Optimizing the genomic-enzymatic model of *Yarrowia lipolytica* for biofuel production: a constraint-based reconstruction analysis approach

Presented by:
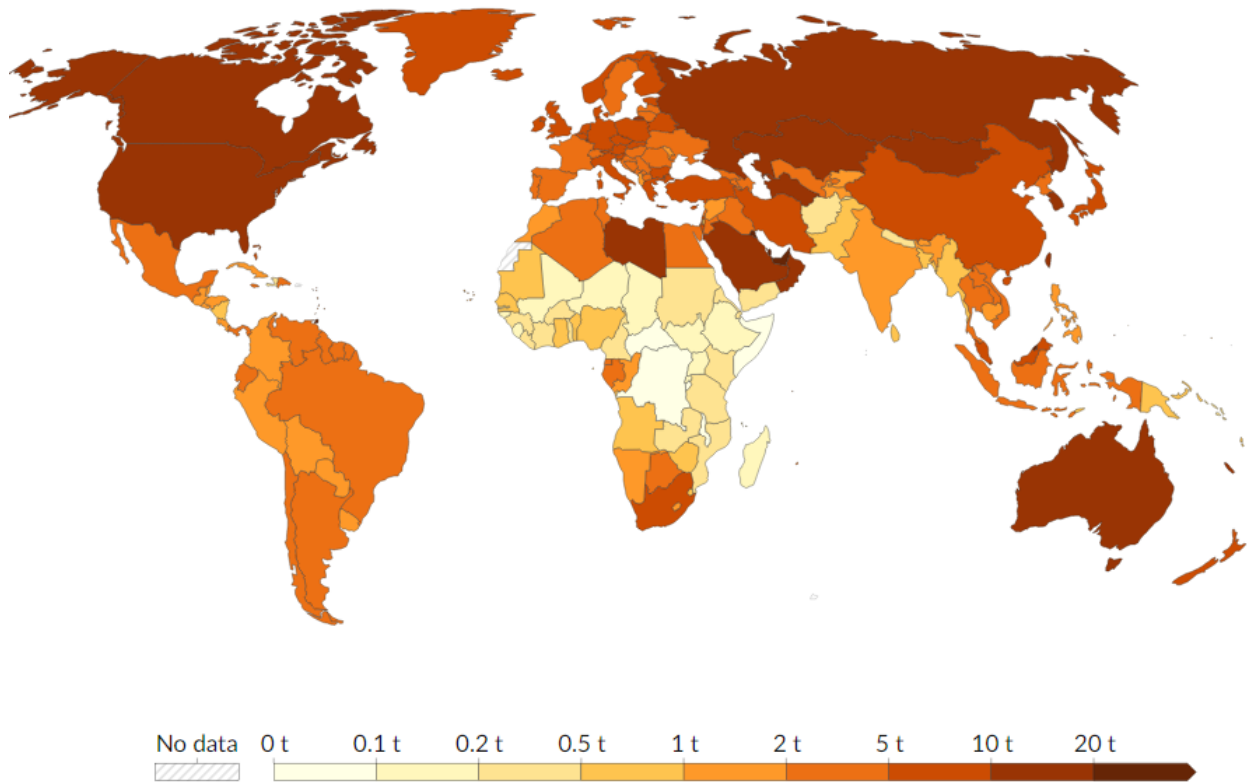
Jeremy GUERRERO

Supervised by:

Paul CARDENAS
Martin GUTIERREZ
Sabine PERES

Introduction

Per capita CO_2 emissions, 2021

The world's largest per capita CO2 emitters are the major oil producing countries

Energy use in industry and transport fuel accounts for 48.2%

Urgent need to find new ways of producing highly valuable compounds in a sustainable manner
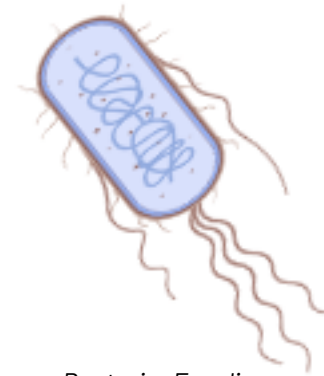
No data | 0 t | 0.1 t | 0.2 t | 0.5 t | 1 t | 2 t | 5 t | 10 t | 20 t

Source: Our World in Data based on the Global Carbon Project (2023)

Alternative industry bioproduction

Micro Algae: Chlamydomonas

Cons:
- Low Productivity
- Difficulty in Scale-Up
- Harvesting and Processing

Bacteria: E. coli
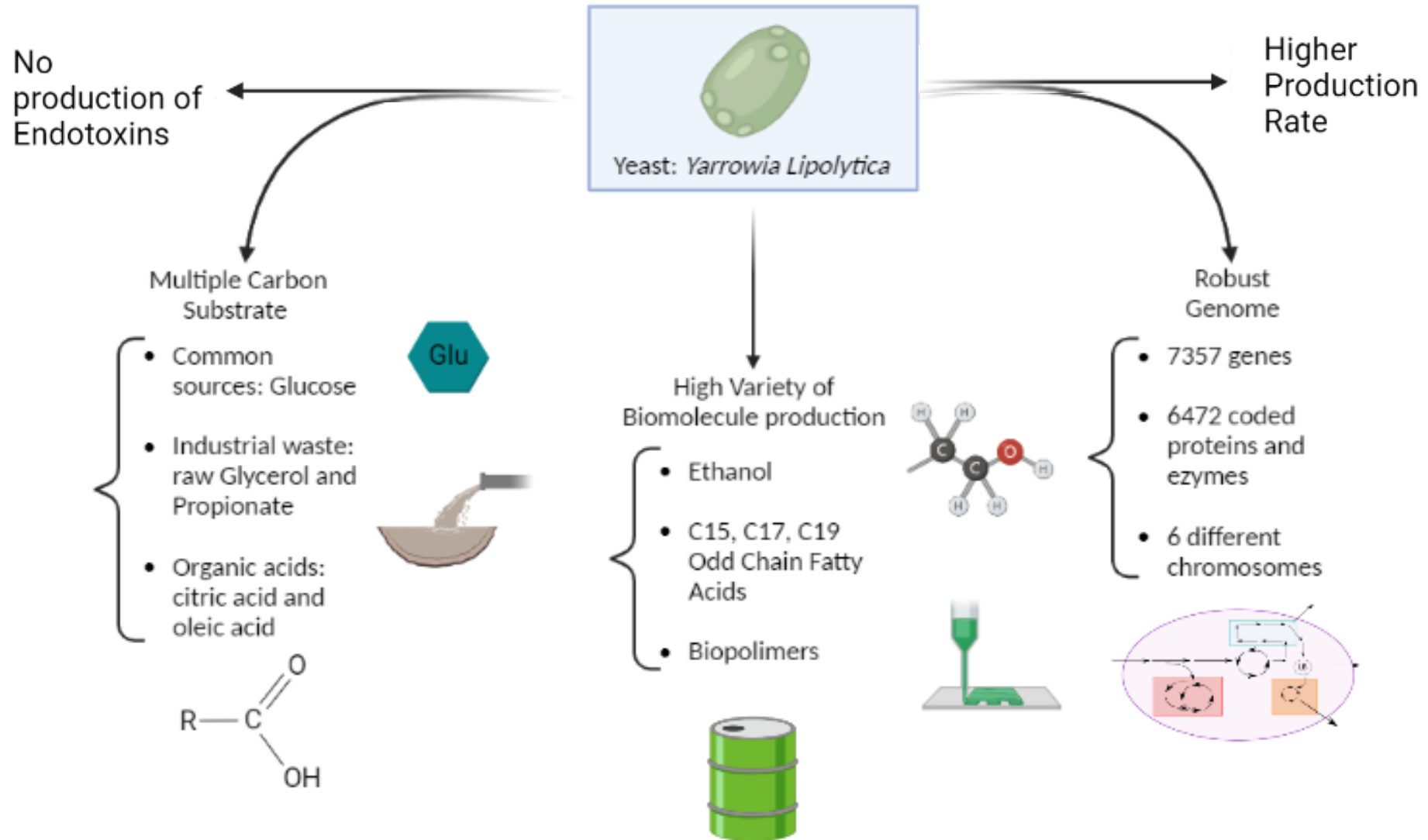
Cons:
- Metabolic Burden
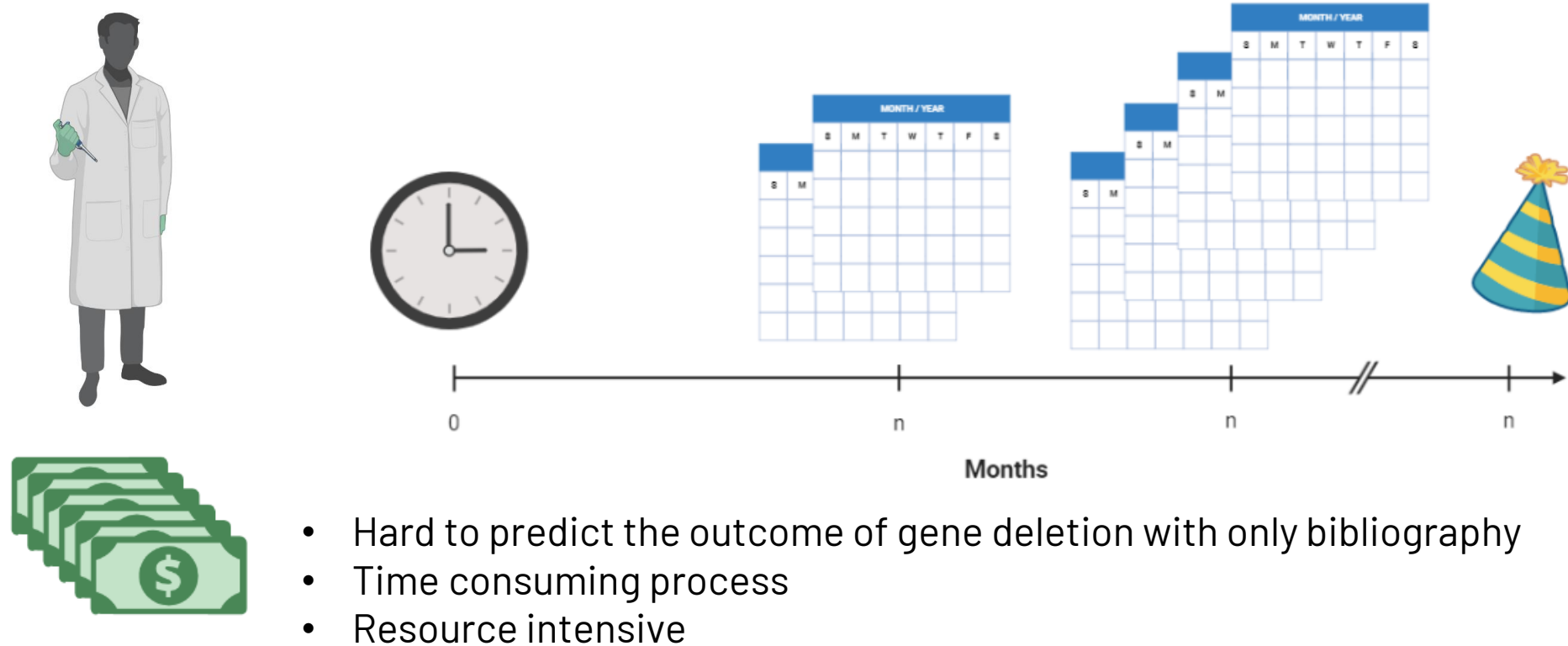- Endotoxins
- Protein Folding

Yeast: S. s. cerevisiae
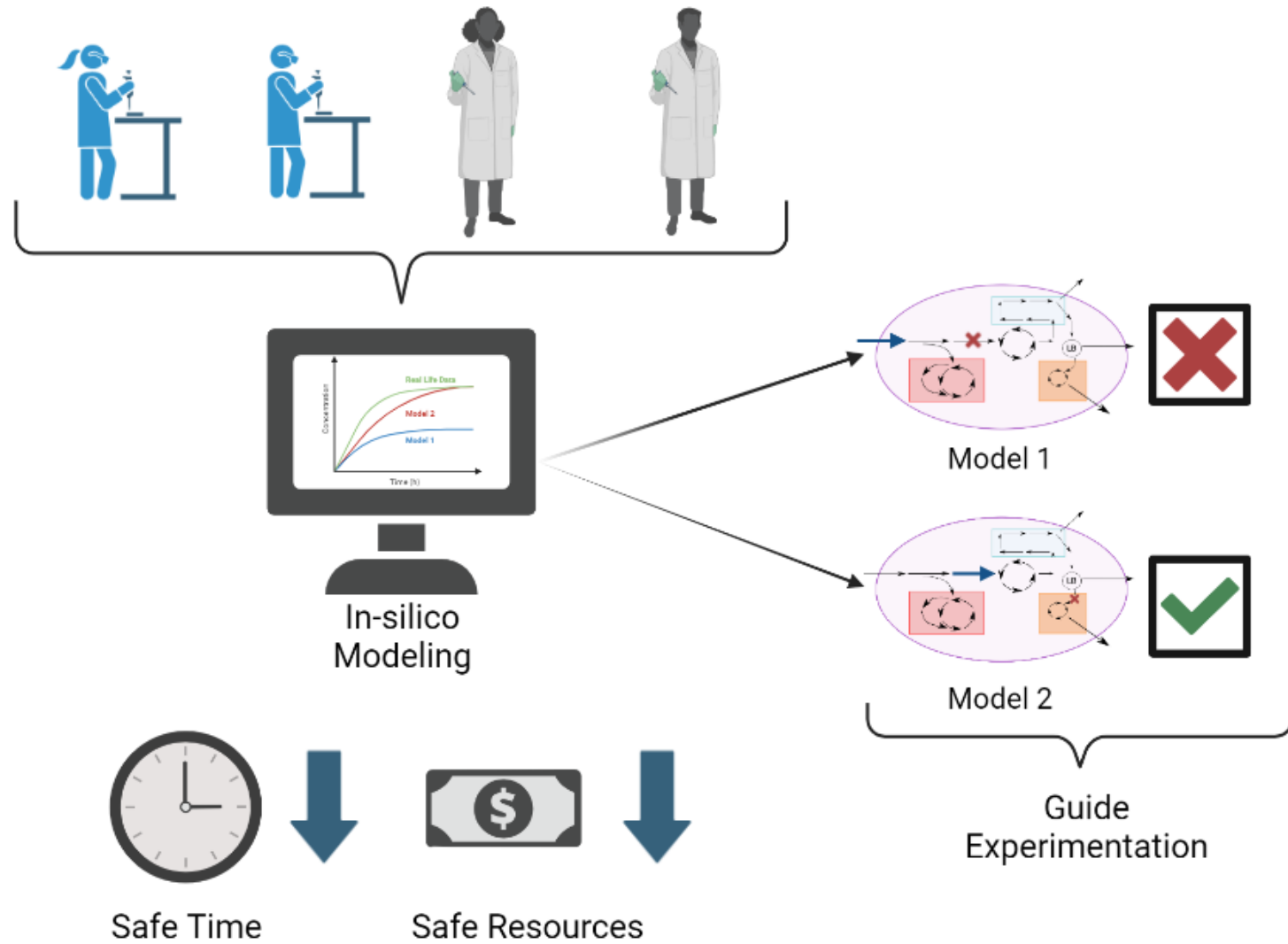
Cons:
- Metabolic Burden
- Endotoxins
- Protein Folding

Novel alternative

No production of Endotoxins

Higher Production Rate

Yeast: *Yarrowia Lipolytica*

Multiple Carbon Substrate

- Common sources: Glucose
- Industrial waste: raw Glycerol and Propionate
- Organic acids: citric acid and oleic acid

High Variety of Biomolecule production

- Ethanol
- C15, C17, C19 Odd Chain Fatty Acids
- Biopolimers

Robust Genome

- 7357 genes
- 6472 coded proteins and ezymes
- 6 different chromosomes

Main Problem: Development of Phenotypes

- Hard to predict the outcome of gene deletion with only bibliography
- Time consuming process
- Resource intensive

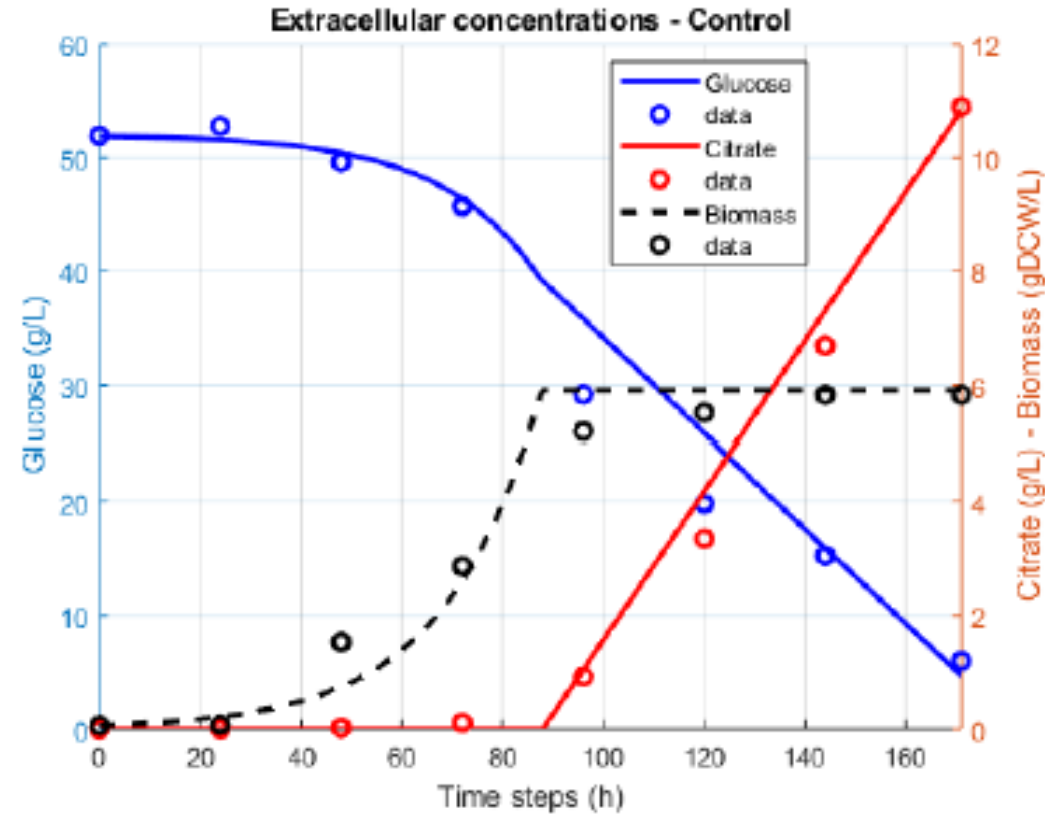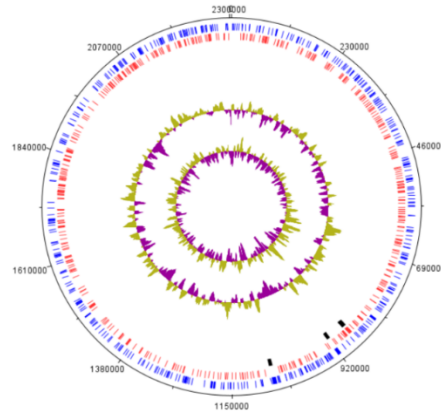# Solution Through Simulation

Antecedentes

Figure 2a: Calibrating dFBA model with experiemental flux rates allowed predictions close to experimental résultats, except for isocitrate wich is not predicted by the model.

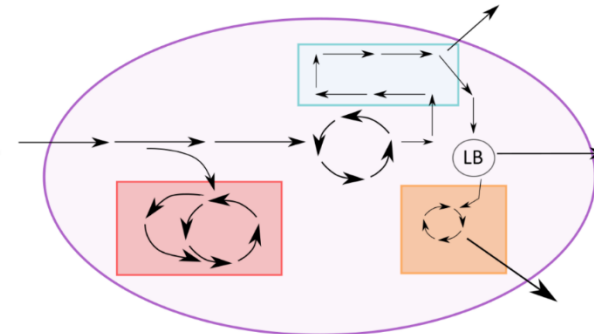# Theoretical Framework

Reconstruction of GEM Model

1) Annotated Genome

2) Recognition of genetic circuits

TF — Enhancer

TF — Promoter — Target gene

3) Metabolic Network Reconstruction

- Metabolites
- Enzymes
- Reactions
- Substrate

Reaction Network

R1: 2A + B → C + 3D
R2: A + 3B → C + E
R3: A → 2B + D + E
R4: 4B → D + A
R5: D + 2B → C + 2E
R6: C + 4E → 3B + D

Stoichiometric Matrix

	R1	R2	R3	R4	R5	R6
A	-2	-1	-1	1	0	0
B	-1	-3	2	-4	-2	3
C	1	1	0	0	1	-1
D	3	0	1	1	-1	1
E	0	1	1	0	2	-4

Mathematic Abstraction

4) Stoichiometric Matrix

Gene–Protein–Reaction in GEM

Linear Programming

FBA Base Equations

Matrix Form

S matrix with n columns m rows

Flux $= v$
Objective $= c$

$$\frac{d(mass)}{d(time)} = S * v(metabolite, substrate, product) = 0,$$

$$S * v = 0,$$

$$v > 0, \forall i \in \text{irreversible reactions},$$

System of Linear Equations Form

maximize c^T,

subject to $Sv = 0$,

flux vector $V = (v_1, v_2, v_3, v_4, ..., v_n)$,

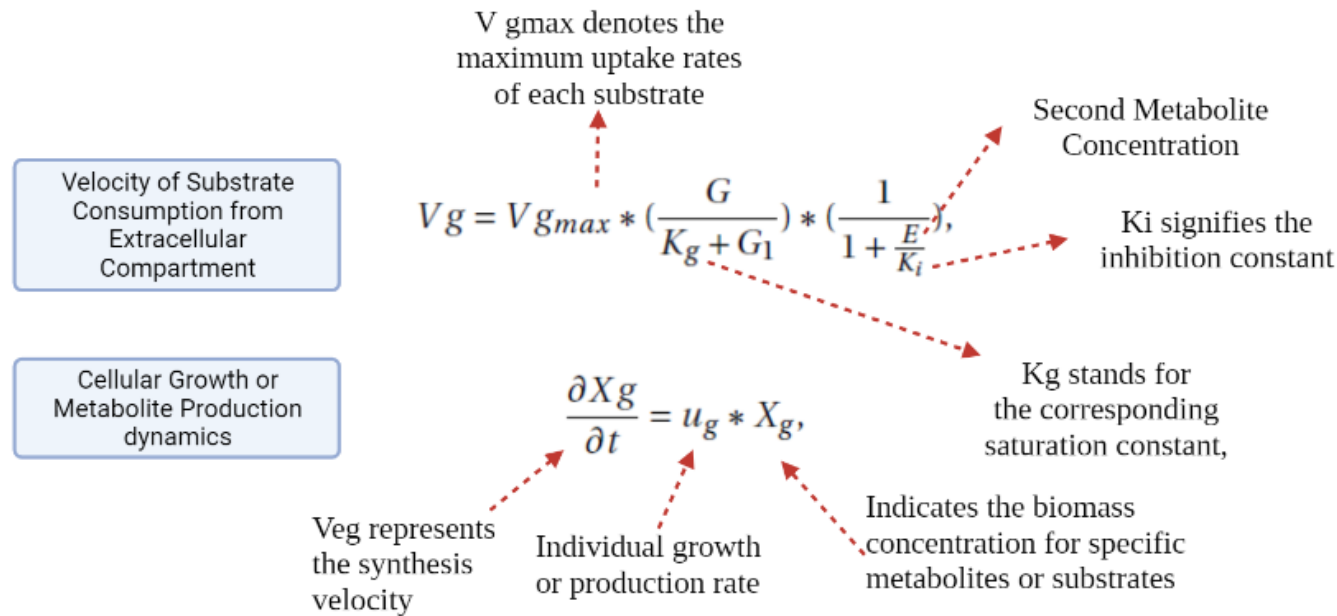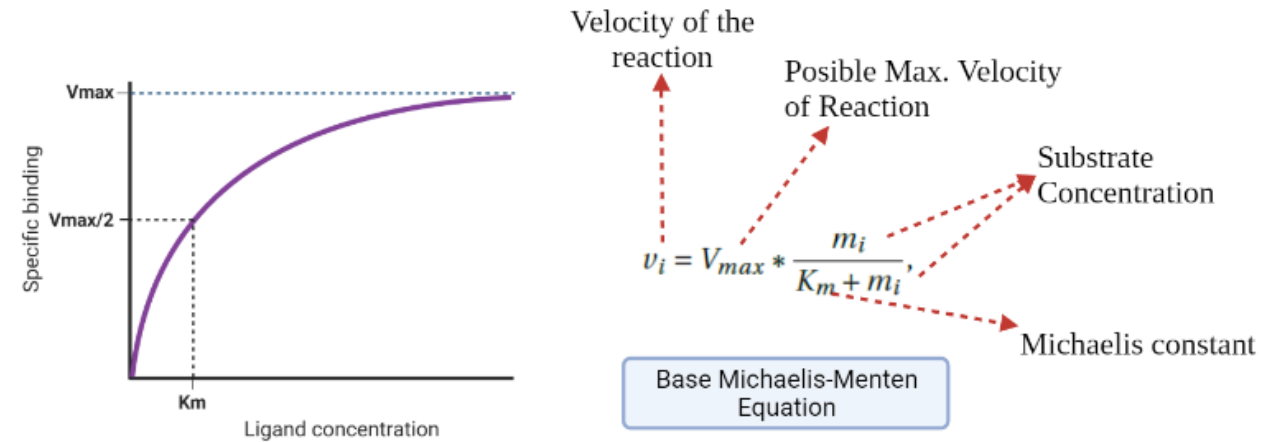constrain $v_{lb} \leq v \leq v_{up}$,

objective function Z, is articulated as a linear combination of our objective fluxes

the metabolite consumption and production within the system conserving mass balance

fluxes vi and exchange reactions ei are encapsulated within a vector

lower vlb and upper vub bounds for our fluxes, thereby constricting the solution space

dFBA Base Equations

Velocity of the reaction

Posible Max. Velocity of Reaction

Substrate Concentration

$$v_i = V_{max} * \frac{m_i}{K_m + m_i},$$

Michaelis constant

Base Michaelis-Menten Equation

V gmax denotes the maximum uptake rates of each substrate

Second Metabolite Concentration

Velocity of Substrate Consumption from Extracellular Compartment

$$Vg = Vg_{max} * (\frac{G}{K_g + G_1}) * (\frac{1}{1 + \frac{E}{K_i}}),$$

Ki signifies the inhibition constant

Kg stands for the corresponding saturation constant,

Cellular Growth or Metabolite Production dynamics

$$\frac{\partial Xg}{\partial t} = u_g * X_g,$$

Indicates the biomass concentration for specific metabolites or substrates

Veg represents the synthesis velocity

Individual growth or production rate

Objectives

Objectives

Main Objective

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.

Specific Objectives

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-vitro 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.

Methodology

Methodology Workflow
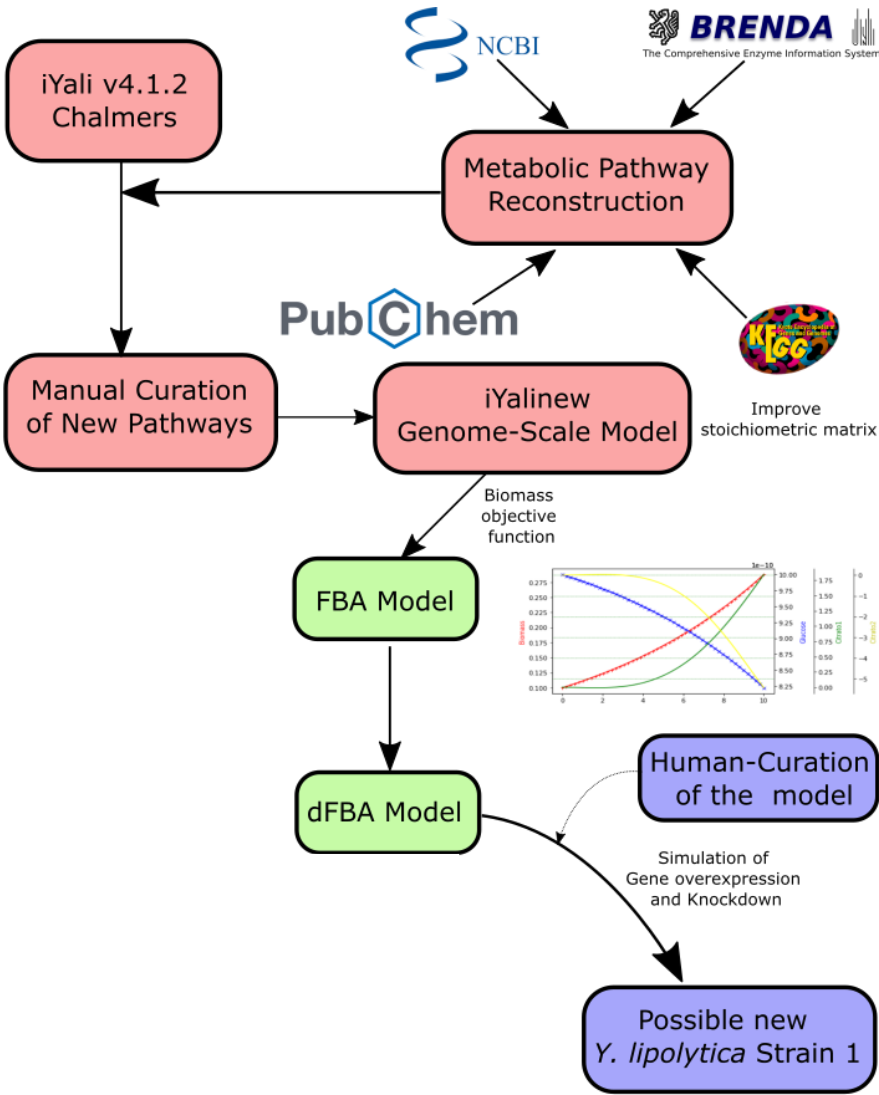
Used Kinetics for dFBA

Parameter	Description	*S. cerevisiae* 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

Contrasting Laboratory Data vs Modelled Data

Laboratory Biologie Intégrative du Métabolisme Lipidique (BIMLip)

- 2020 and 2021
- Unpublished Status

Modelled Performance Indicators

R-Squared

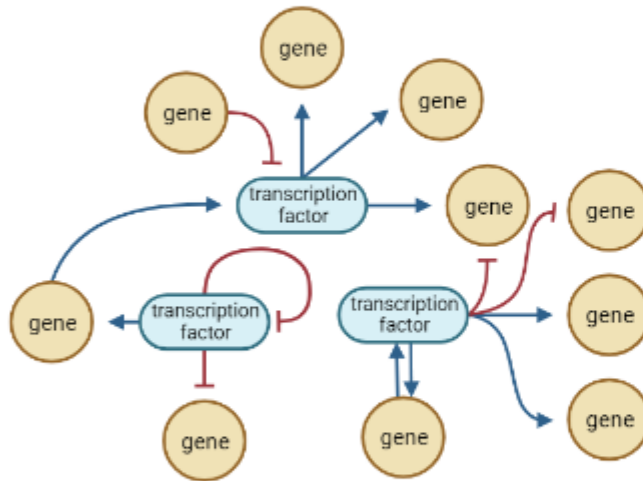$$R^2 = \sum_{t=1}^{t} (Y_{real} - Y_{modelled})^2$$

Mean Absolute Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

Root Mean Squared Error

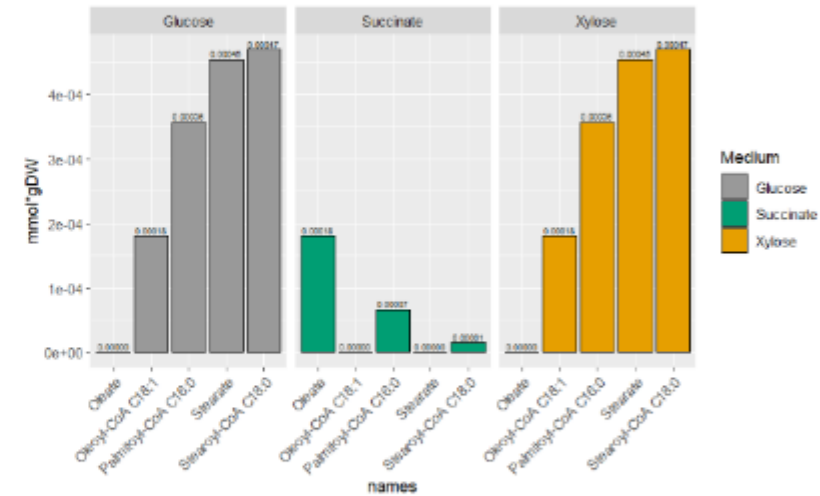$$MSE = \frac{\left| \sum_{i=1}^{N} (Predicted_i - Actual_i)^2 \right|}{N}$$

Independent and Dependent Variables

Results

Reconstructed Pathway: Propionate Absortion
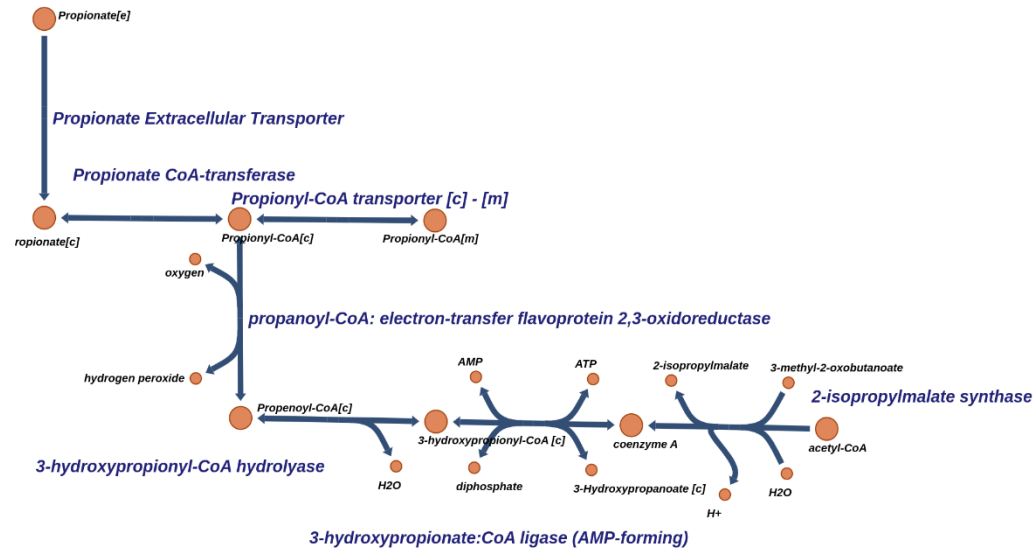
Rx ID	Description	Formula	Lower Bound	Upper Bound
y300090	Propionate Transporter [e] - [c]	s_3717 → s_3718	0.0	1000.0
y300089	Propionate to Propionyl-CoA	s_3718 → s_3719	0.0	1000.0
y300088	propanoyl-CoA:electron-transfer flavoprotein 2,3-oxidoreductase electron transfer flavoprotein	s_3719 → s_1382	0.0	1000.0
y300087	2,3-oxidoredcutase	s_1275 + s_3719 → s_0837 + s_3721	0.0	1000.0
y300086	3-hydroxypropionyl-CoA hydrolyase	s_3721 → s_0803 + s_3722	0.0	1000.0
y300085	3-hydroxypropionate:CoA ligase (AMP-forming)	s_0423 + s_0633 + s_3722 → s_0434 + s_0529 + s_3723	0.0	1000.0
y300084	3-hydroxypropanoate: NAD+ oxidoreductase	s_1198 + s_3723 → 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
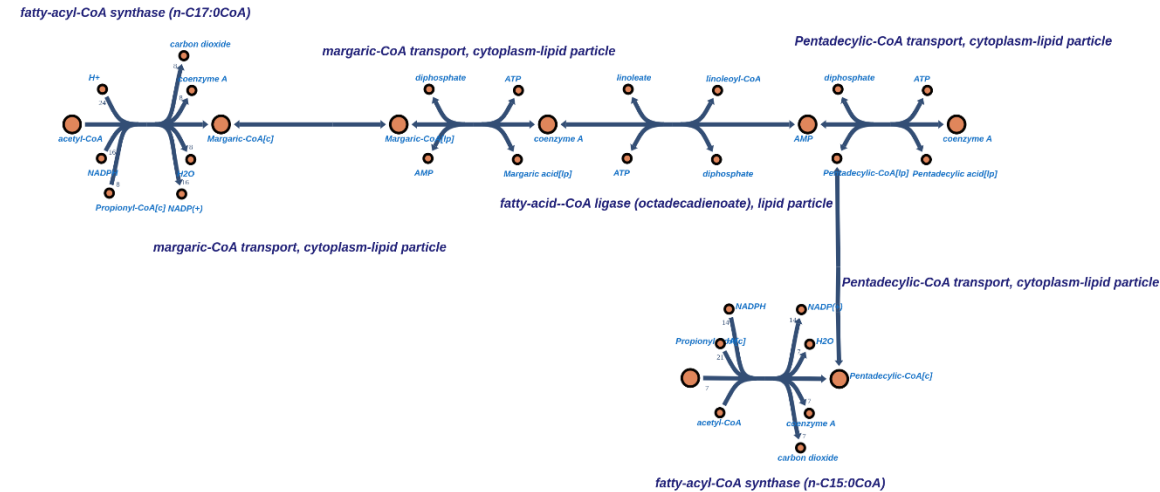
Rx ID	Description	Formula	Lower Bound	Upper Bound
y300099	fatty-acyl-CoA synthase (n-C17:0CoA)	s_0373 + 24.0 s_0794 + 16.0 s_1212 + 8.0 s_3719 → 8.0 s_0456 + 8.0 s_0529 + 8.0 s_0803 + 16.0 s_1207 + s_3728	0.0	1000.0
y300100	margaric-CoA transport, cytoplasm-lipid particle	s_3728 ↔ s_3729	-1000.0	1000.0
y300102	margaric-CoA transport, cytoplasm-lipid particle	s_0635 + s_2842 + s_3729 ↔ s_0531 + s_2840 + s_3730	-1000.0	1000.0
y300104	fatty-acyl-CoA synthase (n-C15:0CoA)	s_0373 + 21.0 s_0794 + 14.0 s_1212 + 7.0 s_3719 → 7.0 s_0456 + 7.0 s_0529 + 7.0 s_0803 + 14.0 s_1207 + s_3731	0.0	1000.0
y300105	Pentadecylic-CoA transport, cytoplasm-lipid particle	s_3731 ↔ s_3732	-1000.0	1000.0
y300106	Pentadecylic-CoA transport, cytoplasm-lipid particle	s_0635 + s_2842 + s_3732 ↔ s_0531 + s_2840 + s_3733	-1000.0	1000.0

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

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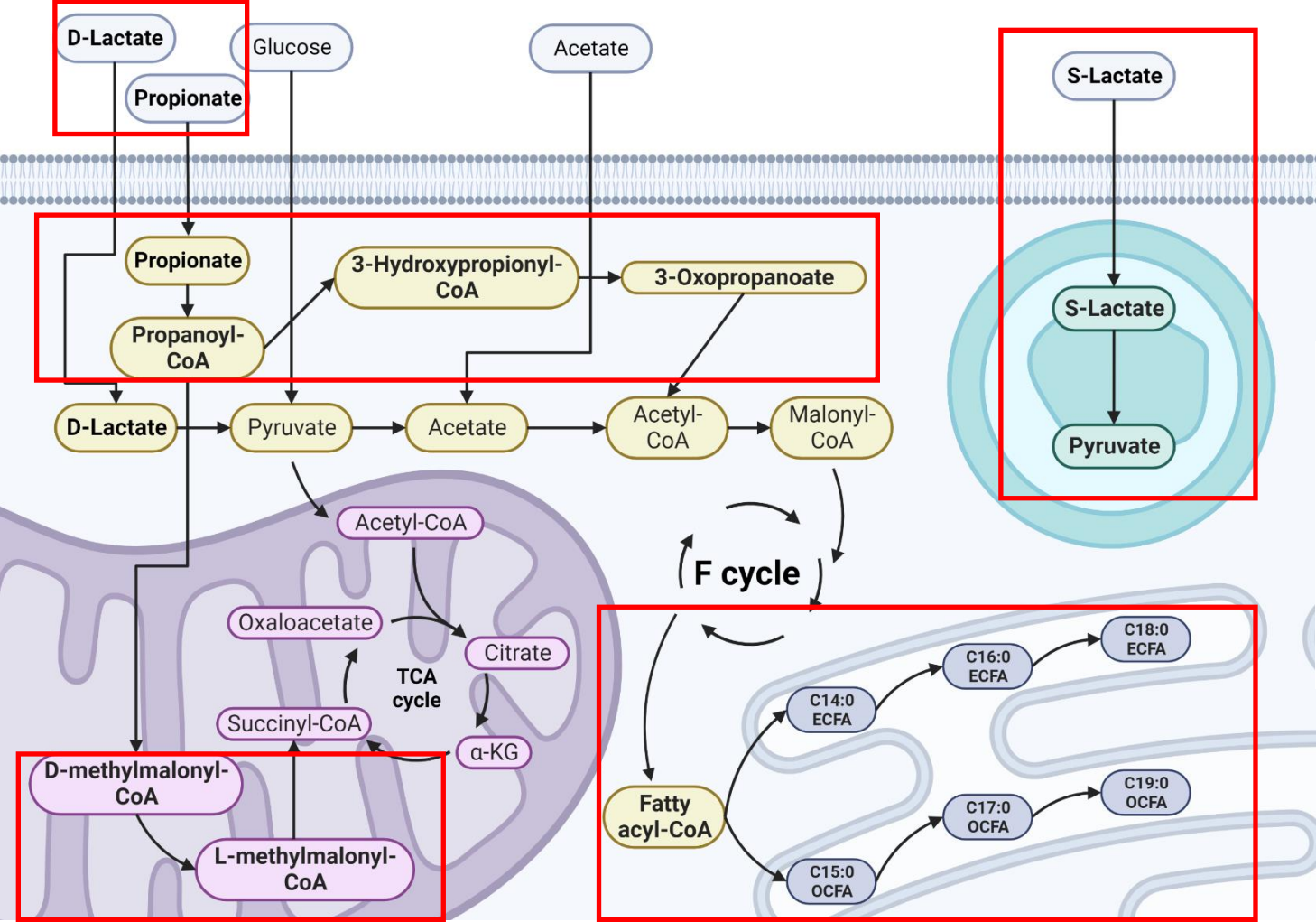
Characteristics	iYali4	new iYali
No. Genes	901	937
No. Metabolites	1,683	1,751
No. Reactions	1,985	2021

Anexos

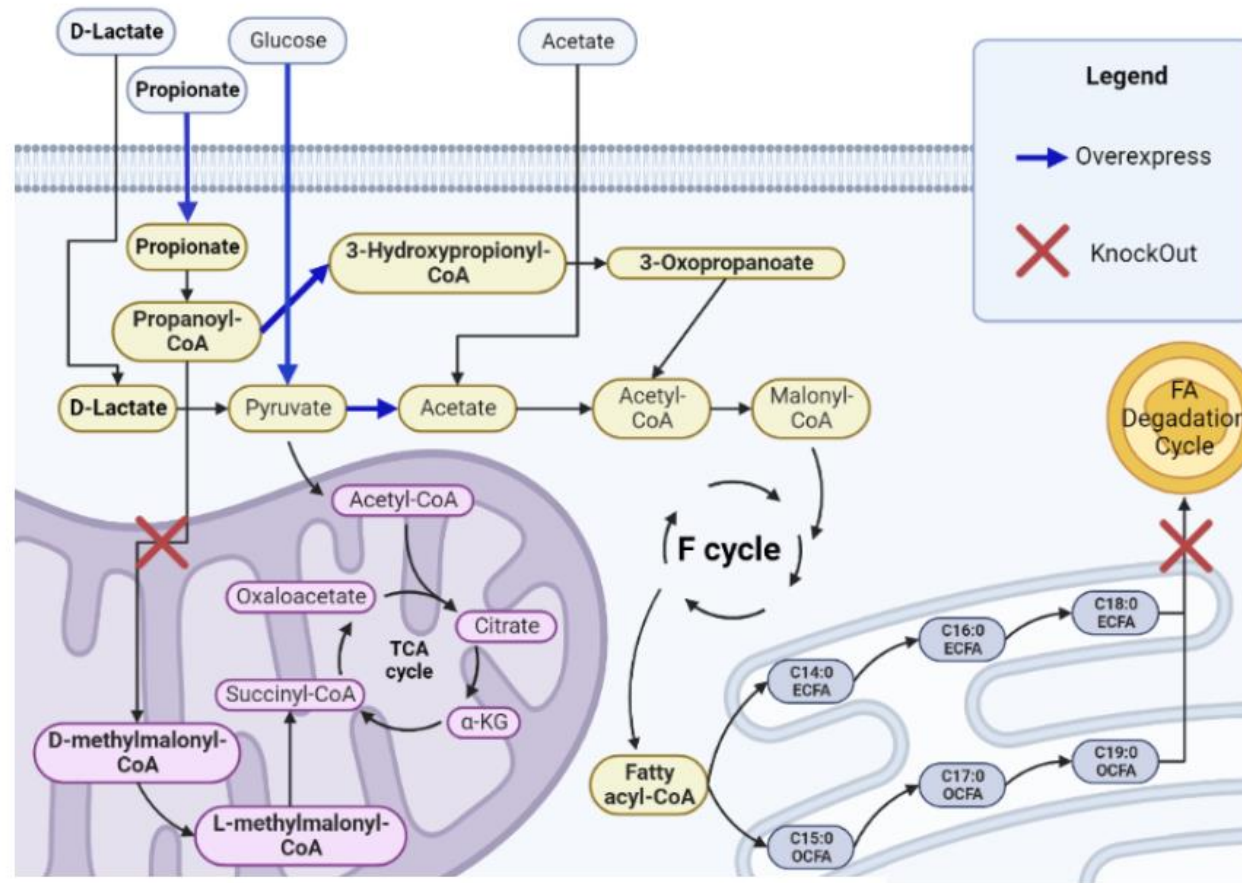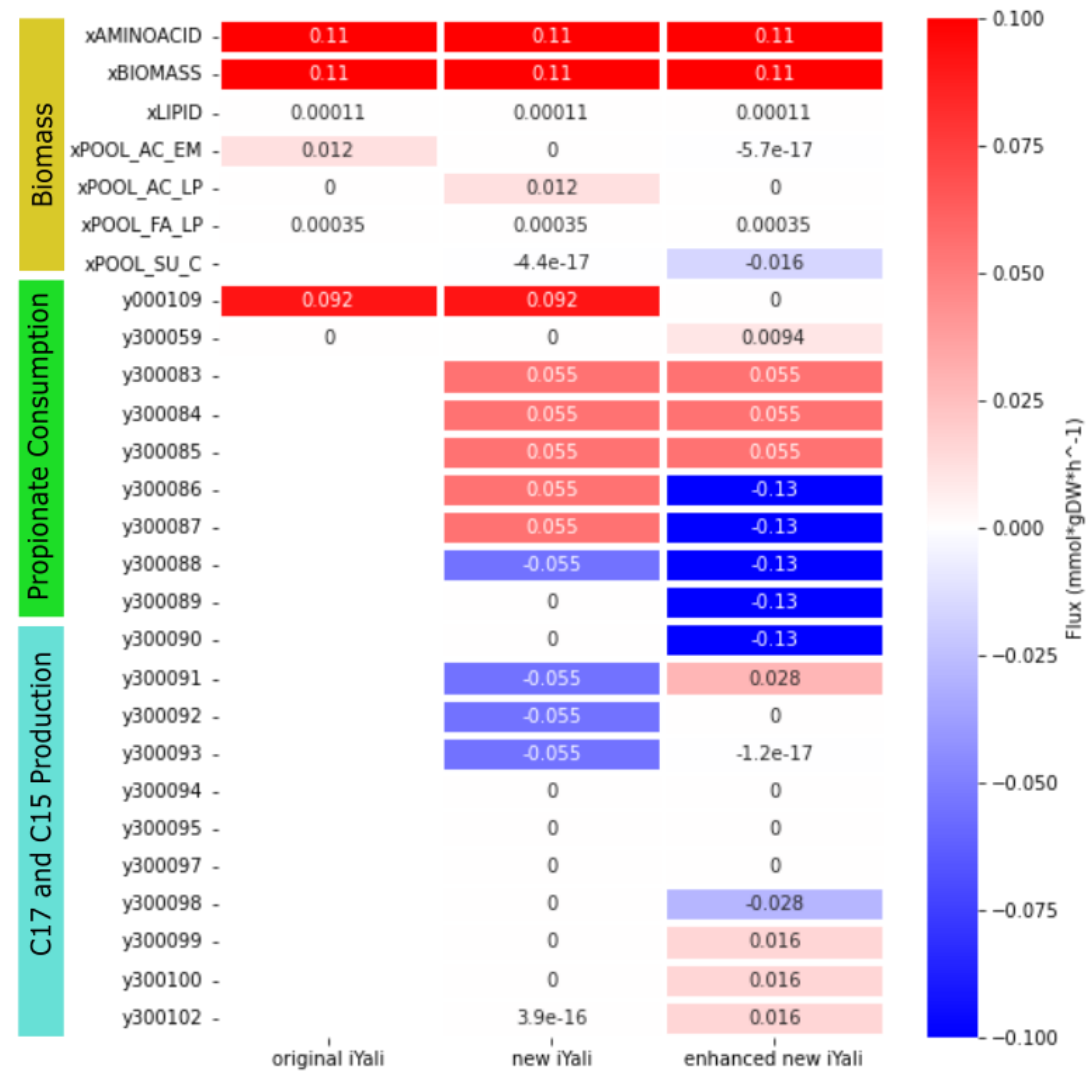Gene Essentiality in Yali Model
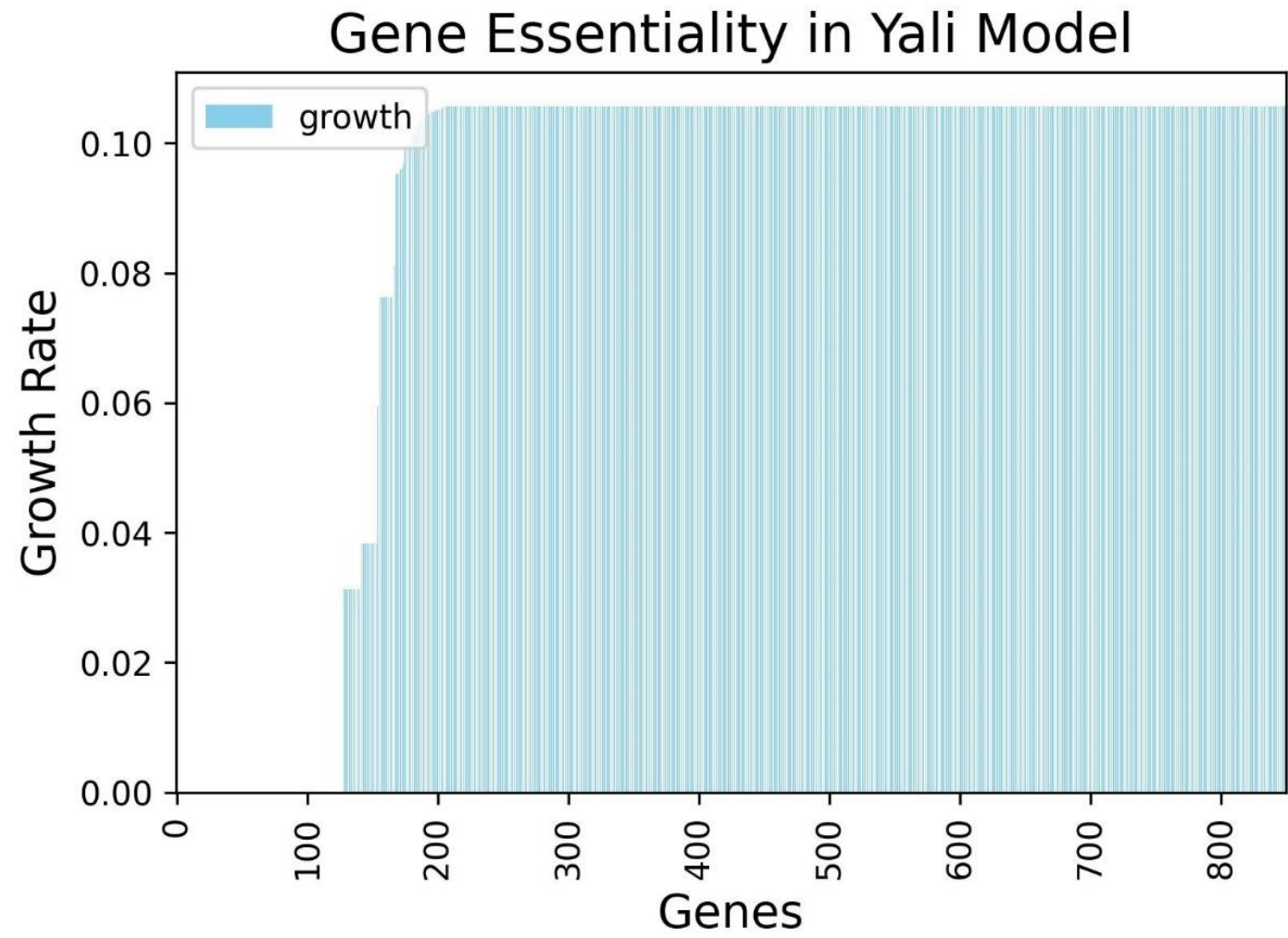
Enhanced new Y. lipolytica Model

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

Flux Analysis of Created Phenotypes
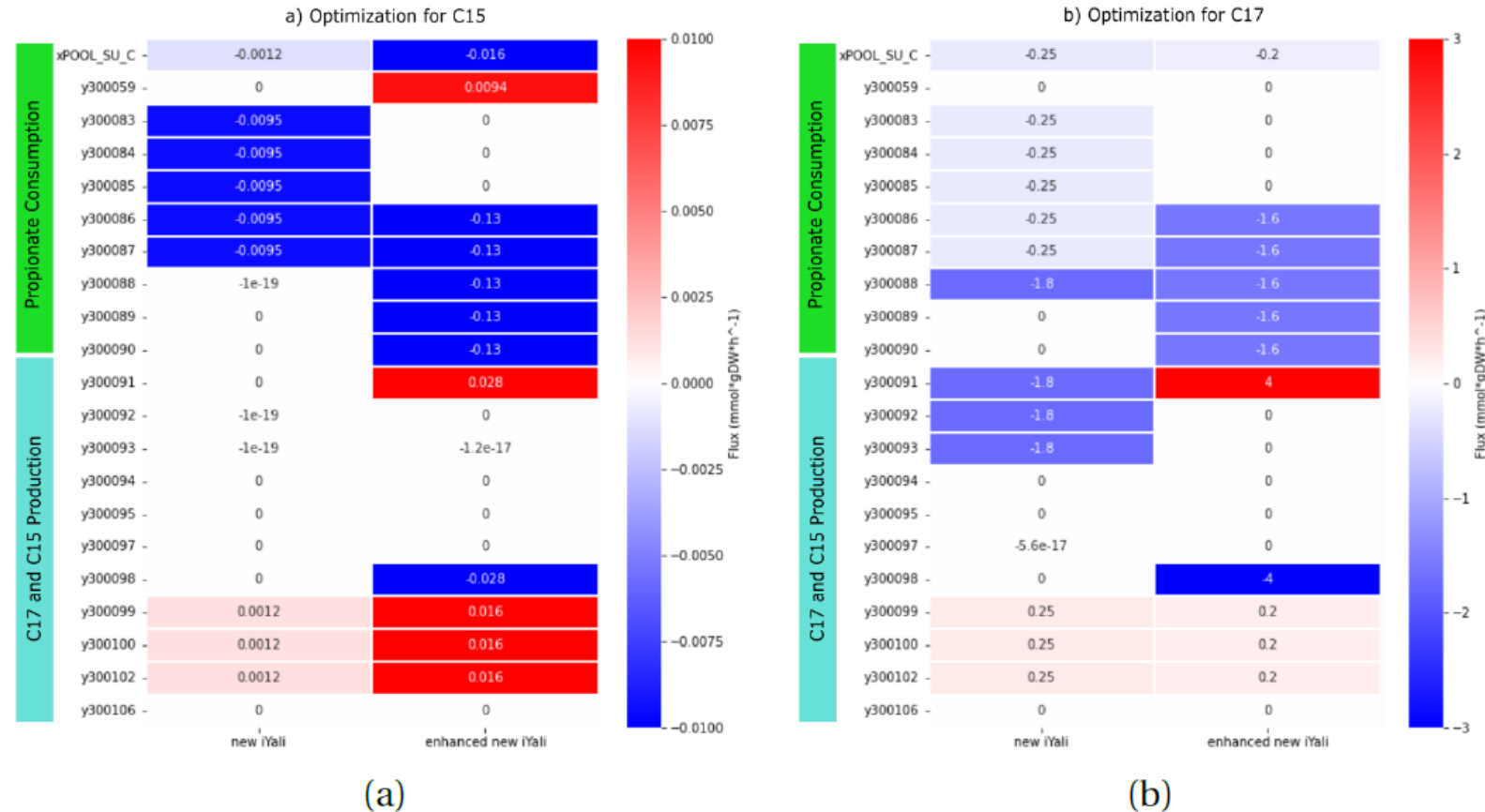
FIGURA 4.5: Flux density of the new iYali model and the enhanced iYali model in 2 different scenarios. (a) Production optimizations of C15. (b) Production optimizations of C17.
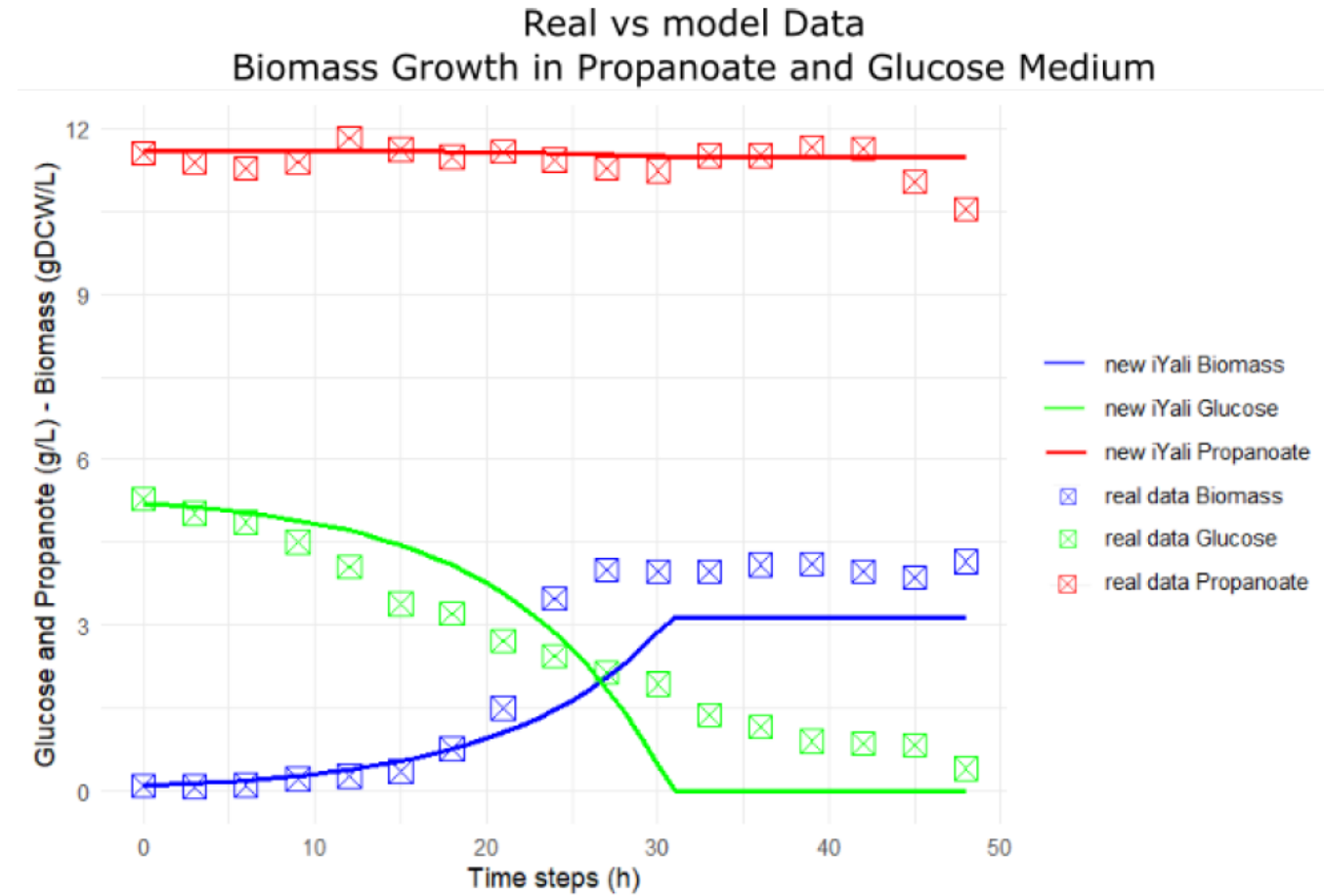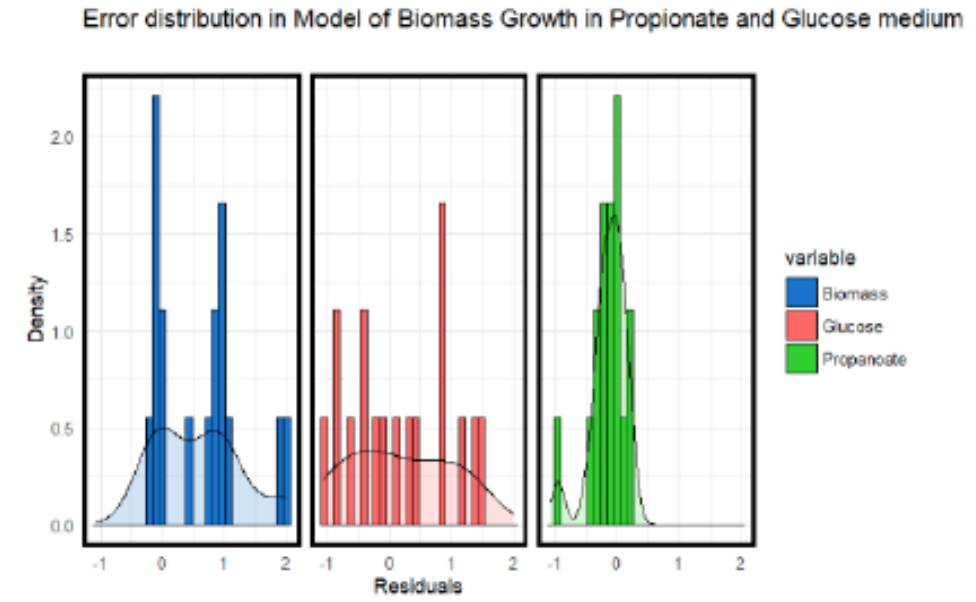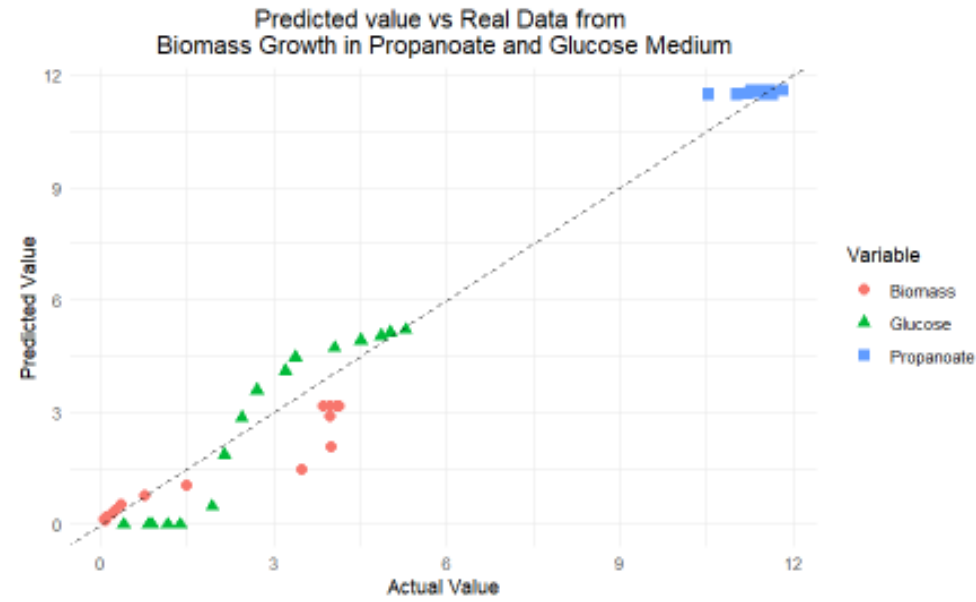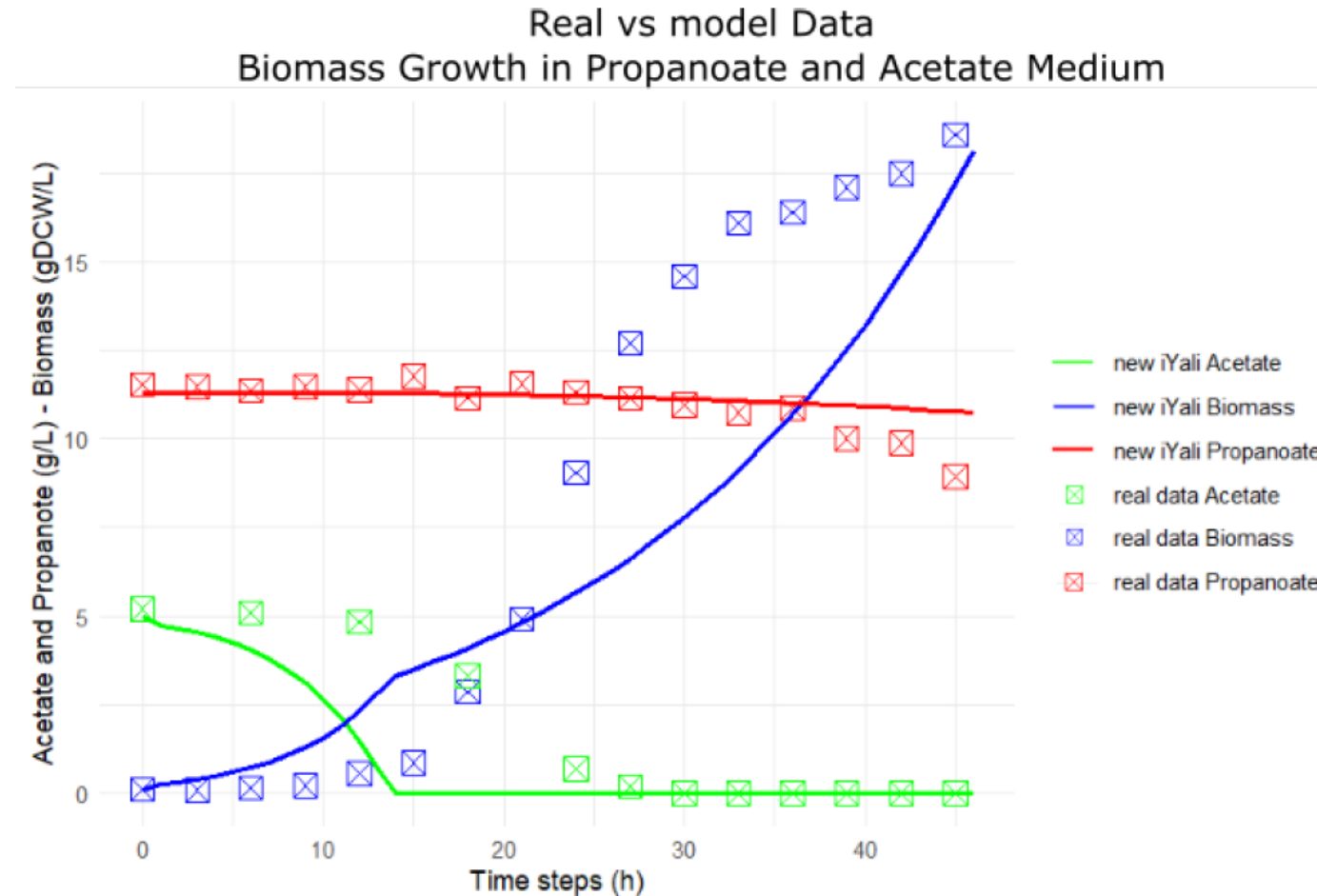
Real Data vs Model Comparison 1

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

Real Data vs Model Comparison 1

Variable	R2	RMSE	MAE
Biomass	0.919380972537853	0.908563069231937	0.665272162044825
Glucose	0.918379765591381	0.815935996969968	0.700499272430421
Propanoate	0.109170166979195	0.304422757370607	0.210529400938935

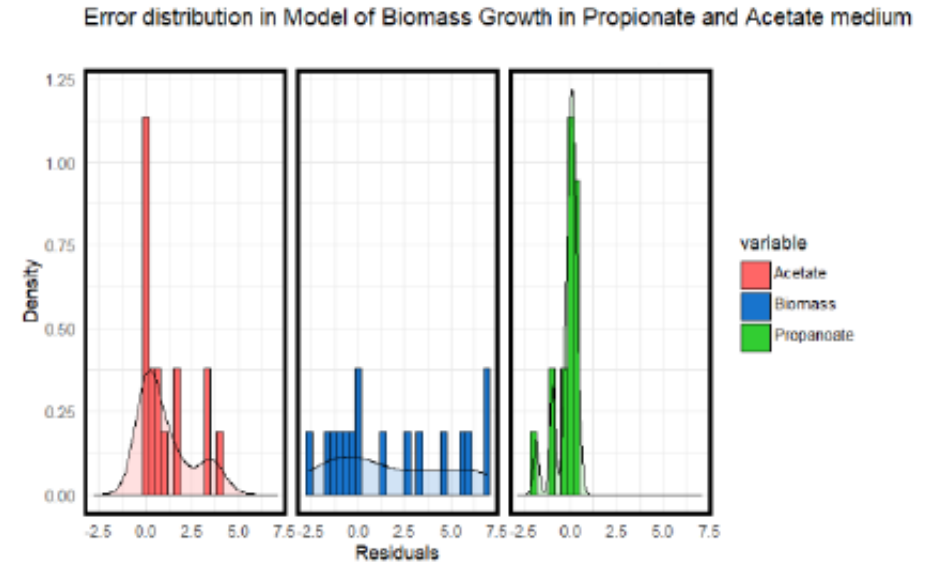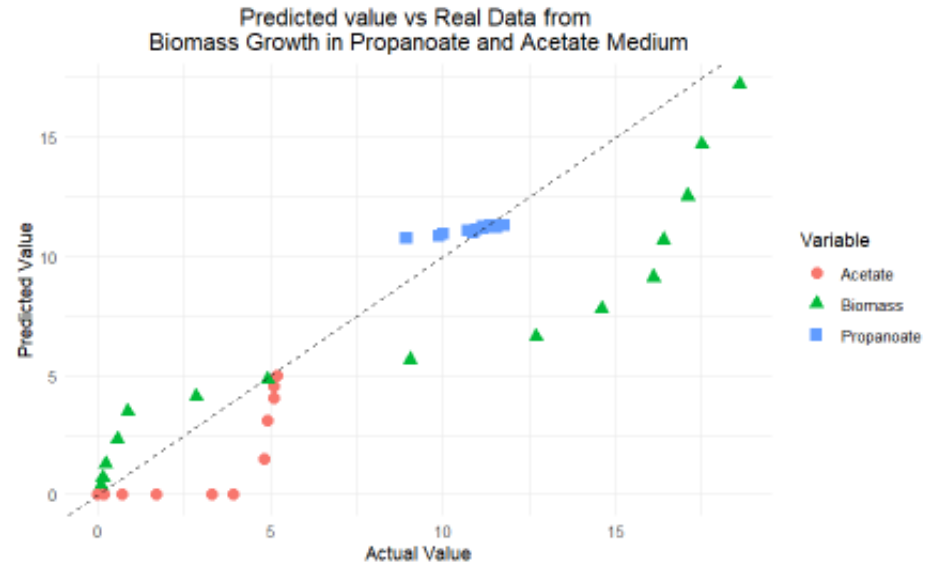TABLA 4.2: List of Performance Indicators of a Predictive Model

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

Real Data vs Model Comparison 2

Predicted value vs Real Data from Biomass Growth in Propanoate and Acetate Medium

Error distribution in Model of Biomass Growth in Propionate and Acetate medium

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

Real Data vs Model Performance

Variable	R2	RMSE	MAE
Biomass	0.919380972537853	0.908563069231937	0.665272162044825
Glucose	0.918379765591381	0.815935996969968	0.700499272430421
Propanoate	0.109170166979195	0.304422757370607	0.210529400938935

TABLA 4.2: List of Performance Indicators of a Predictive Model

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

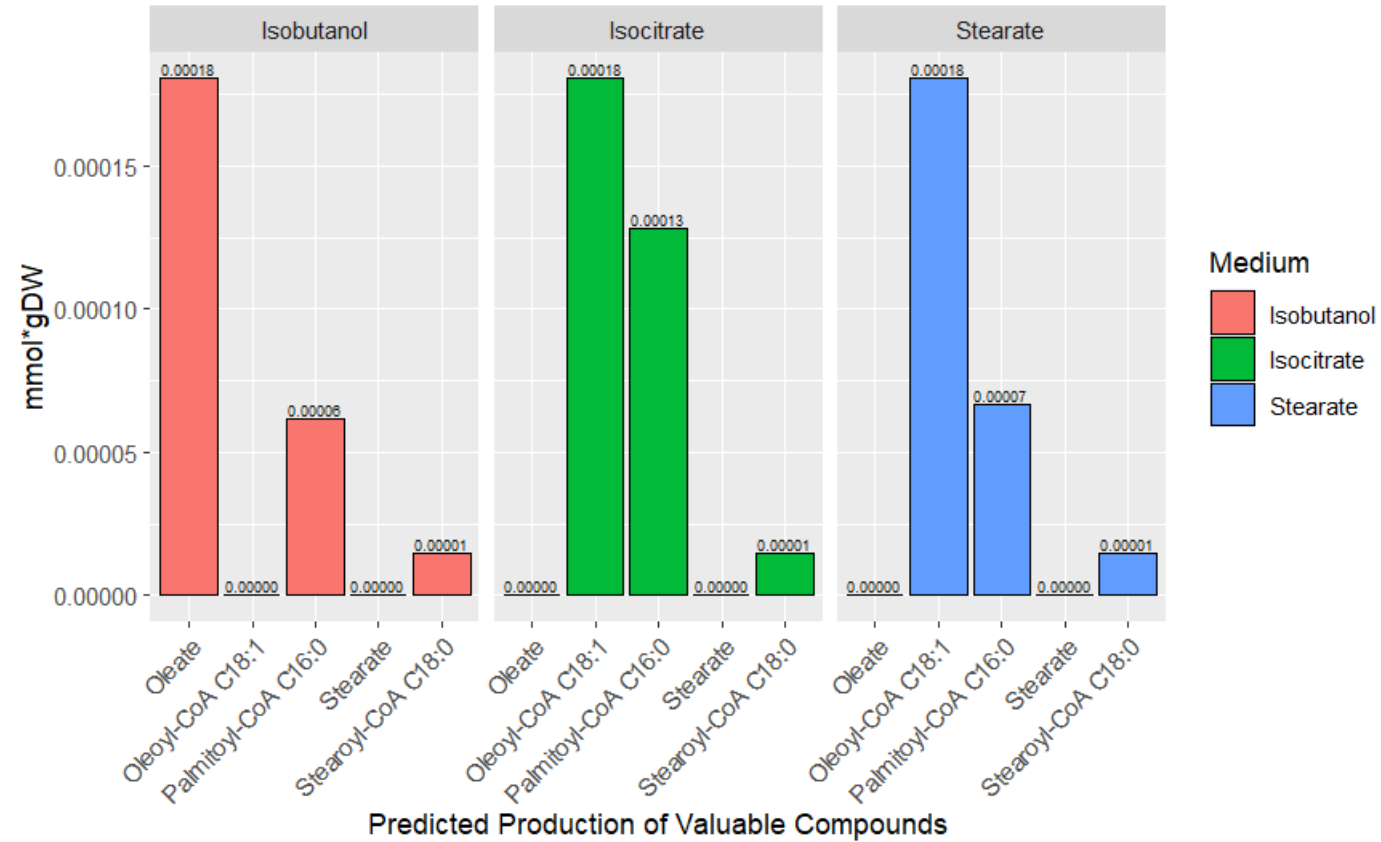TABLA 4.3: List of Performance Indicators of a Predictive Model

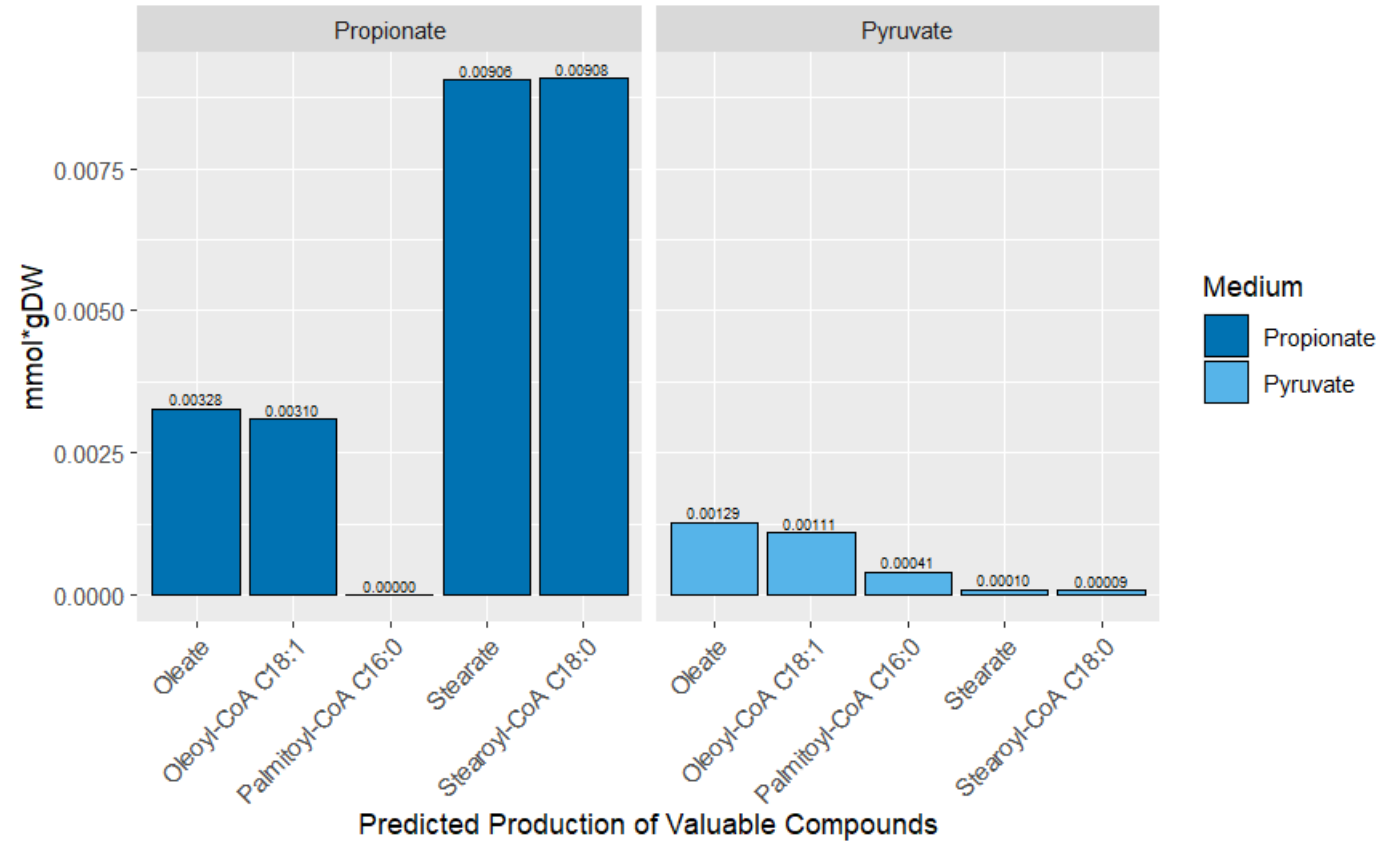# Real Data vs Modelled Production of Lipids and OCFAs
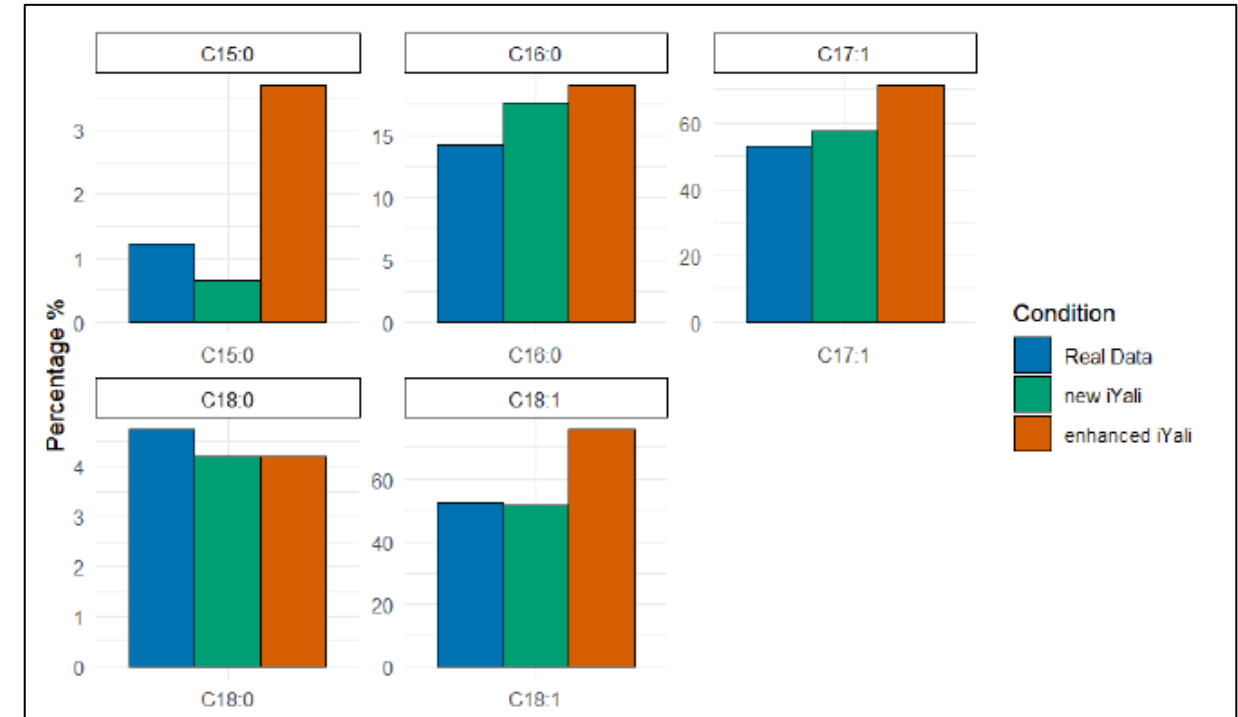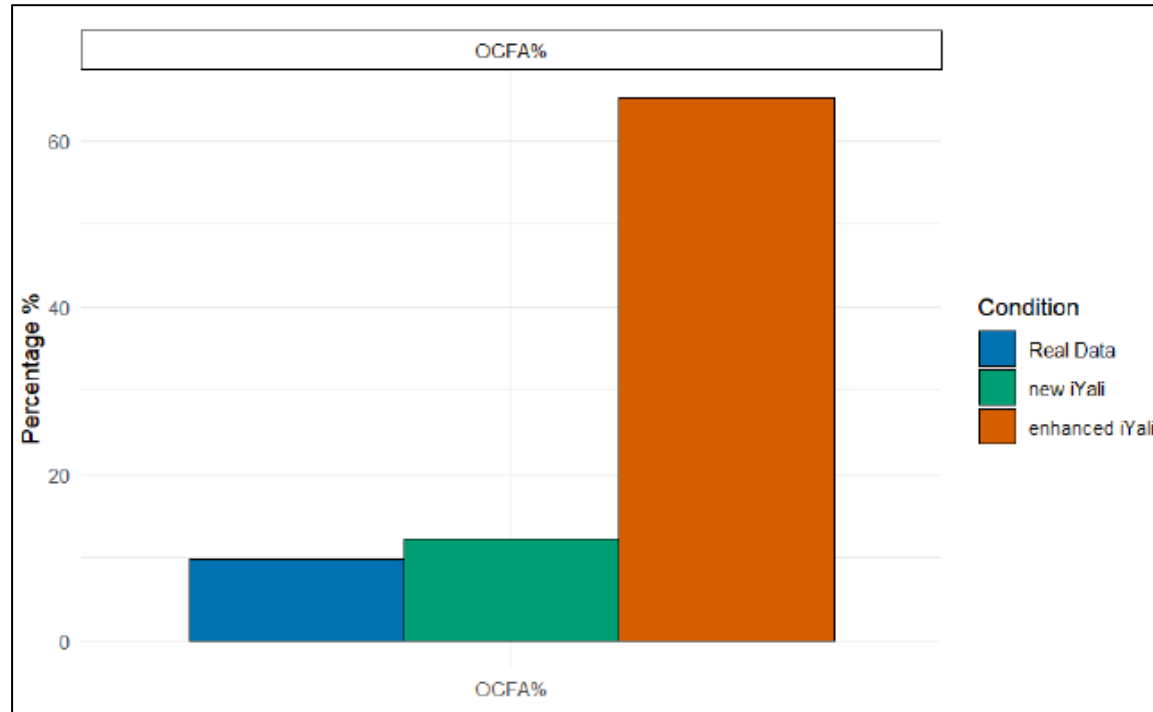
Modelled ECFA Production in different carbon sources
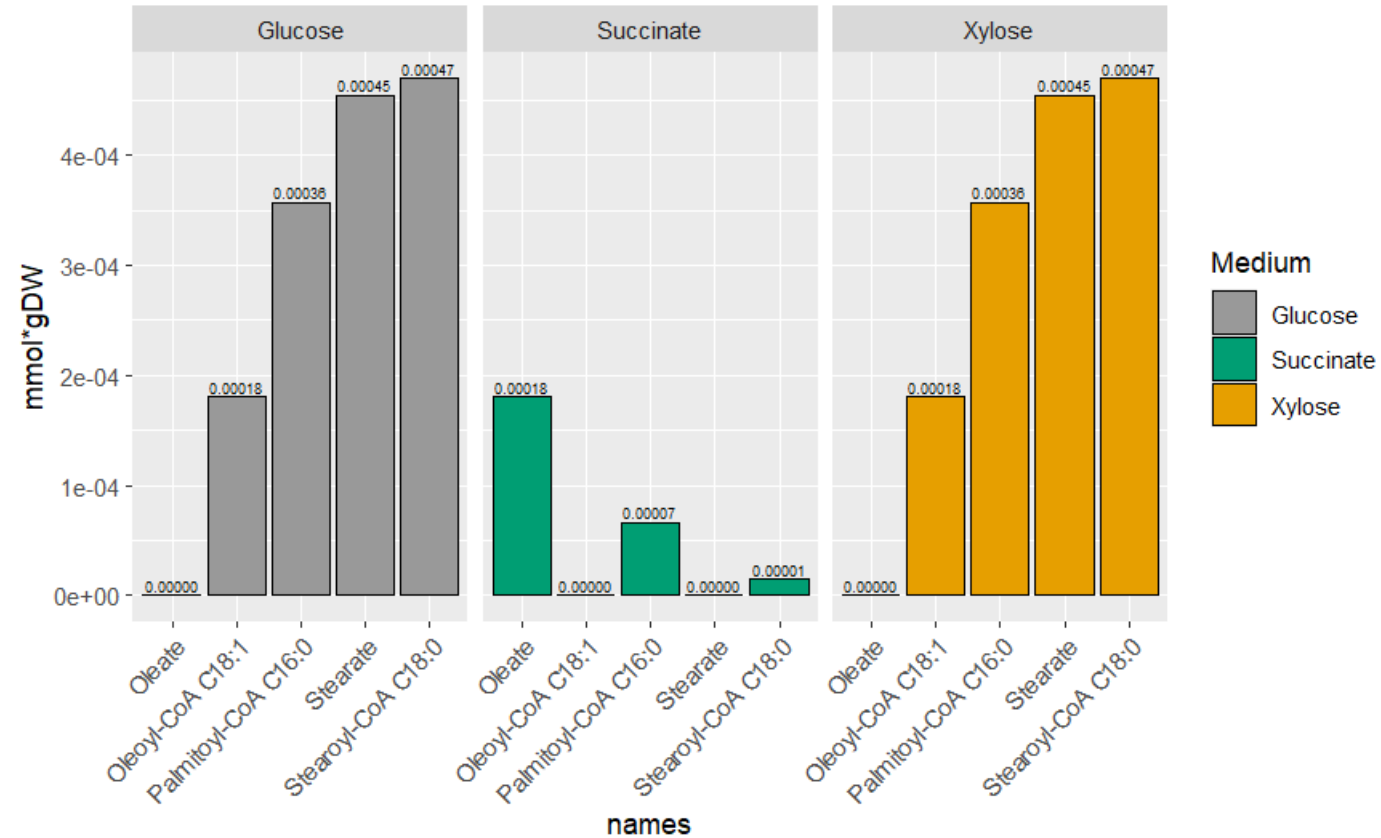
Modelled ECFA Production in different carbon sources

Modelled ECFA Production in different carbon sources

Conclusion

Conclusions

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.
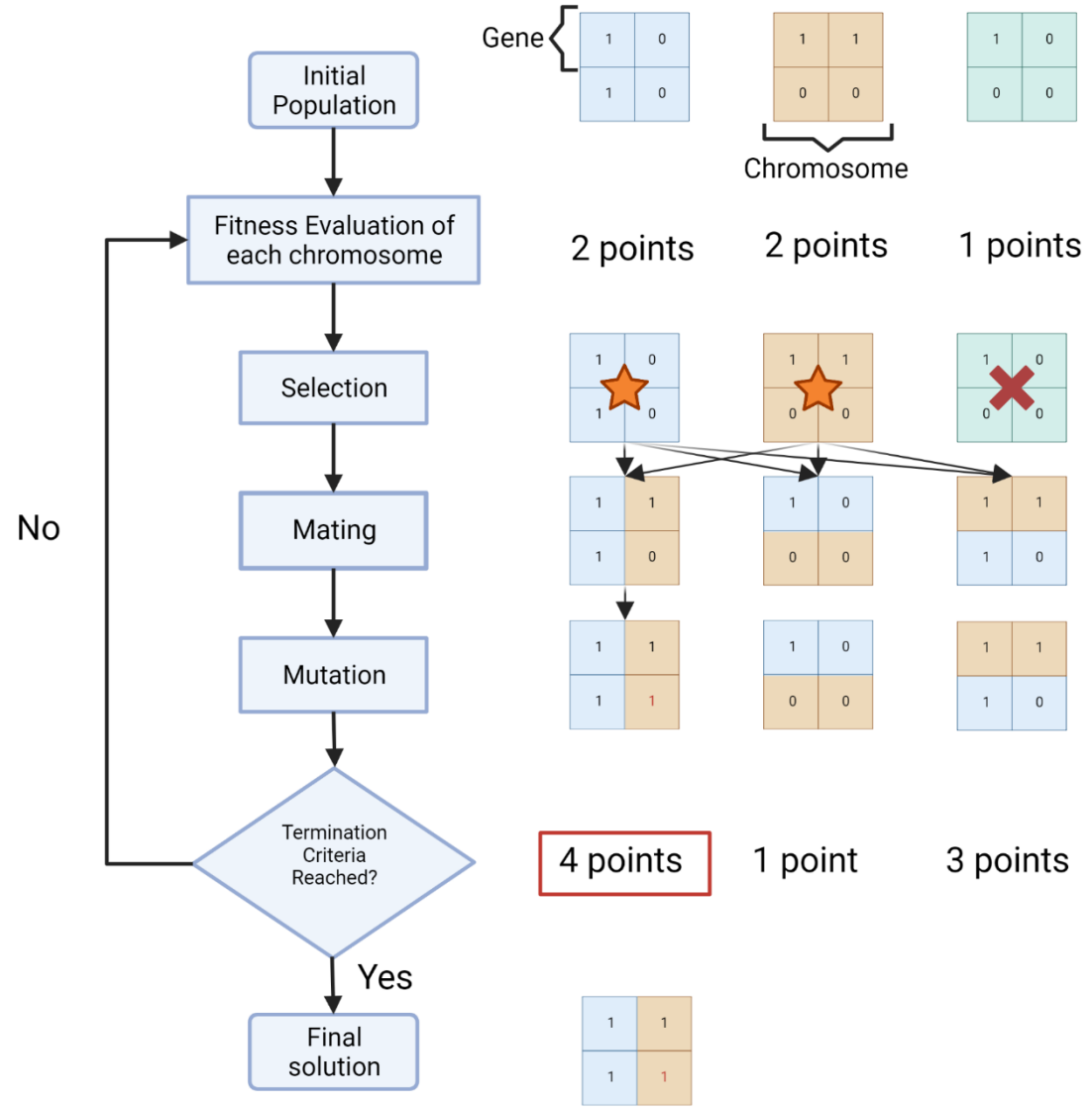
Utilizing the COBRApy toolkit, alongside FBA and dFBA methodologies,we successfully simulated a variety of genetic modifications within the new iYali model. The enhancements centred on the overexpression of key enzymes facilitating propionate metabolism towards acetate and OCFA production.

Our model demonstrated a commendable ability to simulate the organism's metabolic responses with a tolerable degree of precision
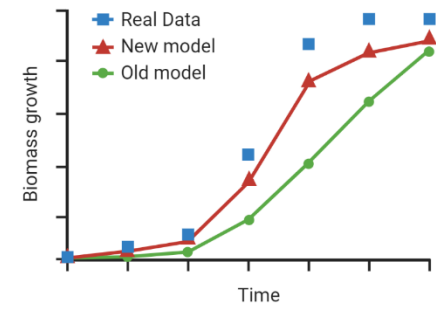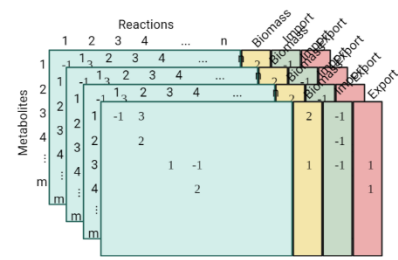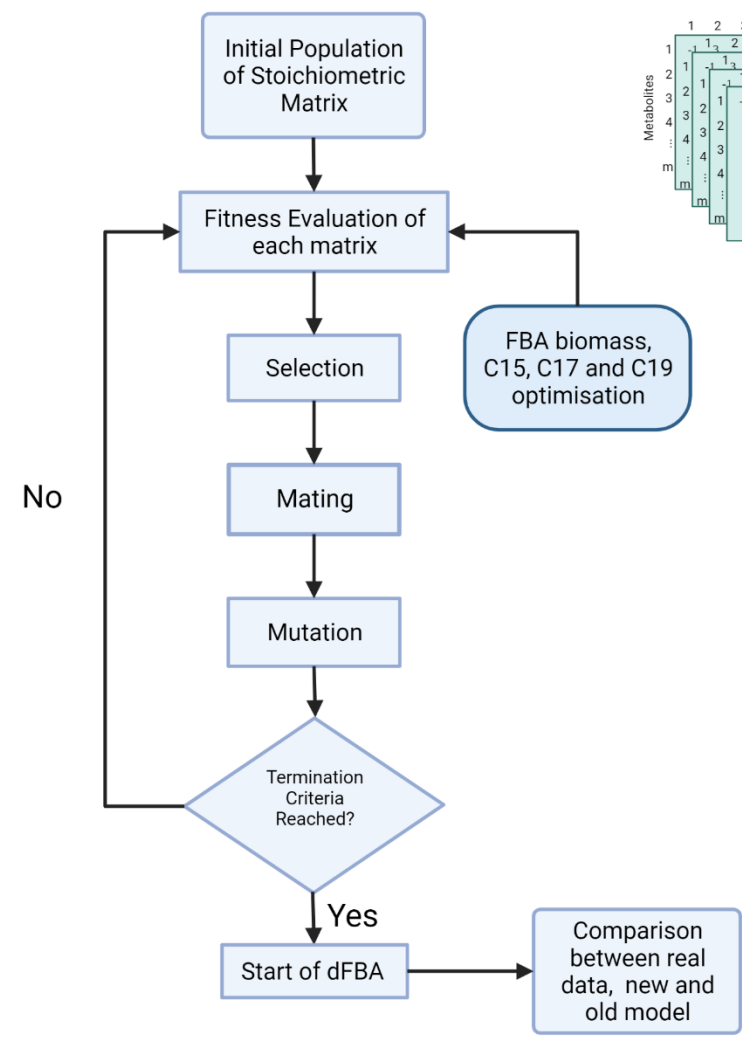
Our findings suggest that the genetic modifications yielded a 20 % increased in OCFA production, specially in C17

Recommendations

Genetic Algorithm
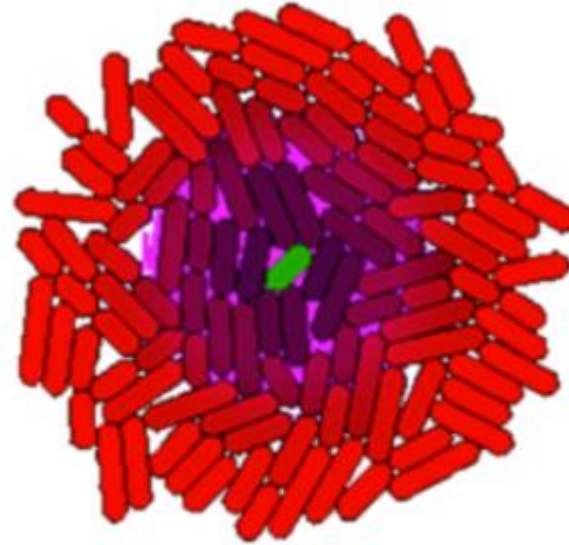

Initial Population → Fitness Evaluation of each chromosome → Selection → Mating → Mutation → Termination Criteria Reached? → (No) back to Fitness Evaluation; (Yes) → Final solution

Gene / Chromosome

2 points   2 points   1 points

4 points   1 point   3 points

Final Algorithm

Gro Production Simulation

gro

Visualize the OCFA and
Biomass Production in Real
Time