acetyl_c: -1.0,
    h_c: -24.0,
    propioco_a_c: -8.0,
    NADPH_c: -16.0,
    car_c: 8.0,
    coenzymeA_c: 8.0,
    h2o_c: 8.0,
    NADP_c: 16.0,
    margariccoa_c: 1.0,
})
reaction17.reaction
yali.add_reactions([reaction17])
yali.reactions.y300099

reaction18 = Reaction('y300100')
reaction18.name = 'margaric-CoA transport, cytoplasm-lipid particle'
reaction18.subsystem = 'Odd Chain Fatty Acid Synthesis'
reaction18.lower_bound = -1000. # This is the default
reaction18.upper_bound = 1000. # This is the default

```

```

reaction18.add_metabolites({
    margariccoa_c: -1.0,
    margariccoa_lp: 1.0,
})

reaction18.reaction
yali.add_reactions([reaction18])
yali.reactions.y300100

reaction20 = Reaction('y300102')
reaction20.name = 'margaric-CoA transport, cytoplasm-lipid particle'
reaction20.subsystem = 'Odd Chain Fatty Acid Synthesis'
reaction20.lower_bound = -1000. # This is the default
reaction20.upper_bound = 1000. # This is the default
reaction20.add_metabolites({
    margariccoa_lp: -1.0,
    amp_lp: -1.0,
    disfos_lp: -1.0,
    margaric_lp: 1.0,
    atp_lp: 1.0,
    coenzymeA_lp: 1.0,
})
reaction20.reaction
yali.add_reactions([reaction20])
yali.reactions.y300102

reaction21 = Reaction('y300104')
reaction21.name = 'fatty-acyl-CoA synthase (n-C15:0CoA)'
reaction21.subsystem = 'Odd Chain Fatty Acid Synthesis'
reaction21.lower_bound = 0. # This is the default
reaction21.upper_bound = 1000. # This is the default

```

```

reaction21.add_metabolites ({
    acetyl_c: -1.0,
    h_c: -21.0,
    propioco_a_c: -7.0,
    NADPH_c: -14.0,
    car_c: 7.0,
    coenzymeA_c: 7.0,
    h2o_c: 7.0,
    NADP_c: 14.0,
    pentidecylcoa_c: 1.0,
})
reaction21.reaction
yali.add_reactions([reaction21])
yali.reactions.y300104
reaction22 = Reaction('y300105')
reaction22.name = 'Pentadecyclic-CoA transport, cytoplasm-lipid particle'
reaction22.subsystem = 'Odd Chain Fatty Acid Synthesis'
reaction22.lower_bound = -1000. # This is the default
reaction22.upper_bound = 1000. # This is the default
reaction22.add_metabolites ({
    pentidecylcoa_c: -1.0,
    pentidecylcoa_lp: 1.0,
})
reaction22.reaction
yali.add_reactions([reaction22])
yali.reactions.y300105
reaction23 = Reaction('y300106')
reaction23.name = 'Pentadecyclic-CoA transport, cytoplasm-lipid particle'

```

```
reaction23.subsystem = 'Odd Chain Fatty Acid Synthesis'
reaction23.lower_bound = -1000. # This is the default
reaction23.upper_bound = 1000. # This is the default
reaction23.add_metabolites({
    pentidecylcoa_lp: -1.0,
    amp_lp: -1.0,
    disfos_lp: -1.0,
    pentadecylic_lp: 1.0,
    atp_lp: 1.0,
    coenzymeA_lp: 1.0,
})
reaction23.reaction
yali.add_reactions([reaction23])
yali.reactions.y300106
yali.objective = 'y300090'
#cobra.io.json.save_json_model(yali, 'yalinepathways_trial_2.json')
#cobra.io.write_sbml_model(yali, "yalinepathways_trial.xml")
cobra.io.write_sbml_model(yali, "iYali-tidy_Obese strain final.xml")
```