# Upgrade of a *Ustilaginomycetes's* genome-scale metabolic model for MEL production by *Moesziomyces antarcticus*

**Fátima Mendes Costa**

Thesis to obtain the Master of Science Degree in

**Biotechnology**

Supervisors:

Prof. Dr. Frederico Castelo Alves Ferreira

Prof. Dr. Nuno Ricardo Torres Faria

## Examination Committee

Chairperson: Prof. Dr. Nuno Gonçalo Pereira Mira

Supervisor: Prof. Dr. Nuno Ricardo Torres Faria

Members of the Committee: Dr. Paulo Jorge Moura Pinto da Costa Dias

Dr. Petar Keković

**November 2022**

# Preface

I declare that this document is an original work of my authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa. The work presented in this thesis was performed at the Institute for Bioengineering and Biosciences of Instituto Superior Técnico (Lisbon, Portugal), during the period September 2021 – September 2022, under the supervision of Prof. Dr. Frederico Ferreira. The thesis was co-supervised by Prof. Dr. Nuno Faria.

# Acknowledgements

I would like to start acknowledging my supervisors. To professor Frederico Fereira, who always was enthusiastic about this work and that arranged the solution to every problem. To professor Nuno Faria, who always found time to answer my (huge amount of) questions, that had a kind word when things did not work that well and who kept a positive thought through the process. Thank you both for letting me take on this challenge.

After, it is mandatory to highlight the people who were with me at the lab, to the people that accompanied and shared all of this path. Thank you, Carlota, for the support, to Gonçalo Catalão, for keeping the spirit up, to Marta, for the laughs, to Tiago, for helping in everything that I needed, and for the talks. An even huge thank you to Gonçalo Monteiro, who was present since the begging, for the help, and for the smiles. And to Ana Ramalhosa. To the person that started on the same day as me, and did most of the path by my side, I am so grateful for everything, from the lectures about models, and the guided tutorials through approaches in the lab, to the weekends in the lab. I am so proud of you all, for the work that every single one of you left in that lab. We did it guys. We have our master's degree.

To Maria Inês, who helped me at the begging of my thesis with everything in the lab, and to Petar, who showed me what is to work in the research world for years. To professor Paulo Dias, that answered all my questions related to modelling software and tools, and that lead me to find the current model to work on. Thank you to each one of you, for teaching me so many things that I required to do this thesis.

At the same time that I acknowledge people that were in the laboratory daily, my next regards go to the people that work with me, and that were so comprehensive through the last months that I was doing my internship, while I was still doing my thesis. Thank you, Angela, Cosimo, Mark, and Miguel. Thank you, Gui, Hugo, Lais, and Tiago, for all the help on the company, for so good moments, and even for helping me with modelling issues. To Moritz and Michael, thank you for trusting me and for allowing me to be an “entrepreneur”.

For the friends that I made at Técnico and for all the people that I meet along the way while organizing the SBE. Thank you for the last two years and for giving me some of the best academic experiences that I ever had. To Catarina and Sofia, thank you for the friendship and for being such a good hangout, and to Ana, Ashni, Joana, and Mariana from SBE, thank you for your patience, for letting me grow as a leader, and for the trust.

For the friends that I made at my bachelor, and that still accompanied my path until now. I am so glad for knowing you. Thank you John, for the last 5 incredible years, to my “godmother” Marta, for the kindness that she had from my first day at the university, and to my two godchildren, Balsinha and Eva, that make me so happy and proud. To Margarida, that always stick by my side through the rough

times, through study sessions, parties, trips, and workouts. You are amazing, and I am beyond grateful for having you as one of my best friends.

For my basketball athletes, but mostly, for the friends that I did on the team. What a journey it was been! It is incredible how now I cannot imagine not meeting you. To Benny, Edu, Luiz, Odair, and one of my best friends Ricardo. You all are so special to me, thank you for each one of you, for all the talks, for worrying with me, for all the good moments (that are so many). I hope this is just the begging of a lot of stories that we are going to make through the years. To Barragan, whom I know even before there was a basketball female team, thank you for not giving up on me, for all the talks, and for being you. To Bia, Liliana, Matilde, Maria, and Su, what a pleasure was to train and to play by your side, you are awesome inside the court, but even more outside of it, I hope that you all stay present after our academic paths end.

Since I already acknowledge so many of my friends, is impossible to not mention my oldest friends. The ones that are with me since I said in my first year of high school, in a physics and chemistry class, “I want to be a scientist”. It seems like I did it guys. Thank you, Catarina, I will say I can appreciate you in so many aspects of my life that is hard to highlight just one, you are one of the best people I know. Thank you, Chula, for always putting me to proof, to having my back, I adore you. Thank you Inácia, for having the coolest personality I know, you saw something in me from day 1, and thank you for trusting in me. Thank you, Tatiana, you are one of a kind. And to finish the Portimão crew, thank you to Inês, that is my friend for almost 10 years. How did that time get by so fast? You will always be my “other” small sister.

To my oldest friend Ana Barrocas. I know you since we were 5 years, and it is hard to describe our friendship in just some words. We grew a lot, together, and also apart, but the most impressive thing about this growth is that we always stayed in contact. And now, that we are starting our professional life, I look behind, and is incredible how we manage to achieve so many incredible things with a lot of work. I am truly proud of us.

To my family. To my aunts and my uncles, but mostly to my cousins. It is strange for me to be concluding my master, and my university path, but is even stranger to see how much all of you have grown. To Alice, Bia, Joana, Marta, and Sara, I love you.

I had to leave a paragraph for my grandparents. I doubt that any time you are going to read this, even so, I have to leave written how much I appreciate you. Not every day was easy living with older people, but for sure it was always funny. I have to thank you from the bottom of my heart for everything you did for me in the last 5 years, especially the fact of always staying so positive and happy for me. I am glad that I had the opportunity to have spent my academic years with you.

To Isabela, who supports me in everything, no matter what, and knows how challenging was for me to make this thesis. Thank you even for the days when we do not agree, I know you will stay by my

side for many more years and I am looking forward to so many things that are yet to happen together. You are the special one.

To Miguel. Thanking you will never be enough for the backup that you gave me during the last year. You were the person that understood better my struggles, and that helped me pursue what I needed and wanted, despite not always being the easiest solution. Thank you for the words, for the care, for the comprehension, and for wanting the best for me. I am so grateful for having you in my life.

And for the end, I left the three people that made all of this possible. All of these experiences through the years, who believed in me from day one and never doubt that I was not capable of doing it. Thank you to my parents, for allowing me to live my academic years with so many good memories, for being present on a daily basis, and for the constant support through the constant challenges that appeared along my academic journey. Thank you, mum, for all the talks, and for understanding with the greatest compassion all difficulties. Mostly, I have to thank my sister, for being my best friend, for supporting me in every decision, even when she disagreed, and for helping me when I needed it. I am so proud of being your sister and I am anxious to see you on the day you deliver your thesis. Is going to be remarkable.

Finally, I would like to thank for the funding received from the FCT granted to the Research Unit iBB - Institute for Bioengineering and Biosciences (UIDB/04565/2020 and UIDP/04565/2020) and to i4HB - the Associate Laboratory Institute for Health and Bioeconomy (LA/P/0140/2020).

# Abstract

Biosurfactants are gaining attention because they are biologically obtained and can substitute for fossil oil-driven surfactants. One important type of glycolipid biosurfactant is the mannosylerythritol lipid (MEL), produced mainly by *Ustilaginomycetes* and *Moesziomyces* spp. Optimal use of the metabolism potential of these organisms can help in biotechnological development to make MEL production competitive. This was achieved with an *in silico* overview and accurate predictions of the experimental situations.

Since *Moesziomyces* metabolic network is not coded in any available database, a recent genome-scale metabolic model for metabolic activity modelling of *Ustilago maydis* needs to be updated for use with *Moesziomyces* strains, as these producers of MEL have high genome similarity. The majority of the metabolic problems in the model are related to specific metabolite production that does not occur in *Moesziomyces* and the presence of dead-end reactions. MATLAB software and the COBRA toolbox were used to obtain a more accurate model.

Exponential batch and fed-batch cultivations with glucose were employed to test growth and MEL production. These values were given as input to the model, which predicted values with higher error values than *U. maydis* values, by applying the constraint-based model approach. With more experimental data and other analytical techniques, this upgraded model will enable both biotechnological applications and the investigation of metabolic responses to various environmental conditions.



**Keywords:** Metabolic modelling; Flux Balance Analysis; Mannosylerythritol lipids; *Moesziomyces*; MATLAB*; Ustilago Maydis*.

# Resumo

Os biosurfactantes estão a ganhar atenção porque são obtidos biologicamente e podem substituir os surfactantes derivados de combustíveis fósseis. Um tipo importante de biosurfactante glicolipido é o Manosileritritolípidos (MEL), produzido principalmente por *Ustilaginomycetes* e *Moesziomyces spp.* A utilização otimizada do potencial metabólico destes organismos pode ajudar ao desenvolvimento biotecnológico, tornando a produção de MEL competitiva. Recorrendo ao uso de simulações *in silico,* foi possível obter previsões para situações experimentais.

Uma vez que a rede metabólica de *Moesziomyces* não está codificada em nenhuma base de dados disponível, um modelo metabólico recente à escala do genoma para modelação da atividade metabólica de *Ustilago maydis* precisa de ser atualizado para utilização com estirpes de *Moesziomyces*, uma vez que estes produtores de MEL têm uma elevada semelhança genómica. A maioria dos problemas metabólicos encontrados no modelo estão relacionados com a produção de metabolitos específicos que não estão presentes no metabolismo de *Moesziomyces* e com a presença de reações incompletas. O software MATLAB e a *toolbox* COBRA foram utilizados para obter um modelo mais preciso.

Fermentações com glucose como fonte de carbono, foram realizadas para testar o crescimento e a produção de MEL em *M. antarcticus*. Os valores obtidos foram dados como *input* ao modelo, que previu valores de erro superiores aos valores de erro obtidos com *U. maydis*, aplicando a abordagem do modelo baseado em restrições. Com mais dados experimentais e outras técnicas analíticas, este modelo atualizado permitirá várias aplicações biotecnológicas, como a investigação de respostas metabólicas a várias condições ambientais.

**Palavras-chave:** Modelação metabólica; Análise de Balanço de Fluxo; Manosileritritolípidos; *Moesziomyces*; MATLAB*; Ustilago Maydis*.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| Ac | Acetyl |
| AcCoA | Acetyl-CoA |
| ATP | Adenosine triphosphate |
| $CO_2$ | Carbon dioxide |
| COBRA | Constraint-based Reconstruction Analysis Toolbox |
| DCW | Dry cell weight |
| FA | Fatty acids |
| FBA | Flux Balance Analysis |
| g | Gram |
| g $L^{-1}$ | Gram per litre |
| GC | Gas Chromatography |
| GSMM | Genome-scale metabolic model |
| $h^{-1}$ | Per hour |
| HPLC | High Performance Liquid Chromatography |
| iUma 22 | First high-quality genome-scale metabolic model for *Ustilago maydis* |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| L | Litre |
| MAC2 | Mannosylerythritol-acyl-transferase |
| MATLAB | Matrix Laboratory sofware |
| ME | Mannosylerythritol |
| MEL | Mannosylerythritol Lipids |
| *M. antarcticus* | *Moesziomyces antarcticus* |
| *M. bullatus* | *Moesziomyces bullatus* |
| NADPH | Nicotinamide Adenine Dinucleotide Phosphate |
| NMR | Nuclear Magnetic Resonance |

| | |
|---|---|
| OAA | Oxaloacetate (Oxaloacetic acid) |
| PPP | Pentose Phosphate Pathway |
| TCA | Tricarboxylic Acid Cycle |
| TLC | Thin Layer Chromatography |
| *U. maydis* | *Ustilago maydis* |

# 1. Scope and Motivation

A vast variety of products, including cosmetics, agricultural chemicals, household detergents, and pharmaceuticals are produced using surfactants. Microbial surfactants, also known as "biosurfactants," are environmentally friendly substitutes for conventional surfactants. As environmental consciousness and emphasis on a sustainable society that works in unison with the environment have increased, natural surfactants have become increasingly important. [3] The increase in the bioavailability of insoluble nutrients, the facilitation of microbial attachment to hydrophobic surfaces, and antibiotic activity are some benefits of biosurfactants, according to the research.

The process of biosurfactant production is affected by knowledge gained in metabolic pathways. The objective of this work was to study and upgrade an available *Ustilago maydis* genome-scale metabolic model into a *Moesziomyces antarcticus* model, with a focus on the biosurfactant Mannosylerythritol lipids (MEL) production, knowing that yeast strains of the genus *Pseudozyma* (currently *Moesziomyces*) and *Ustilago* abundantly produce MELs. This approach will lead to the development of a cultivation strategy that could turn MEL production cost-competitive with fossil-driven surfactants in the future, namely by mixing carbon sources of different origins (vegetable oils and sugars) to improve MEL titres and facilitate downstream.

This should be achieved using an *in silico* approach by designing a metabolic flux model that describes the main metabolic pathway used for MEL production, with the target of the pathways for the production and assembly of MEL building blocks. Using this model, the flux distribution at branch points was established by collecting experimental data on substrate consumption and MEL production from hydrophilic (hexoses and pentoses, such as glucose) substrates via glycolysis, gluconeogenesis, tricarboxylic acid cycle, pentose phosphate pathway, fatty acid synthesis, and *β*-oxidation to generate MEL building blocks. As a result, bottlenecks in MEL production will be recognized, and reasonable solutions for producing this biosurfactant will be developed. The final goal is to make biosurfactant economic costs competitive with fossil oil-based surfactants by increasing substrate carbon use efficiency.

## 1.1. Research questions

The metabolite of interest in this work, mannosylerythritol lipid, is a biosurfactant, namely a glycolipid biosurfactant, which is of growing importance in a different set of industries, because the production yields obtained are substantially higher than those of other forms of biosurfactants derived from renewable resources. Glycolipid biosurfactants are the most promising for commercial manufacturing and use; however, this component is difficult to assess because of its high production costs.

By effectively synthesizing microbial glycolipids for biomedical applications and other uses, the knowledge obtained through the metabolic engineering of microbial lipids for biofuel generation can be applied to lower production costs. A recently released (March 2022) *in silico* model (iUma22 model) obtained for *Ustilago maydis* is, to date, the only available genome-scale metabolic model for the

*Ustilaginomycete* class of organisms. Because the *M. antarcticus* strain does not have any available model, utilising this updated and complete model, even from a different species, is the best way to study the metabolism of this organism, considering that the MEL synthesis pathways are present in the iUma22 model.

## 1.2. Research strategy

Using the *in silico* model obtained from the literature, the results were structured according to three principal objectives:

- Study the *in silico Ustilago maydis* model and the respective pathways. (Chapter 5.1.)

This was possible through the identification of reactions and the respective pathway visualization and parametrization of the model using the flux balance analysis (FBA) approach. Modification to this model as deletion of reactions that do not exist in *Moesziomyces antarcticus*, and addition of reactions that are necessary to achieve a more complete metabolism was also performed.

- Study of possible *Moesziomyces* strains and carbon sources to feed yeast and the respective growth rate, substrate uptake rate, and MEL production rate values. (Chapter 5.2.)

This was achieved in three different experiments. The first experiment had the objective of choosing the carbon source and the strain for use in the following experiments and was realized with three different carbon sources (galactose, glucose, and lactose) and two different strains (*M. antarcticus* and *M. bullatus*). The other two fermentations were assessed to study the effects of the chosen carbon source, glucose, on growth and MEL production. One batch experiment with different initial glucose concentrations was used to obtain *M. antarcticus* growth rate and substrate uptake rate values, which were compared with values from the literature obtained with *U. maydis*. The other experiment, fed-batch fermentation, had different characteristics, with equal initial glucose concentration and addition on day 4 of different glucose concentrations. In this experiment, the values of lipids and MEL production were achieved to later be possible to have the *in silico* model optimized for MEL production.

- Fitting of experimental results to the iUma22 model and to the modified model. (Chapter 5.3.)

To understand the accuracy of the predictions of the iUma22 model for the results obtained with *M. antarcticus*, the growth-rate values and the MEL production values obtained experimentally were compared with values predicted by the model, with and without modifications. In this step was also analysed, by pathway visualization, which routes are active under the optimization of biomass reaction or having MEL production as the objective.

# 2. Introduction

## 2.1. Biosurfactants

Nowadays it is important to have a sustainable alternative to fossil-driven surfactants, since surfactants' manufacture, usage, and disposal can affect the environment. Surfactants are utilised in the production of a wide range of products, including cosmetics, agricultural chemicals, home detergents, and medications. Delivering a more environmentally friendly surfactant alternative to chemical surfactants is possible with microbial surfactants, classified as “biosurfactants”, that were initially recognized as extracellular amphiphilic molecules among natural surfactants.[4]

Microbial-based surfactants are characterised as emulsifying biomolecules that are surface-active and can be an adequate substitute and solution to the use of fossil oil-based surfactants. Biosurfactants are assigned to five major classes, as low molecular weight compounds, including glycolipids (rhamnolipids, sophorolipids, mannosylerythritol lipids, and trehalose lipids) and high molecular weight biosurfactants like polymer type (emulsan). [5] The other types of biosurfactants are lipopolysaccharides or lipopeptides (surfactin, serrawettin), fatty acid type, and particulate surfactants. Low-molecular mass biosurfactants are efficient in reducing surface and interfacial tensions, while the ones with higher molecular mass (such as emulsans) are better in bioemulsification. [6] [7]

Natural surfactants have become increasingly essential as environmental awareness and the emphasis on a sustainable society in harmony with the global environment has grown.[4] According to the literature, some of the advantages of biosurfactants are the increase in the bioavailability of insoluble nutrients, allowance of adhesion of microorganisms to hydrophobic surfaces, and antibiotic activity. [8]

The interfacial features of surfactants are mostly reflected in their structures. Furthermore, biosurfactants are produced stereo-selectively from biomolecules via enzymatic processes, the majority of which are chiral compounds with a unified molecular arrangement. As a result, biosurfactants can demonstrate superior orientation and packing properties in their interfaces. These structural characteristics enable biosurfactants to conduct actions that are not possible with typical chemical surfactants.

### 2.1.1. Glycolipids

Glycolipids are characterised as amphiphilic molecules, with a hydrophobic and a hydrophilic group, namely lipid, and glycosyl moieties. These chemicals perform critical roles in the exchange of energy, substances, and signals across a variety of interfaces, as well as the organisation of biological systems. They have one or more functional groups and chiral centres, complex structures, high surface activity (due to low critical micelle concentration), biological activity, high biodegradability, and low toxicity.[4] The chemical structures of some representative glycolipid biosurfactants are presented in Figure 1.

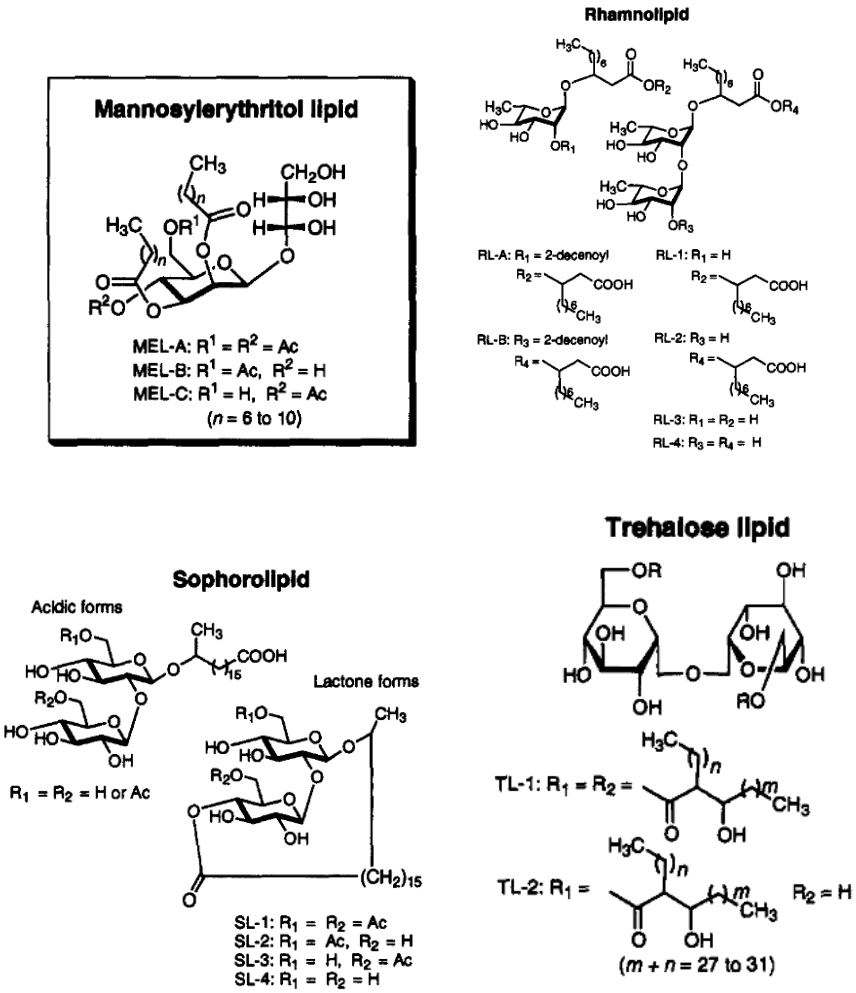



**Figure 1** – Chemical structures of main glycolipid biosurfactants. [4]

Because their production yields are substantially higher than those of other forms of biosurfactants, and as they can be generated from renewable resources, namely from carbohydrate biomass, glycolipid biosurfactants have received the greatest attention. As a result, glycolipids biosurfactants are the most promising for commercial manufacture and use. [4]

There are several classifications for glycolipids, the most useful of which is to divide them into simple and complex glycolipids (Figure 2). Simple GLs (SGLs), also known as saccharolipids, are two-component glycolipids with direct links between the glycosyl and lipid moieties. Complex glycolipids (CGLs) are structurally more heterogeneous, as they comprise residues such as glycerol (glycoglycerolipids), peptide (glycopeptidolipids), or other residues in addition to the glycosyl and lipid moieties. Mannosylerythritol Lipid is one of the SGLs addressed with natural microbial origin. [9]

Nonetheless, glycolipids, due to their high production costs, are hardly assessing the market. The knowledge gained from the metabolic engineering of microbial lipids for biofuel production can be used for the effective synthesis of microbial glycolipids for biomedical applications and other uses, diminishing the production costs problem.

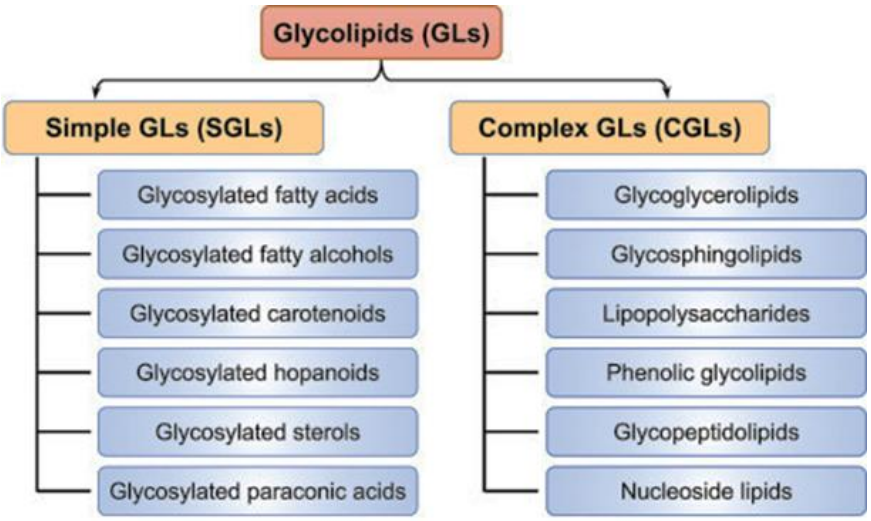



**Figure 2** – Classification of glycolipids. Simple glycolipids (SGLs) comprise glycolipids consisting of glycosyland lipid residues only. Complex glycolipids (CGLs) contain different residues in addition to the glycosyl and lipid residues. [9]

## 2.1.2. Mannosylerythritol Lipid

Mannosylerythritol lipid (MEL) is classified as a glycolipid type of biosurfactant, constituted by a hydrophilic sugar, 4-O-*β*-D-mannopyranosyl-D-erythritol, that compresses a mannose and an erythritol residue and a hydrophobic tail, consisting of two fatty acid chains. These chains are in the C2 and C3 of the pentose sugar structure of mannose (Figure 3) and according to the degree of the carbon acetylation at C4 and C6 position of mannose, it is possible to obtain four MEL homologs. The degree of acetylation and the chirality of erythritol are used to classify MELs as four distinct variants: MEL-A, MEL-B, MEL-C, and MEL-D, ordered from the most hydrophobic to the most hydrophilic, since MEL-A has a di-acetylated structure and MEL-D a non-acetylated one. The length of each fatty acid residue is influenced by the respective producer organism, in this work, the *Moesziomyces* yeast (previously referred to in the literature as *Pseudozyma*), of *Ustilaginaceae* family. Furthermore, mono-acetylated MELs, MEL-B and MEL-C, which are deacetylated MEL-A, are the primary products of *Moesziomyces* species such as *M. tsukubaensis*, *M. hubeiensis*, and *M. graminicola*. [10,11] [12]

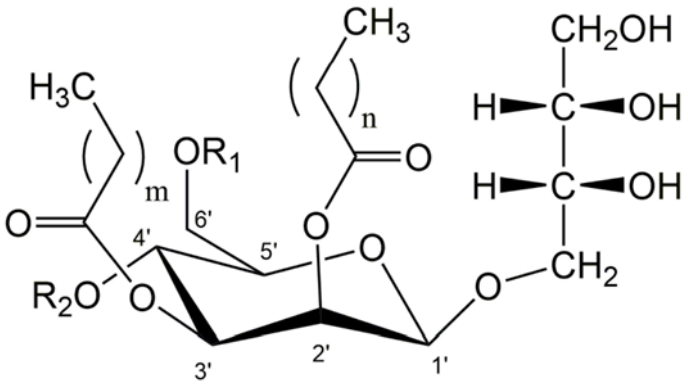



**Figure 3 –** General structure of di-acylated mannosylerythritol lipid. (MEL-A: R1=R2=Acetyl; MEL-B: R1=Acetyl ,R2=H; MEL-C: R1=H, R2=Acetyl, MEL-D: R1=R2=H). Variable chain-length and saturation of fatty acid side-chains at C2 (m= 2–16) and C3 (n= 2–10). [11]

This compound has a variety of applications, making it a much-studied product. MEL can be employed in a variety of contexts, including the agrochemicals industry, pharmaceutical industry, cosmetics industry, and even as surface-modifiers in bioplastics. In terms of molecular properties, MEL can alter cancer cell signalling pathways, increase therapeutic efficacy, and interact with membrane cells, suppressing cell receptors, and also have antimicrobial activity. [13] This microbial glycolipid may also change membrane conformation, giving MEL self-assembling and phase-behaviour characteristics.

To obtain MEL production it is common to have a carbon source that has hydrophobic properties. Examples of the most often used substrates are plant oils (soybean, sunflower, rapeseed), sugars or glycerol, that enhance the growth of biomass and, sometimes, of MEL production. The process engineering of MEL production, however, lacks some knowledge, with few research on the characterization of *Ustilaginaceae* fungal growth factors such as growth rates, substrate consumption, biomass yield, or oxygen demands. As so, it would be interesting to get a deeper insight into how those growth parameters and the resulting biomass concentration are related to successive MEL production.[11]

#### 2.1.2.1. Characterization of MEL and Lipids

The great structural variability of MEL demands the use of many approaches to characterise the structures and quantify the amount of MEL produced in yeast fermentations. Gas chromatography (GC), high performance liquid chromatography (HPLC), mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy, solvent extraction or thin-layer chromatography (TLC) are all widely used procedures.

Gas chromatographic analysis are used to determine the fatty acids incorporated into the side chains of pure MELs and clarification of the structure can be obtained by mass spectrometry. Following the fatty acid profiles, the MEL structures, including their molecular weight, the acetylation pattern, and the fatty acid residue pairings, are possible to investigate through MS technique. [14] Thus, GC-MS is used to assess the fatty acid content of the substrate and MEL product. [15]

High-performance liquid chromatography approach is frequently employed among different research groups for more reliable quantification of MEL. Organic solvents are used to extract MEL and fatty acids or triglycerides from the culture broth, and the organic phases are collected. Following that, the hydrophobic components are separated on an HPLC column and can be analysed, with a distinct peak for each component determined by the retention time. Thin-layer chromatography has traditionally been used to classify MEL, being a qualitative method that allows separation of the four MEL variants (MEL-A, MEL-B, MEL-C, and MEL-D). The TLC method has only been significantly modified by different researchers over the years, and it still relies on MEL staining on TLC plates. This method is still routinely employed today for screening novel organisms and discriminating between MEL variants generated.[16]

Finally, NMR analysis of MEL can be used to establish the detailed structural type of MELs produced by each strain in study, so the structures deduced using GC, MS, or HPLC are confirmed. Typically, as it is shown in Figure 4, resonances of $^{1}H$ NMR and $^{13}C$ NMR are assigned to the atoms in the MEL structure, according to the different values of chemical shifts obtained. [17]

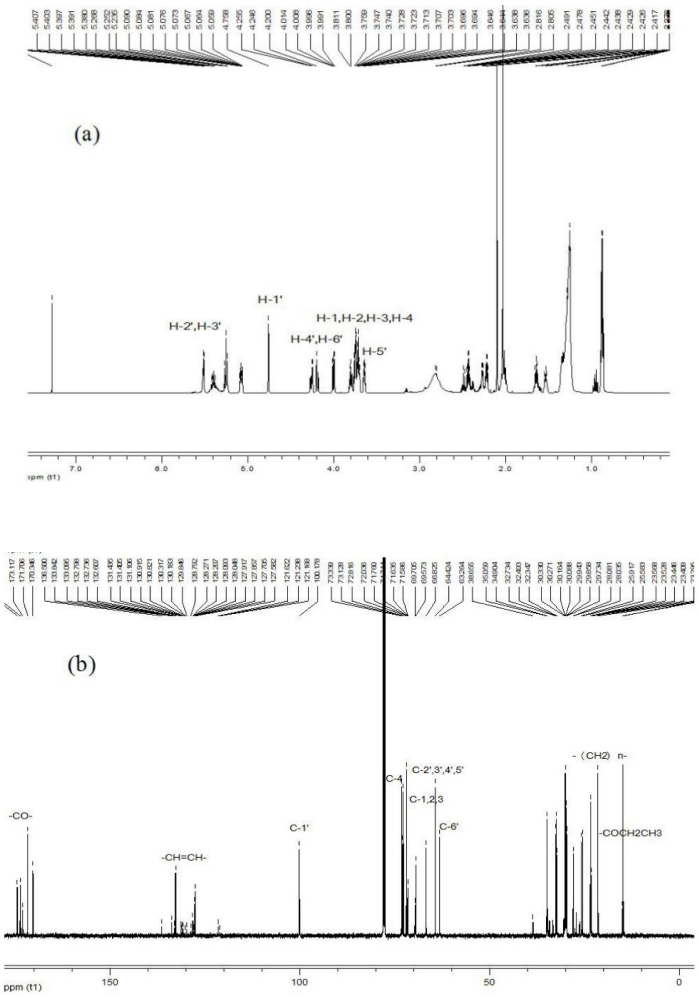



**Figure 4** – $^{1}H$ NMR spectrum and $^{13}C$ NMR spectrum of MEL-A. (a) $^{1}H$ signals at 0 to 7.0 ppm and (b) $^{13}C$ signals at 0 to 180.0 ppm. [17]

## 2.2. MEL producers

The use of yeasts in bioprocesses can be considered one of the most relevant tactics in industrial biotechnology. Microorganisms of the *Ustilaginaceae* family are regarded as a viable source for numerous biotechnologically value-added compounds. These basidiomycetous fungi, which include both plant-pathogenic smuts and non-pathogenic yeasts, have been demonstrated to produce a variety of enzymes, carbohydrates, lipids, organic acids, and biosurfactants that could be commercially valuable. When compared to bacteria strains such as *Bacillus sp*., which are conventionally identifiable by lipopeptide biosurfactants, yeasts belonging to the *Candida* (currently referred to as *Moesziomyces*) genus and *Ustilaginaceae* family are extremely attractive, since they can reach high levels of biosurfactants obtained in diverse fermentation procedures. [11], [18]

Over the last few decades, mannosylerythritol lipid biosurfactants have aroused interest and, according to studies, MELs are generally produced by these anamorphic basidiomycetous yeasts, particularly by the fungi *Ustilago maydis*, in relatively low levels, and *Moesziomyces spp.* (previously *Pseudozyma spp.*), in relatively high numbers, in terms of substrate, fermentation environment, and

downstream processing.[15][19] MEL production is highly related to the taxonomic range of these fungi, the ones that generate primarily MEL-A, such as *P. antarctica*, *P. aphidis*, *P. rugulosa*, and *P. parantarctica*, are closely related on a phylogenetic tree branch, as can be seen in Figure 5. [20]

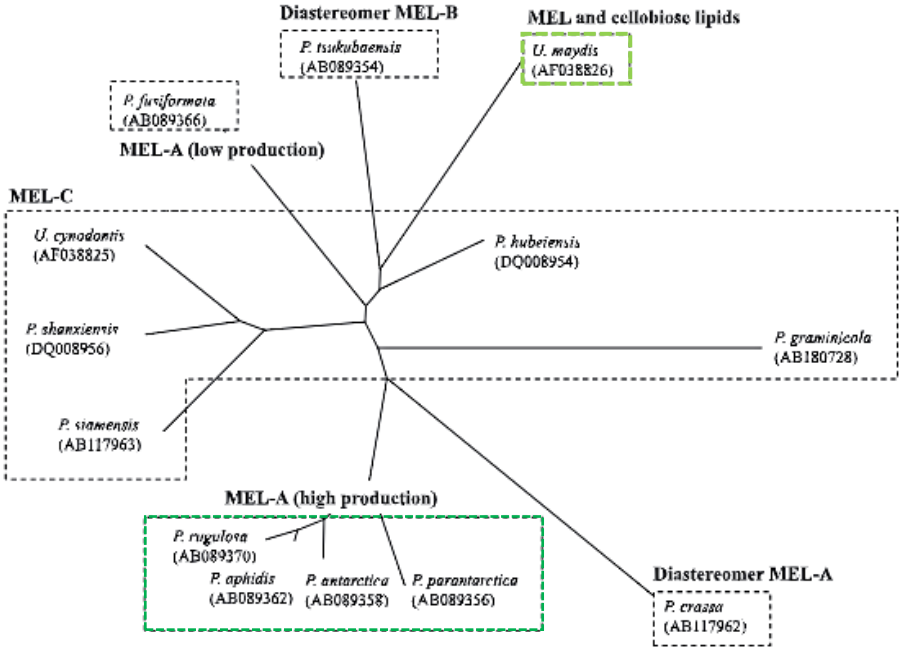



**Figure 5** – Molecular phylogenic tree constructed using rRNA gene and sequences of the genus *Pseudozyma* and *Ustilago*. *P. rugulosa*: currently referred as *M. rugulosus; P. aphidis*: currently referred as *M. bullatus*; *P. antarctica*: currently referred as *M. antarcticus; P. parantarctica*: currently referred as *M. parantarcticus*. [20]

### 2.2.1. *Ustilago*

As previously mentioned, mannosylerythritol lipids are mostly synthesised by strains of the genus *Moesziomyces*, although they are also produced by *Ustilago maydis*. In fact, the MEL biosynthesis pathway was first discovered in the fungus *U. maydis* and was just later discovered in *Moesziomyces/Pseudozyma* spp. Within the *Ustilaginomycetes* class, *U. maydis* is a model organism for plant pathogenic smut fungus. [16]

Corn smut disease (“huitlacoche”) is caused by this basidiomycete fungus, which causes tumours to grow on the cob of the maize plant. The fungus *Ustilago maydis* has a dimorphic life cycle: during the saprophytic stage, it grows as budding yeasts (haploid); when two compatible cells mate, the dikaryotic phase (mycelia) emerges, invading the host tissues and creating tumours (galls) where spore generation occurs. [21]

The genome of *U. maydis* has been thoroughly described, and genetic tools for its molecular biological study as a model microorganism of plant pathogens have been established. Comparative genomic and transcriptome studies between *U. maydis* and a strain belonging to *Moesziomyces* species, *M. antarcticus*, have indicated that the gene expression pattern of *M. antarcticus* differs

dramatically of *U. maydis* under certain conditions, particularly when oil is used as a substrate in MEL biosynthesis. According to this study, even if the gene expression profiles of the two species diverge, they are closely related at the genome level, since the genome organisation and gene content are almost identical. [12] This is relevant to its oleaginous character and in Figure 6 is highlighted the phylogenetic relationship of the two species.

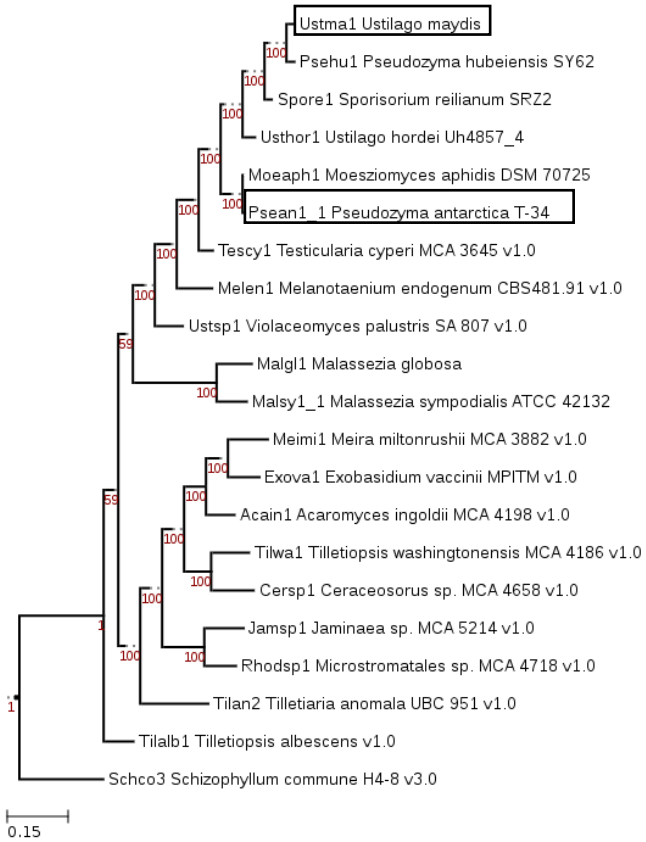



**Figure 6** – Maximum likelihood phylogeny showing phylogenetic positioning of *Ustilago maydis* and *Pseudozyma antarctica* T-34 (currently referred as *Moesziomyces antarcticus*) strains. [22]

Since it has been demonstrated that MEL biosynthesis can be customised by the complementation of genes from other related species, genome sequencing is a powerful method for identifying probable orthologs of MEL biosynthesis genes in other species. [23] The *U. maydis* genome (20.5 Mb) is larger than the *M. antarcticus* genome (18.0 Mb, predicted from the size of total length of scaffolds). While the genomic sequence of *M. antarcticus* was assembled together into 27 scaffolds, *U. maydis* contains 23 chromosomes. Despite only having an average amino acid identity of 72.0% for predicted proteins, the genomes of *M. antarcticus* and *U. maydis* show a remarkable degree of synteny. There was considerable homology between 5987 *M. antarcticus* genes and 5707 *U. maydis* genes, as well as 5,707 of the *U.maydis* genes showed significant homology to gene(s) of *M. antarcticus*, with an overall estimated number of 5482 orthologous genes between the two organisms. [12,24]

Therefore, orthologous genes corresponding to gene clusters associated with the plant invasion in *U. maydis* are partially conserved in the *M. antarcticus* genome. These findings imply that, except from the essential genes for *U. maydis*' plant pathogenicity, the genome organization and gene sets of *M. antarcticus* are nearly identical to those of *U. maydis*. Nonetheless, the efficiency of MEL production

varies across the two organisms, with *M. antarcticus* using vegetable oils to produce MEL at a faster rate than *U. maydis*, despite the fact that they share a significant of chromosomal content. Therefore, knowing that they reach distinct productivity under oily conditions, it is possible to conclude that the gene expression profiles of the two fungal species are different. [12,25]

### 2.2.2. *Moesziomyces*

*Pseudozyma* sp. is the most well-known MEL-producing species in the family of fungi *Ustilaginaceae*, which was initially classified as a genus of anamorphic and asexually reproducing yeasts, based on physiological and morphological similarities. The majority of the genetic techniques adopted for *Pseudozyma* or, as it is currently classified, *Moesziomyces* species, were developed regarding the model organism *U. maydis*, since it is a similarly related phytopathogenic.[16]

In comparison to the plant-pathogenic *U. maydis* genome, a strain of the *Moesziomyces* species, non-pathogenic *M. antarcticus*, has highly conserved synteny, and more than 80% percent of its genes are orthologous to those of *U. maydis*, hence these strains are quite similar. In addition to the relevance of understanding gene expression in *M. antarcticus* or improving genetic engineering and commercial applications, knowledge of fatty acid metabolism will lead to more effective ways of using feedstocks to produce functional bio-based materials. [12]

The physiological role of MEL in these microorganisms is still unknown, however, some research has shown that they can act as emulsifiers for uptake of lipophilic substrates, as a carbon storage molecule, and for surface colonisation via adherence to plant surfaces, cell morphology, or temperature tolerance. Furthermore, when carbon sources of hydrophobic nature are present, MEL formation is greatly induced. Moreover, MEL can be a natural form of protection against powdery mildews because these organisms are mostly segregated from plant surfaces. In addition, numerous *Moesziomyces* species have been found to have biological activity against biodegradable plastics, which are commonly employed in industrial operations, and have been demonstrated to release diacylated MEL in large concentrations, with higher substrate yields and productivities than previously described species. [16]

Different forms of MELs are produced by this type of yeast, and the pattern of production is largely dependent on the producer. MEL-A (the main component, representing more than 70% of all MELs), MEL-B, and MEL-C are mixtures of types of MELs regularly produced in *M. antarcticus*, *M. bullatus*, and *M. rugulosus*. [22]

The basidiomycetous yeast *Moesziomyces antarcticus* (formerly known as *Pseudozyma antarctica*) was isolated in an Antarctic sediment sample and belongs to the *Ustilaginales* order (*Ustilaginomycetes* class and *Ustilaginomycotina* subphylum).[26] The importance of MEL for *Moesziomyces antarcticus* low-temperature endurance was demonstrated, and MEL was also proposed as a carbon storage material similar to triglycerides in this strain. *Moesziomyces antarcticus* was the first-known MEL producer, producing mainly MEL-A, and is asexually typified but closely related to the

smut fungus *Moesziomyces bullatus* (order *Ustilaginales*, previously *Pseudozyma aphidis*). [20] [27] This strain is shown to be able to synthesize MEL only from hydrophobic sources, not water-soluble ones, and has been described to produce large amounts of MELs. [28] However, *M. antarcticus* and *M. bullatus* revealed significant differences in substrate-dependent induction of MEL synthesis compared to that of *U. maydis.* [29]

## 2.3. Metabolic pathways of MEL production

The yeasts previously mentioned, *Ustilaginomycetes*, have a very complex cellular metabolism since they are eukaryotic fungi. The numerous pathways described in that metabolism can lead to commercially valuable molecules, including MEL.

Hydrophilic and hydrophobic precursor molecules are required for the formation of the MEL glycolipid. The sugar core of MELs is produced by binding an erythritol molecule, which is synthesised through the pentose phosphate pathway (PPP), onto GDP-mannose, obtained via glycolysis, from several hydrophilic precursors (Figure 7a). The hydrophobic tail is formed of fatty acids that can be added to the culture medium, as different types of oils, and integrated into the MEL or synthesized *de novo* by the microorganism. Partial peroxisomal *β*-oxidation, also referred to as chain-shortening pathway of fatty acids, is another step of MEL metabolism of great importance, giving MEL's unique fatty acid patterns, since it provides fatty acids for MEL's acylation (Figure 7b). [30]

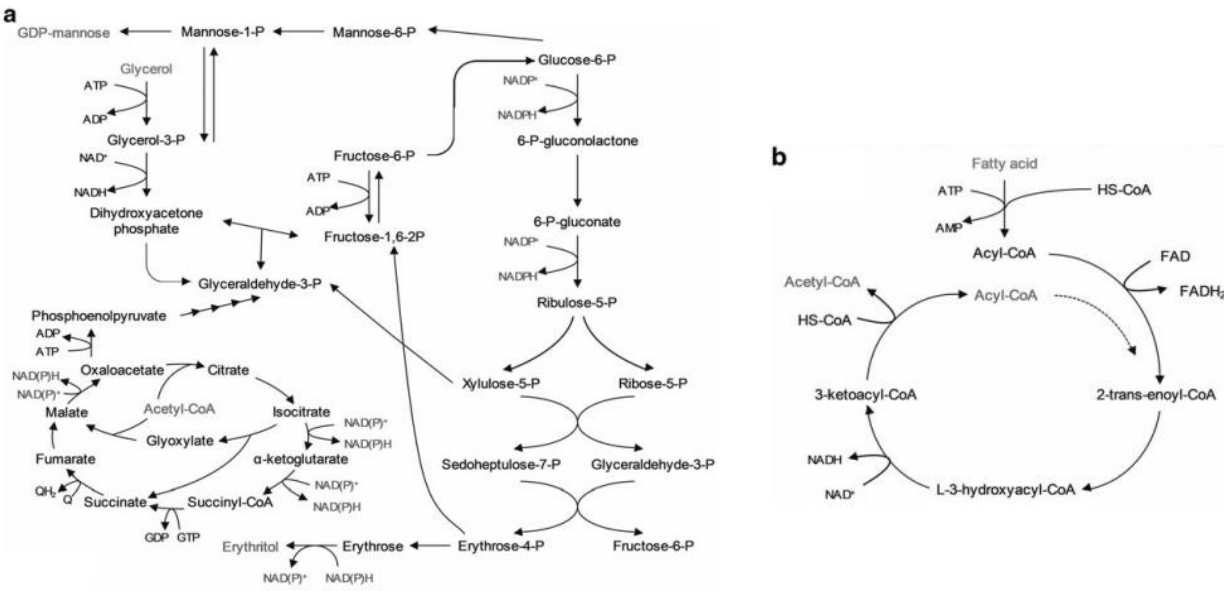



**Figure 7** – Metabolic pathways involved in the production of intermediates for MEL accumulation, (a) GDP-mannose and erythritol and (b) Acyl-CoA and acetyl-CoA, in *Ustilago maydis*. [30]

Finally, MEL synthesis is established via mitochondrial *β*-oxidation, the glyoxylate cycle, and gluconeogenesis, which converts the corresponding precursor molecules to the final product and establishes the link between fatty acid and sugar metabolism. Its regulation will be addressed in the following chapters, and a review of the most important metabolic pathways already stated can be found

in Figure 8, which was constructed using *U. maydis* as the model organism and compared to genomic data for various MEL-producing strains. [16]

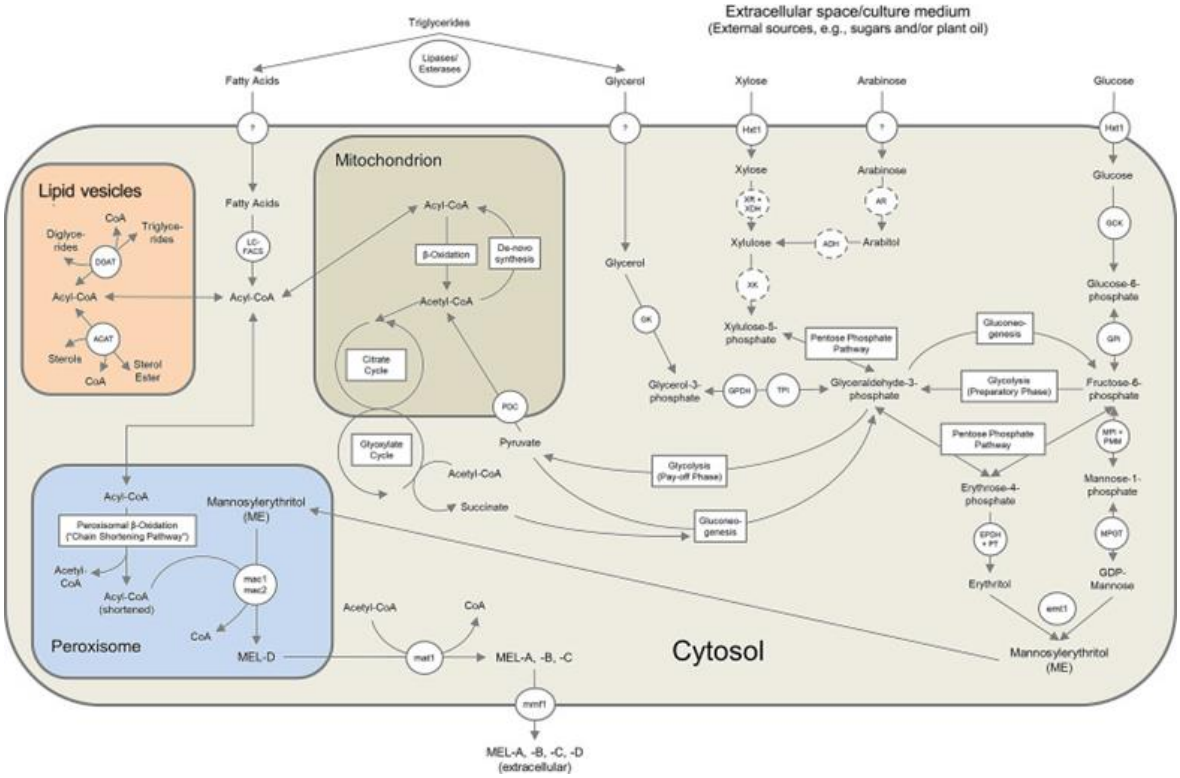



**Figure 8** – Detailed metabolic pathways that lead to the formation of MEL from various substrates in *Ustilago maydis*. [16]

## 2.3.1. Glycolysis/Gluconeogenesis

The first step in MEL metabolism is the formation of mannosylerythritol (ME) from erythritol mannosylation, which is most likely mediated by a glycosyltransferase, the Emt1 (erythritol-mannosyl-transferase), that codes for a protein with similarity to prokaryotic glycosyltransferases involved in the biosynthesis (Figure 9). [31] MEL-producing cells are reported to yield large amounts of the intermediate mannosylerythritol, indicating that this is an essential stage in MEL generation. [32]

Glycolysis pathway, as shown previously in Figure 7a, encodes a series of reactions that converts sugars like glucose or fructose, as the carbon source, into a succession of biomass component precursors, leading to the production of pyruvate as the final metabolite of this pathway. Specifically, it works by converting one 6-carbon glucose molecule into two 3-carbon pyruvate molecules, through a series of ten chemical reactions. In these reactions, substrate-level phosphorylation produces a net two molecules of adenosine triphosphate (ATP), as well as two molecules of reduced nicotinamide adenine dinucleotide (NADH).

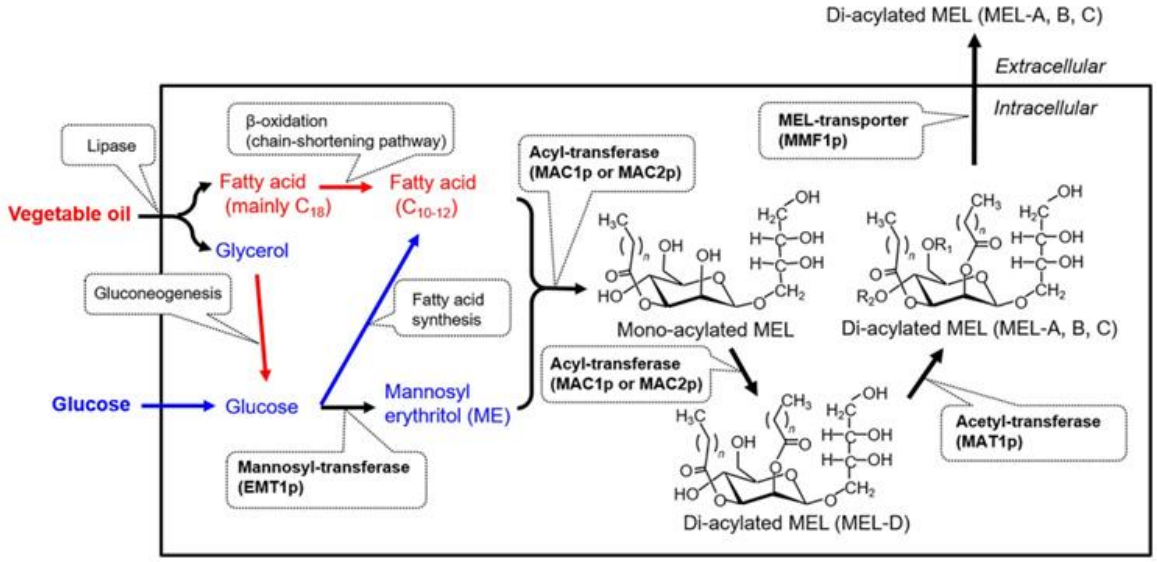



**Figure 9** – Mannosylerythritol lipids biosynthetic route. Emt1p (mannose/erythritol transferase); Mac1p and Mac2p (acyl-transferases); Mat1p (acetyl-transferase); Mmf1p (predicted MEL transporter). [25]

In particular, the synthesised MELs production starts when a monosaccharide transporter carries glucose into the cell, which is then activated by phosphorylation and transformed to fructose-6-phosphate. Fructose-6-phosphate can be glycolyzed or isomerized to mannose-1-phosphate and activated, resulting in GDP-mannose, which is the initial precursor of MEL. When applying plant oils, as the lipid source, glycerol is produced in the broth by lipase or esterase activity. However, as large amounts of glycerol, such as from biodiesel production, are available for fermentation, it can be supplied in addition. Glycerol is transformed to glyceraldehyde-3-phosphate when it is transported into the cell. Following that, it can pursue the same metabolic pathways as previously stated for other substrates, but through the reverse direction of the reactions from glycolysis, via the gluconeogenic pathway. [16]

To have a comparison between the two main producers of MEL, highlighted in the previous chapter, a recent study, comparing the gene expression intensities from *M. antarcticus* T-34 and *U. maydis* UM521, was analysed. Regarding the glycolysis step, when in oily conditions, the majority of the genes were highly expressed in *M. antarcticus*, whilst those in *U. maydis* were suppressed (Figure 10). These results indicate that *M. antarcticus* central metabolic pathways, such as glycolysis and the tricarboxylic acid cycle (TCA), could be related to MEL production, whereas the transcriptional regulatory properties are comparable to MEL biosynthesis genes. [25]

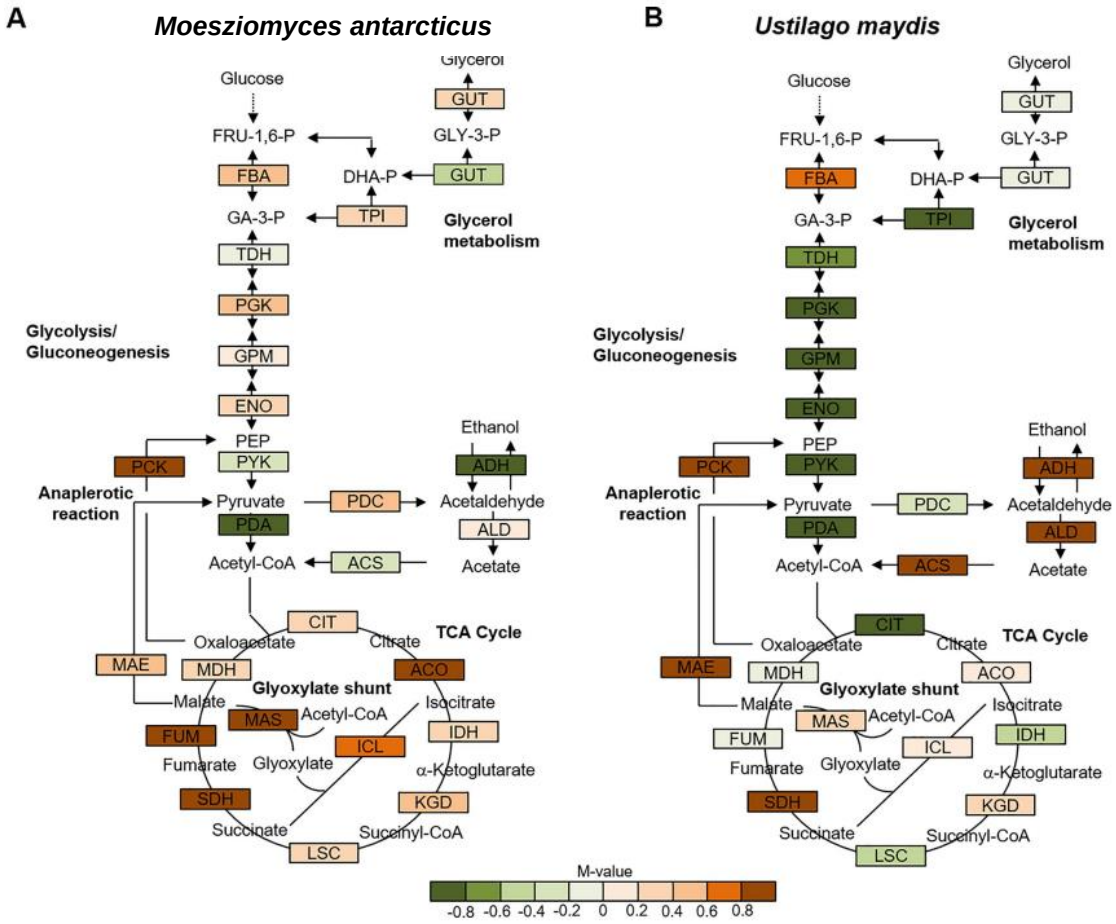



**Figure 10** – Comparison of the expression profiles of genes encoding enzymes that participate in glycolysis and the tricarboxylic acid cycle under oily conditions in *M. antarcticus* T-34 (A) and *U. maydis* UM521 (B). M-values in the central metabolic pathway of *M. antarcticus* T-34 and *U. maydis* UM521. Positive value, in reddish colours: genes induced in the presence of vegetable oil. Negative value, in greenish colours: genes repressed in the presence of vegetable oil. Value equal to zero: genes with no activity in the presence of vegetable oil. [25]

## 2.3.2. Tricarboxylic Acid Cycle

Tricarboxylic acid cycle, also named Kreb's cycle or citric acid cycle, is a cyclical pathway where acetyl-CoA (AcCoA) is oxidized to generate energy. This metabolite can be obtained by oxidative decarboxylation of the final product of glycolysis, pyruvate, considering that it is a cofactor for pyruvate carboxylase, which also is the first enzyme of the gluconeogenesis pathway. [33] Pyruvate enters the matrix of mitochondria, where TCA occurs, and is converted in AcCoa which, together with oxaloacetate (OAA), produces citrate in the TCA cycle. [34]

In this cycle a mole of AcCoA, originated through oxidation of long-chain fatty acids, enters the cycle, two moles of $CO_2$ are evolved, and a mole of OAA is regenerated in a single cycle turn, as it is shown in Figure 11. [35] Acetyl-CoA is required for a variety of metabolic processes in the cell, including energy generation in the citrate cycle, as well as *de novo* fatty acid synthesis. In fungal cells, it represents the linkage between sugar and fatty acid metabolism (Figure 12). [16]

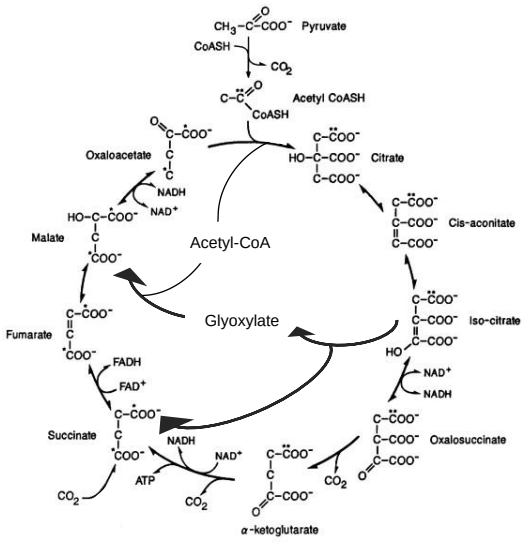



**Figure 11** – Tricarboxylic acid cycle. The asterisks indicate the carbon distribution in a single cycle turn, starting with acetyl-CoA. At the succinate stage, notice the randomization of carbon atoms (adapted from [35]).

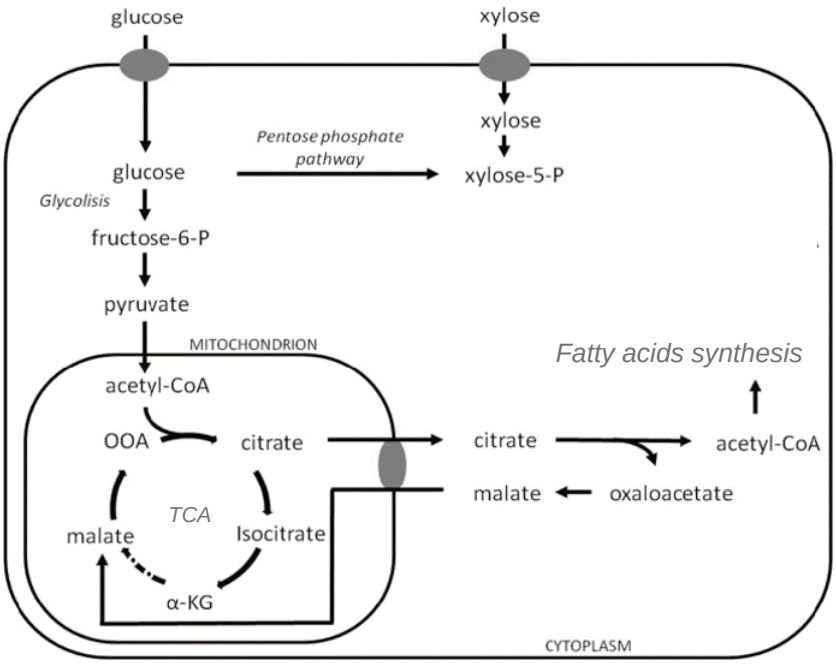



**Figure 12** – Pathways involved in AcCoA synthesis (adapted from [36]).

## 2.3.3. Pentose Phosphate Pathway

Nonetheless, the carbons provided by glucose for subsequent glycolysis and TCA are not exhausted in these pathways. The pentose phosphate pathway (PPP) provides all organisms with a supply of nicotinamide adenine dinucleotide phosphate (NADPH) for use in reductive biosynthesis, such as the production of fatty acids, and it also generates five-carbon sugars, becoming an important pathway in MEL metabolism.

In the nonoxidative branch of this pathway, an enzyme connects PPP to glycolysis and provides sugar phosphates to the main carbohydrate metabolic pathways, along with transaldolase (Taldo1). Glucose is phosphorylated to create glucose-6-phosphate (G6P), a glycolysis or pentose phosphate pathway intermediate. G6P is converted to ribulose-5-phosphate in the pentose phosphate pathway, and this metabolite is transformed to xylulose-5-phosphate or ribose-5-phosphate by isomerization and epimerization processes (Figure 13).[37]

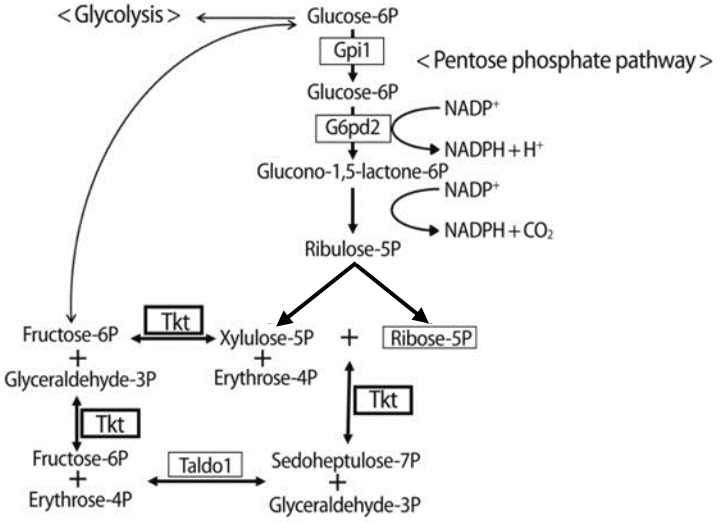



**Figure 13** – Diagram of the pentose phosphate pathway showing substrates and enzymes involved in it. Tkt: transketolase; Gpi1: glucose phosphate isomerase 1; G6pd2: glucose-6-phosphate dehydrogenase 2; Taldo1: transaldoase 1. [37]

Glyceraldehyde-3-phosphate, a glycolysis intermediate, and fructose-6-phosphate can both be supplied into the pentose phosphate pathway, resulting in erythrose-4-phosphate, a precursor of erythrose, which is then converted to erythritol.[16] Erythritol, when linked to active GDP-Mannose, results in mannosylerythritol (ME).

### 2.3.4. Fatty acids Synthesis

Mannosylerythritol needs to be acylated with fatty acid chains to become the final product, mannosylerythritol lipid. The synthesis of fatty acids initiates when malonyl-CoA, the product of the acetyl-CoA carboxylase (ACC) reaction with AcCoa, coming from the TCA cycle, is converted to malonyl-ACP. This conversion is performed by a transacylase (FabD), and the next step, which uses short acyl-CoA primers as substrates, is catalysed by the enzyme *β*-ketoacyl ACP synthase III (FabH).[38] The elongation of the developing acyl chains is achieved in four steps, which are catalysed by the enzymes shown in Figure 14, leading to the production of the metabolite acyl-ACP. This pathway is also nominated as type II fatty acid biosynthetic pathway or chain elongation. [16]

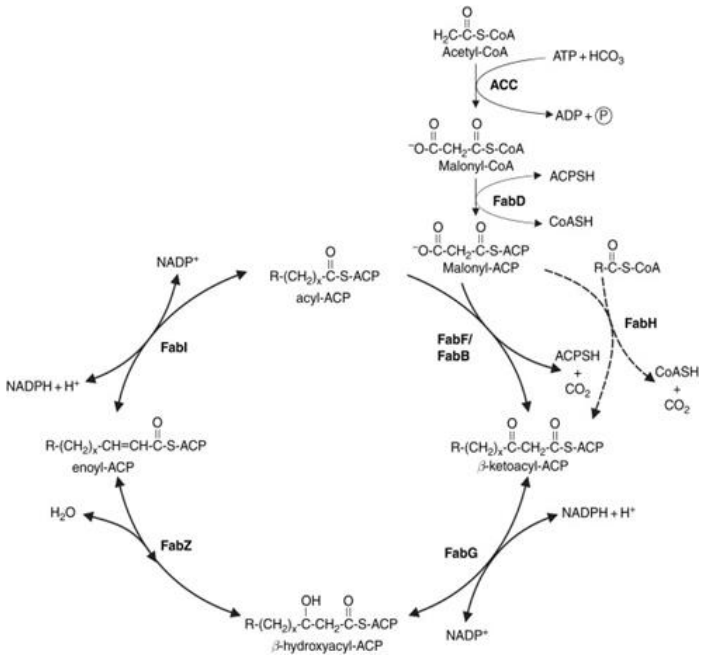



**Figure 14** - Initiation and elongation cycle of fatty acid biosynthesis. ACC: acetyl-CoA carboxylase; FabD: malonyl-CoA transacylase; FabH: *β*-ketoacyl ACP synthase III; FabB or FabF: *β* -ketoacyl ACP synthase I or II, respectively; FabG: *β*-ketoacyl ACP reductase; FabZ: *β* -hydroxyacyl-ACP dehydrase; FabI: enoyl reductase. [38]

When vegetable oil was used as the substrate instead of glucose in the experiment previously mentioned, with *M. antarcticus* and *U. maydis* genes, the induction ratios of the genes for fatty acid synthesis were decreased in both strains. Because fatty acids must be generated from glucose to meet physiological demands and for MEL production (Figure 15), that findings suggest that fatty acid synthesis is required just in the presence of glucose. [25] The opposite happens under oily conditions, since this substrate can be degraded and fatty acids are obtained via the beta-oxidation pathway, which will be explained in the next chapter, and not from fatty acid (FA) synthesis.

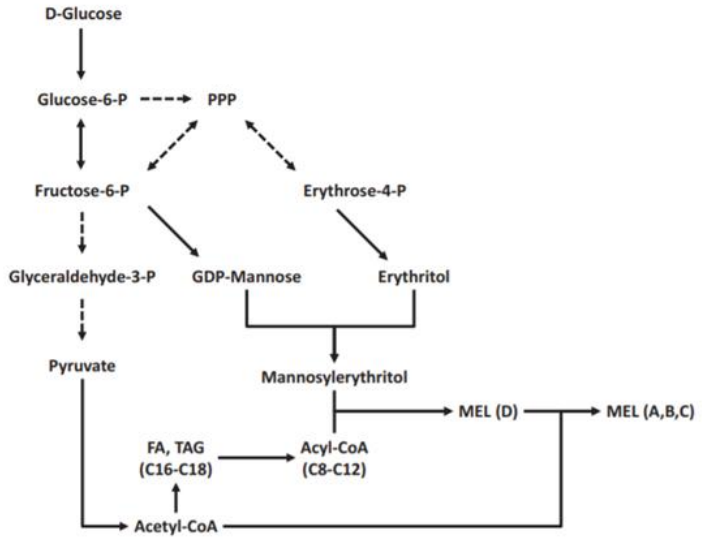



**Figure 15 -** Metabolic pathways for the biosynthesis of mannosylerythritol lipids (MEL) from glucose. PPP: pentose phosphate pathway; FA: fatty acids; TAG: triacylglyceride. [39]

### 2.3.5. *β*-oxidation

Different fatty acid pathways exist in microorganisms, each leading to the synthesis of MEL or the storage of intracellular triacylglycerides. Supplementing plant oils into the culture medium is the most typical technique to supply FAs for MEL synthesis, with the cleavage of triacylglycerides (TAGs) into FAs and glycerol as the initial step, by extracellular lipases or esterases. Fatty acids are directed to lipid metabolism while glycerol is integrated into the above-mentioned pathways, particularly at gluconeogenesis. [16]

Thus, either complete *de novo* synthesis from acetyl-CoA (type I fatty acid biosynthetic pathway), chain-elongation from medium-chain acids (described in the previous chapter), or direct integration (type III fatty acid biosynthetic pathway), as seen in other oleaginous yeasts, are likely to play a role in intracellular triacylglyceride production. In type I fatty acid biosynthetic pathway, the FA parts of the MELs can be formed via beta-oxidation (*β*-oxidation), with long-chain acids being broken down and FAs composition of MEL dependent on the oil supplied as carbon source. [12]

Findings on MEL production by *M. antarcticus* demonstrated that the FAs in MEL are not produced by *de novo* synthesis, but rather are intermediates in a *β*-oxidation pathway that can take place in two distinct organelles. In *M. antarcticus*, as well as in *U. maydis*, genes involved in peroxisomal and mitochondrial *β*-oxidation were increased, but the induction rates of genes involved in fatty acid synthesis were inhibited in both strains, under oily circumstances. This research suggests that both strains degraded vegetable oil via the *β*-oxidation pathways and that all the genes involved in the metabolism, with the exception of lipid metabolism genes, were expressed alongside MEL biosynthesis genes, implying that *M. antarcticus* efficiently derived energy from vegetable oil via respiratory metabolism. [25]

In an experiment to better understand MEL production, cerulenin (a potent inhibitor of *de novo* fatty acid synthesis) was used and resulted in the production of MELs remaining unaffected. In contrast, the use of 2-bromooctanoic acid (a potent inhibitor of the fatty acids *β*-oxidation pathway), blocked the synthesis of MELs, with the degree of inhibition directly correlated with the chain length of the substrate used. These findings revealed that the production of TAGs involved more than one of the three fatty acids biosynthetic pathways, and that the choice of pathway is influenced by the chain length of the substrate.[23]

#### 2.3.5.1. Complete *β*-oxidation

The *β*-oxidation pathway that occurs in mitochondria is also named as complete *de novo* synthesis, or type I fatty acid biosynthetic pathway. In this pathway fatty acids have to be generated from acetyl-CoA, leading to MEL production from non-lipidic substrates.

However, it is reported in the literature that medium-chain fatty acids are derived through the chain-shortening pathway in *M. antarcticus* MEL production, and that *de novo* synthesis is not required.

When cells were cultured on long-chain fatty acids, suppression of *de novo* fatty acid synthesis had only a minor influence on MEL synthesis, however growth was partially reduced. In fact, blocking *de novo* synthesis pathway increased MEL formation slightly, likely by channelling carbon flux to MEL synthesis. With or without this inhibitor, the fatty acid distribution in MEL was identical. [16]

### 2.3.5.2. Partial *β*-oxidation

While limiting *de novo* fatty acid synthesis had little effect on MEL production, adding a partial *β*-oxidation inhibitor to *M. antarcticus* resting cells reduced MEL generation. As a result, the long-chain substrate (C18) had the highest inhibition rate. Three reported chemicals that limit *β*-oxidation of fatty acids are 2-bromooctanoic acid (BOA), thioridazine, and 4-pentenoicacid. [40] In Figure 16, the already mentioned presupposed pathways, responsible for the shortening of fatty acid chains in MEL production, are represented.

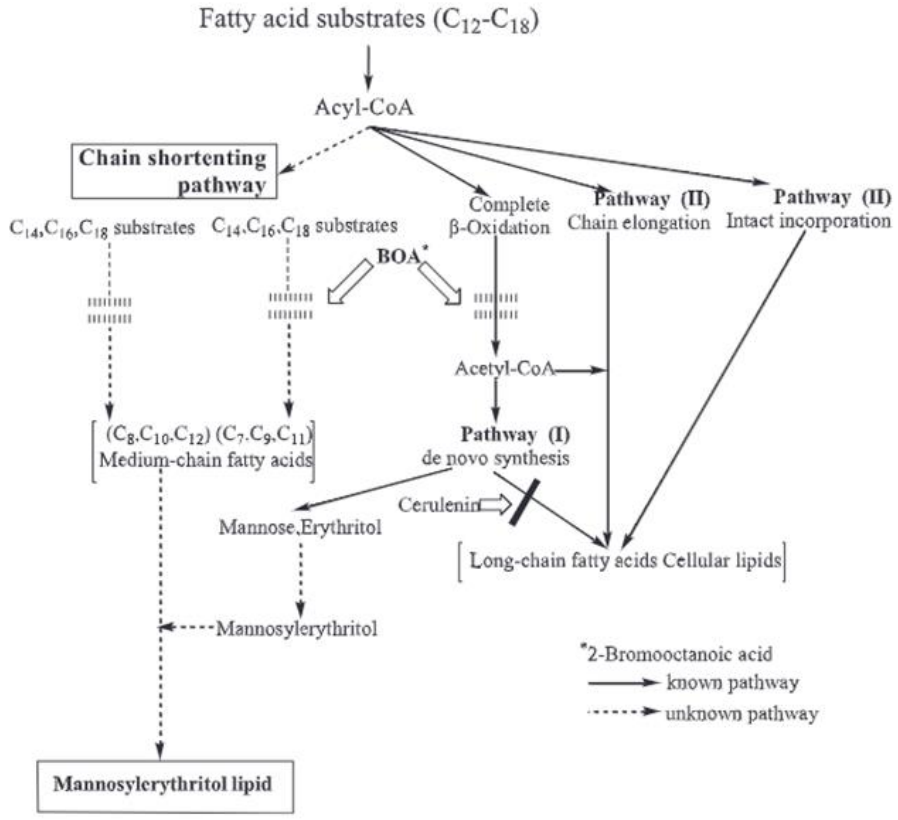



**Figure 16** – MELs presumptive chain-shortening pathway in *Moesziomyces antarcticus.* [41] [40]

Peroxisomal activation of acyl-CoAs, also known as partial *β*-oxidation or chain-shortening pathway, is required for the activation of long-chain fatty acids and acetate as a growth-dependent carbon source.[34] This partial *β*-oxidation, which takes place in cellular peroxisomes, is important for MEL specific fatty acid patterns. In contrast to total *β*-oxidation, which occurs in the mitochondrion, the peroxisomes are key organelles for secondary metabolite synthesis in fungi in general.

Two acyl-CoA oxidases, two thiolases, and two bifunctional proteins are involved in the peroxisomal beta-oxidation of the various acyl-CoA esters. According to current knowledge, the key enzyme responsible for oxidating very long-chain fatty acids (VLCFAs) and dicarboxylic acids (DCAs) is acyl-CoA oxidase 1 (ACOX1), as shown in Figure 17. [42]

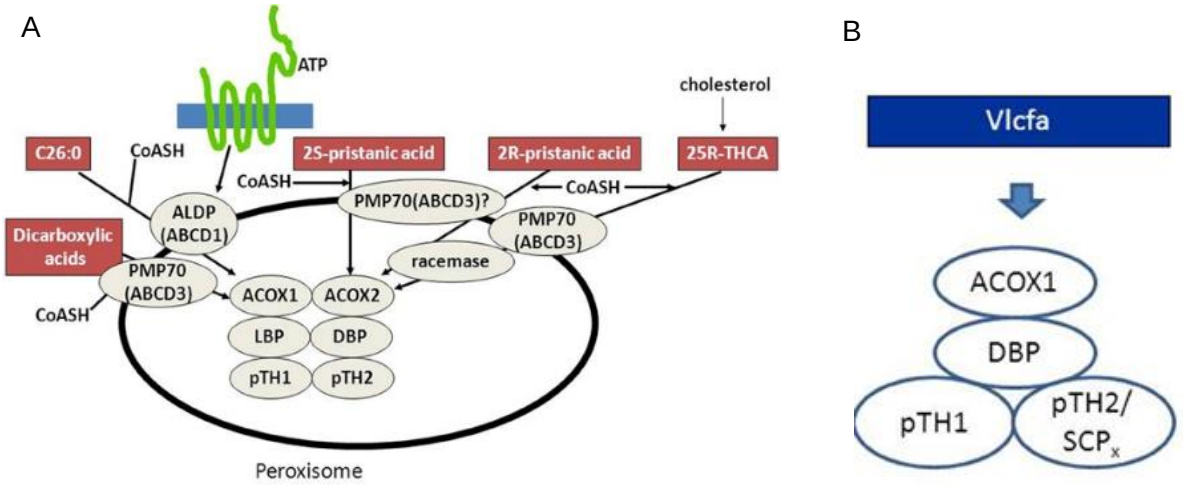



**Figure 17 –** (A) Substrates that are exclusively known to be oxidized in peroxisomes, together with the transporters and enzymes involved in their degradation. (B) Diagram showing the peroxisomal enzymes responsible for the degradation of very long-chain fatty acids (VLCFAs). (adapted from [42])

Recently, it was discovered that, in *U. maydis*, deletion of two genes involved in peroxisomal *β*-oxidation resulted in a significant reduction in MEL production, showing that the chain-shortening pathway and its localization in peroxisomes are critical for MEL biosynthesis. [16] The compartmentalization in peroxisomes of MEL biosynthesis is evolutionary conserved and necessary for the assembly of MELs with the typical acylation pattern. This location of MEL biosynthesis is required not only for the creation of the natural spectrum of MELs, as well as for the assembly of several glycolipids in a single cell. [8]

### 2.3.6. MEL Synthesis

The MEL synthesis is catalysed by different enzymes and the encoding genes for these enzymes are clustered together. The gene cluster is comprised of five genes that code for a glycosyltransferase (Emt1), two acyltransferases (Mac1 and Mac2), one acetyltransferase (Mat1), and one cellular exporter (Mmf1), which have first been discovered in *Ustilago maydis*.[14] As a result, the complete assembly of the MEL molecule consists of four major steps, described in Figure 18.

The glycosyltransferase Emt1 (erythritol-mannosyl-transferase) connects activated GDP-mannose and erythritol for the first time and this glycosylation reaction takes place in the cytosol, which has an abundance of sugar precursors. The resultant mannosylerythritol (ME) is subsequently acylated with two fatty acid chains at the mannose moiety's C2′ and C3′ positions, generating diacylated MEL-D.

The two acyltransferases Mac1 and Mac2 (mannosylerythritol-acyl-transferases), catalyse these two acylation processes, which require activated fatty acids in the form of acyl-CoA. Both enzymes are considered to be highly regioselective, preferring distinct acyl-CoA chain lengths. [32]

Because deletion of either Mac1 or Mac2 lead completely disabled MEL synthesis, this acylation process appears to be required for secretion. Moreover, Mac1 and Mac2 were discovered to have particular targeting sequences that coordinate their localization in peroxisomes. Chain-shortening of fatty acids, the peroxisomal *β*-oxidation, also occurs in this organelle, so the colocalization of fatty acid metabolism and acyl-transfer into MEL makes this compartmentalization understandable. [8]

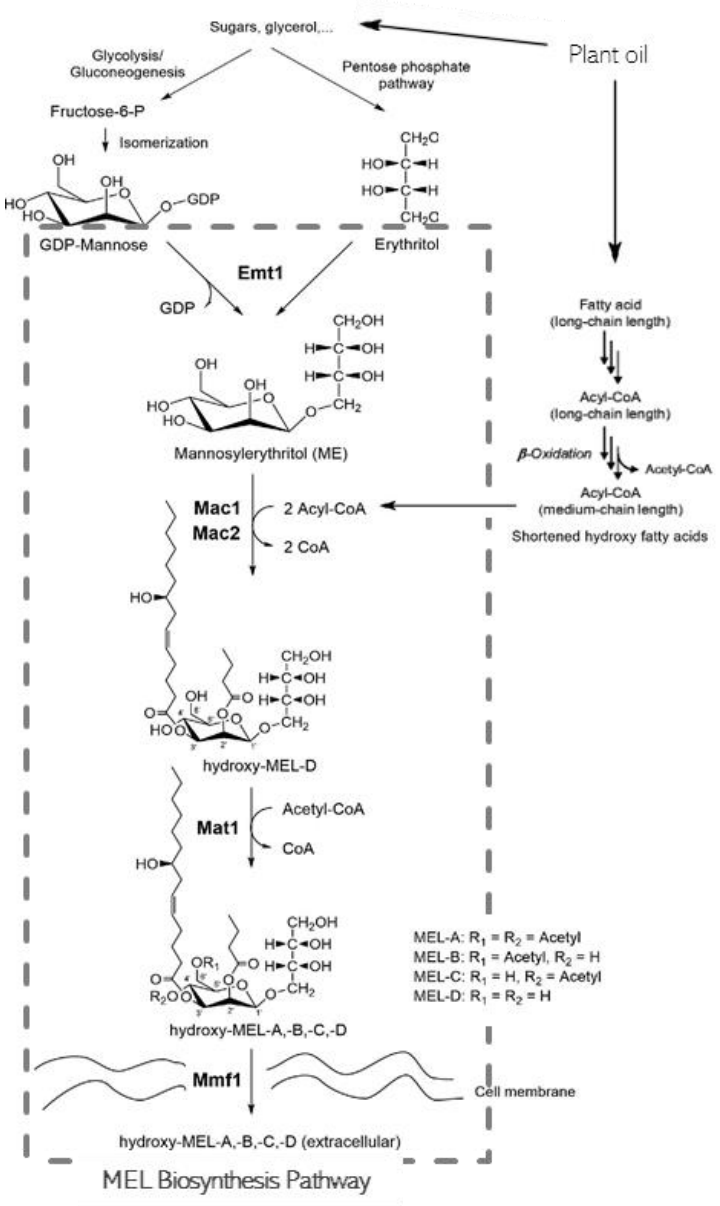



**Figure 18** – The biosynthetic proposed metabolic pathway for MEL production from plant oil. Chain-shortened hydroxy fatty acids from oil get esterified at C3′ of the mannose moiety by action of an acyltransferase Emt1: erythritol/mannose transferase; Mac1 and Mac2: acyltransferases; Mat1: acetyltransferase; Mmf1: membrane-bound transporter. (adapted from [7,14] [43])

Subsequently, the single enzyme Mat1 (mannosylerythritol-acetyl-transferase), which requires acetyl-CoA as a substrate, acetylates the MEL-D molecule to variable degrees at C4′ and C6′. The three acetylated versions generated, MEL-A, MEL-B, and MEL-C, are produced in different quantities depending on the activity or regioselectivity of Mat1 in distinct strains. This enzyme is present at the plasma membrane, where it catalyses the acetylation of MEL-D before the membrane-bound transporter Mmf1 exports MEL produced. MEL can be exported from the cell by this membrane transporter Mmf1 (mannosylerythritol-major-facilitator protein), which belongs to the major facilitator protein superfamily. Mmf1 is thought to have poor selectivity because all four MEL variants can be identified outside the cell. [16]

In *M. antarcticus*, high expression of the MEL biosynthetic gene cluster was seen in oily conditions, leading to the conclusion that it may be able to synthesize MELs, whereas *U. maydis* gene cluster expression is susceptible to a variety of nutrients, including carbon and nitrogen sources. [12] The gene cluster for MEL biosynthesis was strongly expressed regardless of whether the carbon source in *M. antarcticus* was glucose or oil, leading to the conclusion that the metabolism of MEL should be studied in this organism, rather than in *U. maydis*.

With this, it is possible to understand that the pathways of MEL synthesis should be studied in detail, and a model must be designed for predictions of the metabolism routes followed, according to the substrate given to the organism. To forecast genetic modifications that rearrange the metabolism toward the generation of the compound of interest, metabolic modelling is required, which is a *in silico* approach. Even if a *Ustilago maydis* general metabolic model already exists, at the moment there is not a described MEL metabolism model for *M. antarcticus*.

## 2.4. Metabolic Modelling

Biological research, in general, can be aided by metabolic and regulatory models. Under a variety of conditions, computational assumptions can be compared to experimental observations, and if there are discrepancies, it is because the model is incomplete or erroneous in some way. These discrepancies can be investigated, assisting in the discovery of novel biological characteristics, where the most up-to-date genome-scale models are usually used. [2]

Thus, the implementation of experimentally determined constraints can be used to convert a metabolic model into a condition-specific model. Constraints can be established by defining flux bounds for each reaction, for example. In a metabolic model, limitations can be applied to simulate specific cellular conditions, such as biomass maintenance requirements, environmental limits, or maximum enzyme capabilities, that can predict the outcome at *in vitro* experiments. [44]

Mathematical modelling is therefore used to study the dynamic interactions between various components of a biological system to better understand the system's overall behaviour. Hypothesis formulation and the application of prediction models are becoming increasingly important in understanding the mechanisms underlying complex biological systems, disorders, and drug actions, thanks to high-throughput omics data and network analysis. [45]

Reconstructions of cellular metabolism for a wide range of microorganisms and certain mammalian genomes are publicly accessible. These reconstructions, which are genome-scale, aim to include both reactions supported by direct experimental evidence and those implied by the genome annotation. Typically, they are designated genome-scale metabolic (GEM) networks, frequently reconstructed to include all identified metabolic genes and pathways in a given organism. Thereby, the metabolic reactions that are active in the organism at any particular time are a superset of these reconstructions. [46]

### 2.4.1. Constraint-based modelling and Flux Balance Analysis

Models are physical systems converted in a mathematical form and can be characterised by assumptions, committing certain understanding. Mostly, models are approximations of reality, or representations of data chosen to focus on essential features of the cell or organism in study, that help to understand large and complex systems. Models are built to make predictions, understand the system under research, and can explain how the system behaves if some conditions change, which enables the prioritisation in the laboratory experiments that should be done. The objective is to make well-founded and testable predictions of biological systems. An accurate metabolic reconstruction subsequently can be adjusted as a computer model and used for a growing variety of applications. To assess a metabolic network's capabilities, flux balance analysis (FBA) and other constraint-based methodologies might be applied. The biomass reaction can be optimised to the maximum value in FBA by using linear programming software to simulate growth. The steady-state mass conservation of metabolites assumption, which dictates that metabolite concentrations derivative is zero, implying that the system does not have an accumulation of compounds, imposes the main constraints in FBA and the mathematical issue is transformed into a linear system. [2] [47]

Thus, in the FBA approach, the metabolic network is defined as a linear programming optimization problem. The calculation of metabolite fluxes through a metabolic network is achieved by constraint maximisation of an objective function. A stoichiometric matrix (S) represents the metabolic reactions mathematically, having n rows, where n is the number of metabolites in the model, and m columns, which are the number of reactions. The stoichiometric coefficient of each metabolite, for each reaction, is filled into each position of the matrix (Figure 19), with this matrix and the constraints defining a linear system of equations, which is solved by linear programming. [48]

Stoichiometric modelling is a type of mathematical modelling that is applied to all approaches that use a metabolic network based on reaction stoichiometry to describe cell metabolism. However, other inputs, such as limitations, are also required, with the underlying biochemical network's stoichiometry constraining the solution in FBA. The fundamental disadvantages of these models are the lack of regulatory and kinetic input, which restricts the precision of the predictions, and that it only predicts fluxes, rather than metabolite concentrations. [49] [50]

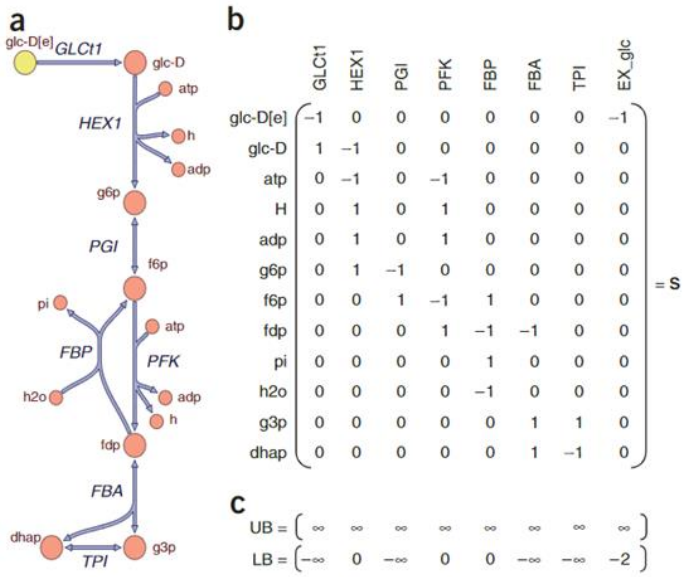



**Figure 19** – Metabolic networks represented stoichiometrically. (a) A graphical representation of glycolysis reactions in *E. coli*. (b) The matching stoichiometric matrix (S). Each column corresponds to a specific reaction and each row to a specific metabolite. An exchange reaction is shown in the last column. (c) Inequality constraints, reaction's upper (UB) and lower (LB) boundaries. [51]

Constraint-based metabolic models can be employed in synthetic biology and metabolic engineering fields due to their predictive capabilities. Constraint-based models are particularly useful since they can forecast when a specific metabolite will be overproduced and released, excluding impossible behaviours or impossible flux solutions. By adding or deleting reactions to the network, gene knockouts and knock-ins may be simulated, and the behaviour of the modified network can be predicted through FBA. Because FBA considers the complex interacting effects that a knockout has on all pathways at the same time, these model-based predictions are very accurate. Likewise, the range of possible states a biochemical reaction system might achieve can be defined by a variety of physical, chemical, and biological constraints, with these states correlating to distinct flux distributions through the metabolic network. [51]

## 2.4.2. Program and tools used in metabolic modelling based on FBA

Systems biology modelling efforts result in models with an ever-increasing number of components and interactions. As these models get more complex, software that provides sophisticated numerical analysis tools to study their behaviour is required. MATLAB is a programming and numeric computing platform with a large number of auxiliary toolboxes, a standard in many scientific fields, commonly used in systems biology. It includes a user-programmable platform with several state-of-the-art mathematical and numerical methods, as well as a simple, but powerful, high-level scripting language, that minimises the programming overhead typically associated with independent software applications.[52]

The toolboxes available for this program provides a flexible and open environment for systems biologists to cope with different ideas, develop and share new algorithms, and create applications for the study and modelling of biological and biochemical systems.[52]

A wide range of methodologies for assessing genome-scale models have been developed, and the methods have been used to investigate a growing variety of biological problems. FBA can be used to analyse constraint-based models by the constraint-based reconstruction and analysis (COBRA) strategy. COBRA is a modelling technology that creates manually curated stoichiometric network reconstructions. Models can then be constructed and examined using equality and inequality constraints, as well as computing functional states. Mass conservation and thermodynamics (for directionality) are among the constraints, but so are the constraints based on experimental procedures and regulatory ones. [40] [43]

COBRA is a framework for mechanistic integrative analysis that can be applied to any biological system having prior mechanistic information, including systems with inadequate information, as mechanistic models describe cellular behaviour based on simplifications.[53] The requirement for easy reproducibility and request for usage of COBRA methods was recognized early in the creation of the COBRA framework. This requirement prompted the creation of COBRA Toolbox, an open-source software tool that operates in the MATLAB environment and allows for quantitative metabolic phenotyping utilising a variety of COBRA methodologies. [1]

A tool that allows visualization of the fluxes through the different metabolic pathways present in the model is Escher, a web application for visualizing data on biological pathways. This application enables the users to quickly construct new pathway maps, creating paths semi-automatically by using Escher's pathway suggestions, which are based on user data and genome-scale models. By employing rules that specify which enzymes catalyse each process, users can see information related to genes or proteins on the associated reactions and pathways, using the associated reactions and pathway data.

A source with the names, stoichiometries, and associated genes for each metabolic reaction in an organism is required in order to construct a pathway map. A COBRA model, which is a compilation of all the processes, metabolites, and genes that are known to exist in an organism, provides this information. Escher might be used to display pathways like gene expression and membrane translocation, which is added into COBRA models (Figure 20). Although COBRA models have typically focused on metabolism, the COBRA modelling technique can be used to any biological reaction network. [54]

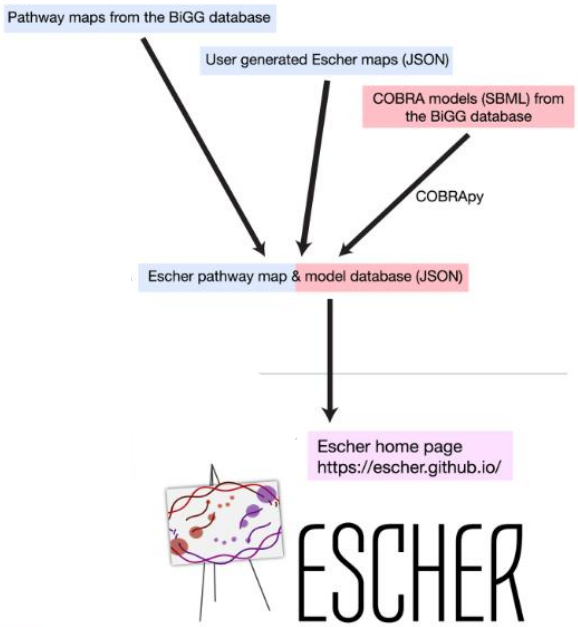



**Figure 20** – The structure of the Escher project. Escher maps can be created by users or retrieved from the Biochemical, Genetic and Genomic (BiGG) database. COBRApy is used to produce COBRA models. On the Escher website or locally via one of the several Python package methods, the Escher web application can be viewed. (adapted from [54])

While FBA provides predictions for metabolic networks with thousands of components, making important changes in FBA solutions challenging to detect, the easiest method to understand FBA is to explore simulations interactively by changing parameters and receiving real-time feedback. For beginners, understanding how FBA functions is challenging due to these difficulties. By adding on-going FBA calculations to the Escher program for pathway visualization, Escher-FBA, an extension of the Escher application, satisfies these requirements. In order to provide interactive FBA simulations within a pathway visualization, Escher-FBA creates the network and reaction data using the same input files as Escher. Without downloading any software or writing any code, users can also create high-quality figures, upload metabolic models, knock out reactions, modify objective functions, and set flux bounds, constraining specific reactions. Hence, it is feasible to reproduce several FBA simulations that lead to valid scientific theories.[55]

### 2.4.3. *In silico* models

To stay consistent with genome sequencing and eliminate a major bottleneck in metabolic analysis, computational approaches for metabolic reconstruction must be able to generate models that require minimum curation, while yet accurately predicting metabolic phenotypes. The conversion of *in silico* models into mathematical models, which are utilised to predict physiological characteristics and states, is frequently an iterative process. The model is made publicly available and then modified with continuous input from the scientific community. [47] [56]

Biochemical network reconstructions are databases with biochemical, genetic, and genomic structures. The information obtained from biochemical characterization of processes and their

substrates is frequently used in these reconstructions. To discover the encoding enzymes responsible for each catalytic reaction, genetic research is also required. The fact that biochemical network reconstructions can be converted into a computational model makes them useful. As a result, the computer model may be used to answer a growing number of biological issues. In some cases, model predictions can be used to identify profitable pathways for experimental confirmation, as well as give qualitative and statistically correct projections of experiments. [2]

The outcome is an *in silico* forecast of steady-state flow through each reaction in the model, such as a prediction of the maximum optimal growth rate of the cell. Growth can be simulated under a variety of conditions, including aerobic and anaerobic circumstances, as well as growth on glucose or other substrates. By varying the limits on the exchange reactions that operate as sources of substrate and waste metabolites, different circumstances can be simulated as well.

Accessible, interoperable, and reusable *in silico* models are required for the study of complex system biology. BioModels, a repository for mathematical models, was created to provide a platform that supports universal sharing, easy access, and model reproducibility. The models submitted to BioModels are curated to ensure that the computational representation of the biological process is appropriate and that the simulated results are repeatable, as shown in Figure 21. Modelers benefit from BioModels as they have access to trustworthy and semantically enriched curated models in standard formats that are simple to share, copy, and reuse. [57]

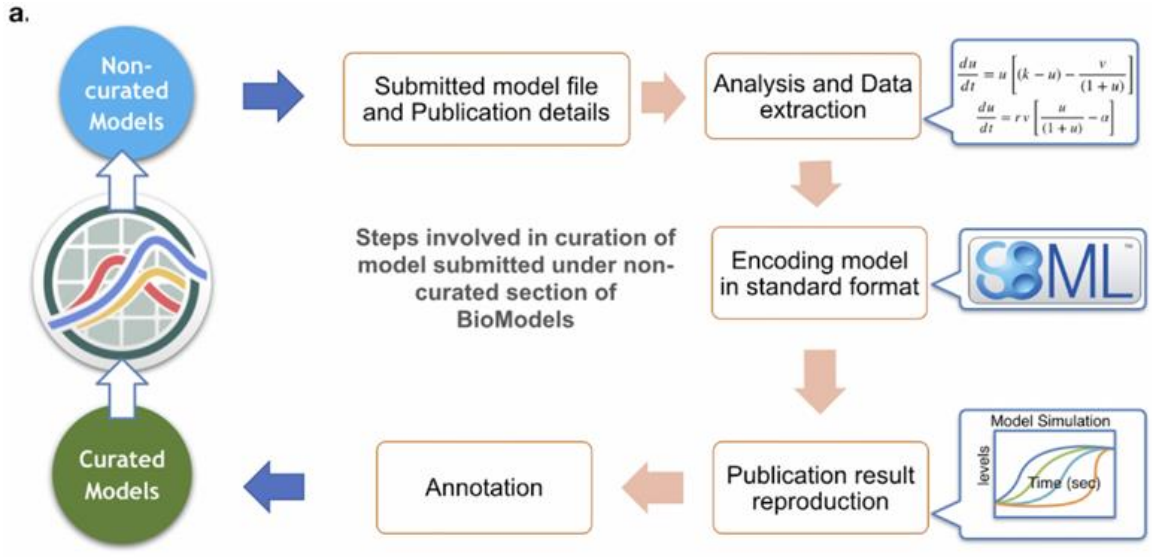



**Figure 21** - Model curation workflow in BioModels database. [57]

After models are submitted to BioModels, they undergo a thorough curation process that includes reproducing literature results and annotating them with recognized ontologies or databases. SBML (Systems Biology Markup Language) models can be stored in the BioModels database and MATLAB can provide an interface for importing these models, with toolboxes that support this language, allowing models to be utilised.

### 2.4.4. MEL biosynthesis modelling and validation

Until the date, there is not any available *in silico* model for *Moesziomyces antarcticus*. For this reason, modelling of a current model, available for *Ustilago maydis*, has to undergo mathematical approaches that upgrade it. The goal is to have a MEL metabolic model for *M. antarcticus*, which revealed being able to use a variety of raw materials, such as crude vegetable oils, with mass production of MELs. This ability is important because allows lowering production costs, by utilising raw resources, which is crucial to extending industrial applicability. [12]

As so, the differences between the MEL metabolism in these two strains must be summarised. This is possible due to a research that compared the gene expression of both strains, which was covered in detail through the previous chapters related to the metabolic pathways.

Regarding the expression of genes giving oil as substrate, the genes involved in peroxisomal and mitochondrial *β*-oxidation were more induced in *M. antarcticus* T-34 than in *U. maydis*, at the same time that the induction ratios of the genes for fatty acid synthesis were suppressed in both strains. These results suggest that the vegetable oil is degraded via the *β*-oxidation pathway and that fatty acid synthesis was required in the presence of glucose, because fatty acid must be synthesized from glucose to supply cellular demands and for MEL biosynthesis, but not with oil as substrate. The principal difference between the two strains, already shown in Figure 10 - Section 1.3.1., is the induction of glycolysis and gluconeogenesis pathways in *M. antarcticus*, and the suppression of these pathways in *U. maydis*, as the consequent pathway, the TCA, as a similar induction ratio pattern (Figure 22).

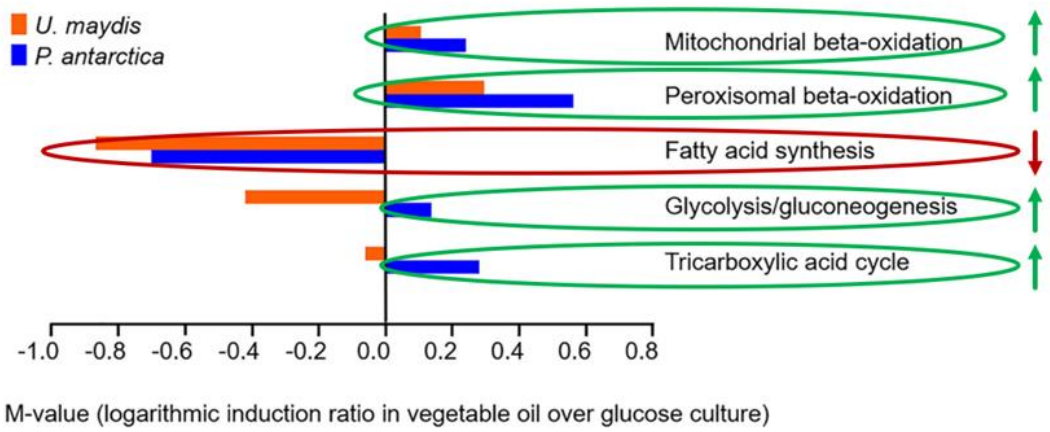



**Figure 22** – Induction of genes associated with oil degradation and conversion, as well as primary metabolism. The M-value averages of genes responsible for MEL biosynthesis pathways: mitochondrial beta-oxidation, peroxisomal beta-oxidation, fatty acid synthesis, glycolysis, tricarboxylic acid cycle, pentose phosphate pathway, and malate/pyruvate cycle are displayed. In orange is *U. maydis* UM521 and blue is *P. antarctica* T-34 (currently referred as *M. antarcticus*) values. Values below 0 represent less induction ratio, while values above 0 represent higher induction ratio, with vegetable oil as substrate, when compared with glucose as substrate (adapted from [25]).

# 3. Materials

## 3.1. Laboratory Material

Reagents: Agar (José M Vaz Pereira, S.A.),CHCl (Sigma-Aldrich®), D-glucose (Fischer), Heptanoic acid (Sigma-Aldrich®), $KH_2PO_4$ (Chem-lab), $MgSO_4$ (Sigma-Aldrich®), $NaNO_3$ (LabKem), Sulphuric acid (Sigma-Aldrich®), Yeast extract (Oxoid).

Organic Solvents: Absolute ethanol, Acetone, Ethyl Acetate, Hexane, Isopropanol, Methanol, all from Fischer®; Acetyl Chloride from Fluka.

Equipment: Autoclave (AJC, Uniclave 88); Centrifuge 1 (Sartorius 1-15P, Sigma, rotor 12000 rpm); Centrifuge 2 (Eppendorf®, 5810 R); GC (Hewlett Packard, HP5890); HPLC (Hitachi Elite LaChrom); Lyophilizer (Alpha 1-2 LD plus CHRIST®); Microscope (Leica DMLB); Oven 60ºC (Memmert); Oven 80ºC (Memmert); Pipettes (Biopette PLUS / Eppendorf®); Shaking incubator 2000 (Optic ivymen system); Spectrophotometer (UH5300 Hitachi).

## 3.2. Softwares

- COBRA Toolbox v3.0 [1] available from https://opencobra.github.io/cobratoolbox/stable/ ;
- Escher version 1.7.1 [54] available from https://escher.github.io/#/ ;
- Microsoft® Excel® for Microsoft 365 MSO commercially available from http://microsoftstore.com;
- MATLAB R2021a Update 5 (Mathworks) commercially available from http://mathworks.com.

## 3.3. Databases

- Biomodels [57] available from https://www.ebi.ac.uk/biomodels/ ;
- Github available from https://github.com/ ;
- KEGG [58,59] at http://kegg.jp/ (last date of access: 28.10.2022).

## 3.4. Genome-scale metabolic model

*Ustilago maydis* genome scale metabolic model (GSMM), iUma22, for simulation of metabolic activities. Available from https://github.com/iAMB-RWTH-Aachen/Ustilago_maydis-GEM/blob/master/code/iUma22_EscherMapping.ipynb. [60]

# 4. Methods

## 4.1. MEL production

### 4.1.1. Microorganisms and maintenance

Mannosylerythritol lipids were produced by *Moesziomyces* yeast strains, *Moesziomyces antarcticus* PYCC 8538$^{T}$ (CBS 6678) and *Moesziomyces bullatus* (previously referred as *M. aphidis*) PYCC 5535T (CBS 6821) provided by the Portuguese Yeast Culture Collection (PYCC), CREM, FCT/UNL, Caparica, Portugal. The strains were plated in yeast malt (YM) agar (yeast extract 3 g/L, malt extract 3 g/L, peptone 5 g/L, D-glucose 10 g/L, and agar 20 g/L) and incubated for 3 days at 30°C. Stock cultures of each species were prepared from the plates, by growing each strain in liquid medium and stored in 20% (v/v) glycerol aliquots at - 80ºC.

### 4.1.2. Media and cultivation conditions

An inoculum was prepared according to the procedure referred to in the literature. [61] The glycerol stocks of *M. antarcticus* and *M. bullatus* were transferred to an erlenmeyer flask, with 1/5 of the working volume (50 mL) of a mineral medium consisting of 3 g/L NaNO3, 0.3 g/L $MgSO_4.7H_2O$, 0.3 g/L $KH_2PO_4$, 1 g/L yeast extract (OXOID) and 40 g/L D-glucose, at initial pH 6.0. All were previously sterilized in an autoclave at 121 °C and 1 bar, for 20 minutes.

The inoculum was then incubated in an orbital (Shaking incubator 2000) for 48 hours at 200 rpm and 27°C. After, 10% (v/v) of the inoculum was used to start the fermentation in fresh media. Both the inoculum and the fermentations were conducted in 250 mL erlenmeyer flasks, with a working volume of 50 mL.

Initially, three conditions were tested, for each strain, with 40 g/L of D-glucose, 20 g/L of galactose, and 40 g/L of lactose, all added at day 0. Duplicates were made for each experiment. After, 10% (v/v) of the inoculum was used to start the fermentation in fresh media, with five different conditions of glucose concentration, 5 g/L, 20 g/L, 40 g/L, 80 g/L, and 120 g/L, with an initial OD below 0.1 in all the 250 mL erlenmeyer flasks. The inoculated erlenmeyer's were shaken at 200 rpm at 27°C up to 168 hours.

For the second set of experiments, an initial 40 g/L of D-glucose was used as carbon source, with feeding at day 4 of 20 g/L, 40 g/L, 80 g/L and 160 g/L of D-glucose, using the *M. antarcticus*. The inoculated erlenmeyer's were shaken at 200 rpm at 27°C up to 240 hours (Shaking incubator 2000, Optic ivymen system). Growth rates and glucose uptake rates were estimated using experiments carried out for this study and data from the literature.

## 4.1.3. Yeast cultivation parameters

### 4.1.3.1. Cell dry weight

To analyse biomass growth, the dry cell weight (DCW) was measured in the samples taken during the cultivation time. The culture broth was recovered, and 1 mL samples were centrifuged (Sartorius 1-15P, Sigma) at 10000 rpm for 5 min, resulting in the supernatant, which was removed and stored, and the pellet, which was washed twice with Milli-Q® water and left to dry at 60°C in the oven (Memmert) for, at least, 48 hours. The dry biomass was then weighed, and the DCW value was calculated.

### 4.1.3.2. Optical density (OD)

Optical density (OD) determination at 600 nm on a spectrophotometer (UH5300 Hitachi)allowed monitoring cell development in real-time. The OD measurement values of growth from the literature were converted into units of gDCW/L using the empirical relation from the yeast of 0.62 gDCW/L/OD (BNID 111182). [62] The DCW conversion coefficient value used for growth in *M. antarcticus* was 0.30 g DCW/L/$OD_{600}$. [63]

A linear relationship of the exponential growth phase was reached with the ln(OD) over time. The growth rate of each condition was obtained with the slope value of each condition, as represented in Equation 1. The corresponding units are reported assuming time was measured in hours and biomass was measured as OD units. Doing a linear least squares regression between $lnX$ and ($t-t0$) will yield estimates for $\mu$ (the slope) and $lnX0$ (the intercept). [64]

$$lnX = lnX_0 + \mu\,(t - t_0)$$

*Equation 1* – Exponential growth phase equation. $X$ = cell concentration (units of OD); $X_0$= initial cell concentration at the beginning of the phase (units of OD); $t$= time; $t_0$= time the phase starts (normally corresponds to the first time point within a growth phase); $\mu$= growth rate with units of 1/h or OD/h for exponential and linear growth, respectively.[64]

## 4.1.4 Substrate quantification

The previously collected supernatants were first diluted with $H_2SO_4$ 0.05 M solution, in a proportion of 1:20, and centrifuged (Sartorius 1-15P, Sigma) at 10000 rpm for 5 min, to precipitate any cellular content that remained in the sample. Following that, they were transferred to a high-performance liquid chromatography (HPLC) vial. The sugar quantification was performed by HPLC, using a system (Merck Hitachi, Darmstadt, Germany) equipped with a refractive index detector (L-7490, Merck Hitachi, Darmstadt, Germany) and an Rezex ROA Organic Acid $H^+$ (8%) column (300 mm× 7.8 mm, Phenomenex) at 65°C. Sulfuric acid (0.005 M) was used as mobile phase at 0.5 mL/min.

#### 4.1.4.1. Substrate uptake rate

The substrate uptake rate was estimated with a linear equation, with the corresponding units of time measured in hours, biomass measured as OD units, and substrate concentration measured as g/L.[64] The substrate consumption rates were converted into mmol/gDCW/h using the equation below:

$$\frac{\mathrm{X\ g\ glu/L/OD/h} * \frac{1}{0.30\ \frac{\mathrm{g\ DCW}}{\mathrm{L/OD}}}}{0.18\ \mathrm{g\ glu/mmol}}$$

*Equation 2* – Conversion of substrate uptake rate units of g glucose/L/OD/h to mmol/g DCW/h. *X* = Substrate uptake rate. 1 OD = 0.30 gDCW/L; Molar weight of glucose = 180 g glu/mol.

### 4.1.5. MEL and fatty acids quantification

In this study, the methyl ester derivatives were prepared to allow the quantification of MEL and fatty acids on the second growth rate experiment, with feeding of different concentrations at day 4 (chapter 4.1.2). At day 2, day 4, day 7 and day 10, samples of 3 mL of solution broth were recovered and lyophilised (Alpha 1-2 LD plus CHRIST®) for 24 hours. The obtained mass of each sample was weighted, giving the theoretical mass value, and transferred to a glass tube, where the sample was once more weighted, giving the experimental mass value that later was transesterified.

The purified MEL fraction was mixed with 2 mL of HCl/methanol solution and 10 µL of standard solution, composed of *n*-hexane and heptanoic acid, and reacted at 80ºC (Memmert) for 60 minutes. Afterward, the transesterification reaction was stopped by adding 1 mL of water and 1 mL of hexane. This generated a two-layer solution, the lower an aqueous layer and the upper the ester layer. Subsequently, the organic phase was dried with anhydrous sodium sulphate, filtered, and collected in vials. 1 µL of the organic phase containing isolated MELs was injected into the Gas Chromatography (GC) system (Hewlett-Packard, HP5890) equipped with a FID detector and an Agilent HP Ultra2 capillary column. The oven temperature gradients were set, starting from 140 ºC to 170 ºC at 15 ºC/min, 170 to 210 ºC at 40 ºC/min and 210 to 310 ºC at 50 ºC/min and a final 310 ºC for 3 minutes.

MEL's distinctive characteristics, including its two lipid chains with carbons ranging from 8 to 14, were taken into account when it was quantified (other lipid chains of fatty acids existing in the yeast are usually longer and soluble organic acids chains have usually less than eight carbons). MEL were measured using the quantities of C8, C10, C12 and C14 fatty acids with molecular weights ranging from 144 to 242 g/mol depending on the length of the two-acyl chain (C8–C14). Fatty acids chains concentration values were quantified by C16 and C18 quantities on each sample.

#### 4.1.5.1. MEL production and fluxes determination

The flux of MEL, also denominated productivity of the reaction, was obtained from the values obtained by GC analysis. The values of MEL and fatty acid concentration, with units of g/L, were used to find the fluxes, using mmol/g DCW/h units. This was possible by considering the absorbance values (OD units) obtained at the same time (in hours) that the GC samples were collected, following the Equation 3 presented. Having MEL production values, as g/L, was necessary to reach the corresponding flux units, that later allowed a fit of the experimental values to the model.

$$\frac{\mathrm{Y\ g\ MEL/L}}{\frac{X}{t}} * \frac{\frac{1}{0.30\ \frac{\mathrm{g\ DCW}}{\mathrm{L/OD}}}}{0.62\ \mathrm{g\ MEL/mmol}}$$

*Equation 3* – Conversion of MEL concentration units of g MEL/L to mmol/g DCW/h. Y= MEL concentration (g/L); $X$ = cell concentration (units of OD); $t$= time (hours). 1 OD = 0.30 gDCW/L; Molar weight of MEL = 616 g MEL/mol.

### 4.1.6. MEL and fatty acids characterization

Normal phase thin layer chromatography (TLC) plates (Macherey-Nagel Alugram Xtra SIL G/UV254) and a solvent system consisting of chloroform-methanol-water (6.5:1.5:0.2) were used for separating the MEL samples by polarity. The plates were sprayed with a matrix solution consisting of 1.5 g of naphthol, 51 mL of ethanol, 4 mL of water and 6.5 mL of sulfuric acid. [14] [16]

## 4.2. Simulation of yeast metabolism in MEL producing cells

*In silico* simulation of yeast metabolism was performed using COBRA (Constraint-based Reconstruction Analysis) toolbox and MATLAB R2021a software. MATLAB has a scripting language considerably easier when compared to learning how to build operating system-dependent standalone programs.[52] The metabolic model used was iUma22, the first genome-scale metabolic model (GSMM) of *Ustilago maydis*, downloaded from https://github.com/iAMB-RWTH-Aachen/Ustilago_maydis-GEM/tree/master/model. iUma22 was reconstructed from sequencing and annotation for the simulation of metabolic activities. [60]

This model was retrieved from literature, where the authors utilized the automated PathwayTools workflow and build the biomass equation using literature values and codon composition, to have more than 25% of annotated genes in the sequenced genome present in the final model. In order to test growth predictions, exponential batch cultivations were used and substrate usage was corrected using arrays. Through experiments with glucose growth, the model's quality was evaluated, and the precision of rate predictions was examined. The pan-genome of various *U. maydis* strains that had been annotated and allowed for the reconstruction of KEGG pathways was used to compare the metabolic capability of the model iUma22. The model can thus be used to improve metabolic

engineering strategies and to determine the biotechnological potential of the overproduction of metabolites. [60,65]

The GSMM model is characterized by 1855 reactions and 1233 species of metabolites. All the reactions are assigned with abbreviations, depending on the objective of the reaction, and all the metabolites are coded as abbreviations of the name, followed by the letter of the compartment where they belong, with [e] being Extracellular space, [c] the Cytoplasm and [m] the Mitochondrial Lumen. The metabolites also have the respective ID code, if registered, on KEGG (Kyoto Encyclopedia of Genes and Genomes) database, with each metabolite being represented as CXXXXX, with X being a number between 0 and 9. It is also available the corresponding ID to two other databases: PubChem ID and ChEBI ID.

### 4.2.1. Pathways Reconstruction

Taking this information into consideration, the available model for *U. maydis*, retrieved from GitHub database, has to be modelled according to previous chapter approaches (Chapter 2.4.), with MATLAB as the programming computing platform where fluxes constraints values should be added. However, the desired model for MEL synthesis has to consider the metabolism characteristics of *M. antarcticus*, the desired organism to do this biosurfactant production, and the highlighted differences between the two microorganisms (Chapter 2.3.). A general overview of the metabolic pathways necessary to have MEL production is presented in Figure 23, covering the pathways previously mentioned.

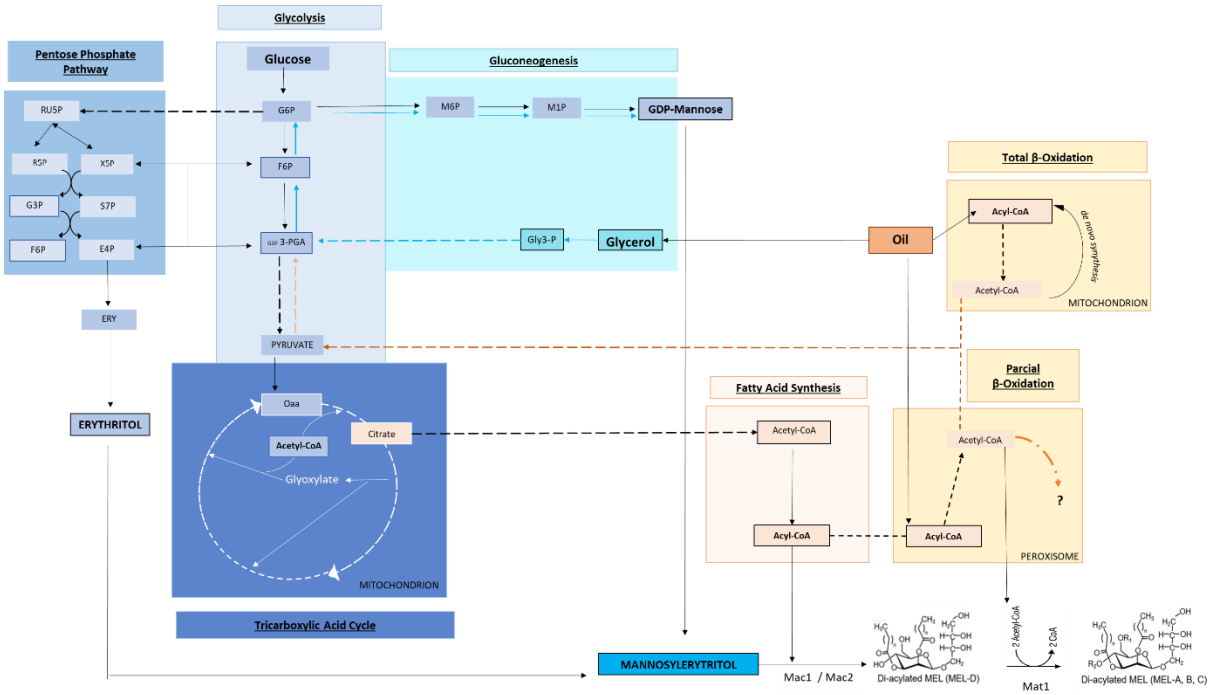



**Figure 23** - Simplified overall metabolism for MEL production, with glycolysis, tricarboxylic acid cycle, pentose phosphate pathway, gluconeogenesis, fatty acid synthesis, mitochondrial *β*-oxidation and peroxisomal *β*-oxidation. The solid lines represent reactions between metabolites. The dashed lines represent reactions between various metabolites that are not present in the image.

Of the 1855 coded reactions, the reactions that belong to pathways leading to the production of the two principal key building blocks, D-mannose, and erythritol, were manually selected. The metabolic pathways of these reactions are glycolysis with glucose as substrate, gluconeogenesis from glycerol derived from oil substrate, the pentose phosphate pathway, and the tricarboxylic acid cycle. Of the 1233 metabolites, at least 140 metabolites were selected as belonging to the desired metabolism (Table A1). Mannosylerythritol (ME) and all the homologs of MEL are present in this model but are not coded in KEGG database. This means that the final product, the objective of the metabolic model of this work, is not yet coded in a database.

To have a better insight into the model, a graphical figure with the most important pathways leading to the production of MEL has to be constructed. The first approach was to highlight all the reactions and manually construct an image of the metabolic pathways that produce MEL, in *Excel*. Afterward, converting the metabolic model iUma22 from SBML to a JSON file, it became possible to have a building construction with *Escher*. A script retrieved from GitHub also has the main pathways of this model, nevertheless, it was necessary to add to the reactions from gluconeogenesis, MEL synthesis, and from fatty acids in the build option of Escher software (Figure 24).

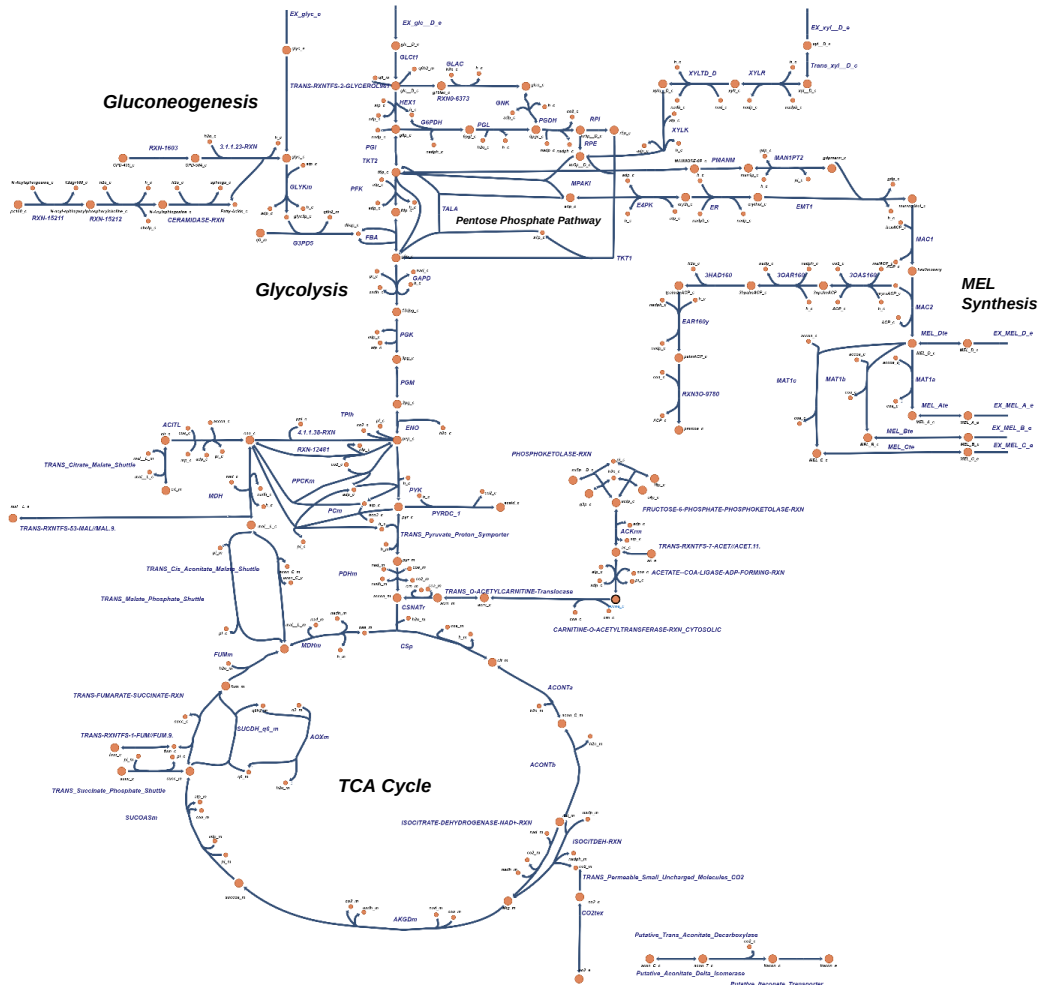



**Figure 24** - Escher visualization of metabolism for MEL production in *Ustilago maydis*, with reactions and metabolites dataset loaded from the model iUma22. Highlighted pathways: glycolysis, gluconeogenesis, pentose phosphate pathway, tricarboxylic acid cycle and MEL synthesis reactions.

### 4.2.2. Flux Balance Analysis

The prediction of an ideal steady-state flow vector that maximizes a microbial biomass production rate is a common use of the flux balance analysis (FBA).[66] The versatile function `optimizeCbModel` can be used to compute FBA and many of its variants. That is, optimizeCbModel's default method is the FBA approach, but depending on the optional parameters provided to `optimizeCbModel`, a variety of methods that are modifications on FBA, are also implemented and accessible with minor adjustments to the input parameters. There are two possibilities: either calculate an FBA solution value or a unique flux balance solution.

Even if the meanings of the fields differ, depending on the optional input arguments to the function, the solution structure `FBAsolution` from `optimizeCbModel` always has the same form and a standardized solver status is stored in `field.stat` (Table 1). The solver-specific status `origStat.` is translated into the `field.stat`. This status is unique to each numerical optimization solver, and it is converted to the standardized solver status for other COBRA Toolbox functions to operate in a manner that is as invariant as feasible, with respect to the underlying solver. [1]

**Table 1** – Solver status in standardized form values and meaning for each one (adapted from [1]).

| `FBAsolution.stat` | **Meaning** |
|---|---|
| -1 | Because of a time constraint or numerical difficulties, no solution is reported. |
| 0 | The constraints in the problem restrict any feasible steady-state flux vector from being generated, therefore no optimal solution can be found. |
| 1 | Optimal solution has been discovered and is returned. |
| 2 | The lower and upper bounds are inadequate to constrain the objective function value, and the problem is unbounded, hence no optimal solution is provided. |

The flux balance analysis approach can be used to calculate growth rates of *U. maydis* on glucose, or it can also be used to simulate growth on other substrates (Figure 25). Nonetheless, the objective of this work is to upgrade the available GSMM for MEL production by *Moesziomyces antarcticus*.

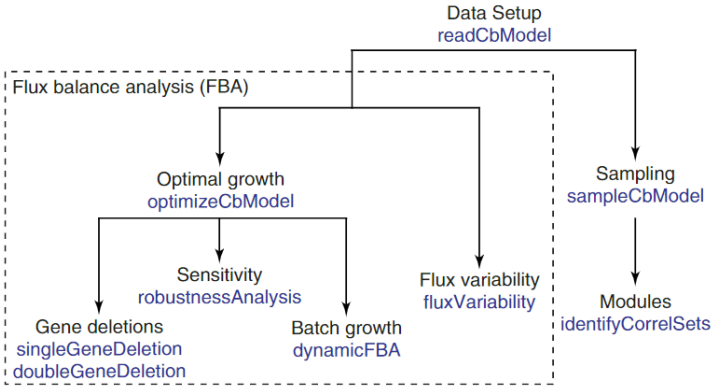



**Figure 25 -** Workflow to use the COBRA Toolbox. The internal implementation of the methods at higher levels is relied on by functions at lower levels of the hierarchy. For example, the gene deletion functions calculate optimal growth for each feasible metabolic network and the reactions associated with one or two genes deleted. The majority of the methods are based on FBA principles. [51]

The model presented contains some reactions that do not exist on *M. antarcticus* and need to be deleted. As so, the first procedure was to delete 15 reactions (Table A2.1) and consequently eliminate 11 metabolites (Table A2.2), related to the production of Ustilagic Acid, which is graphically shown in Figure 26. This was possible using the command:

```
>> model = removeRxns(model, rxnRemoveList);
```

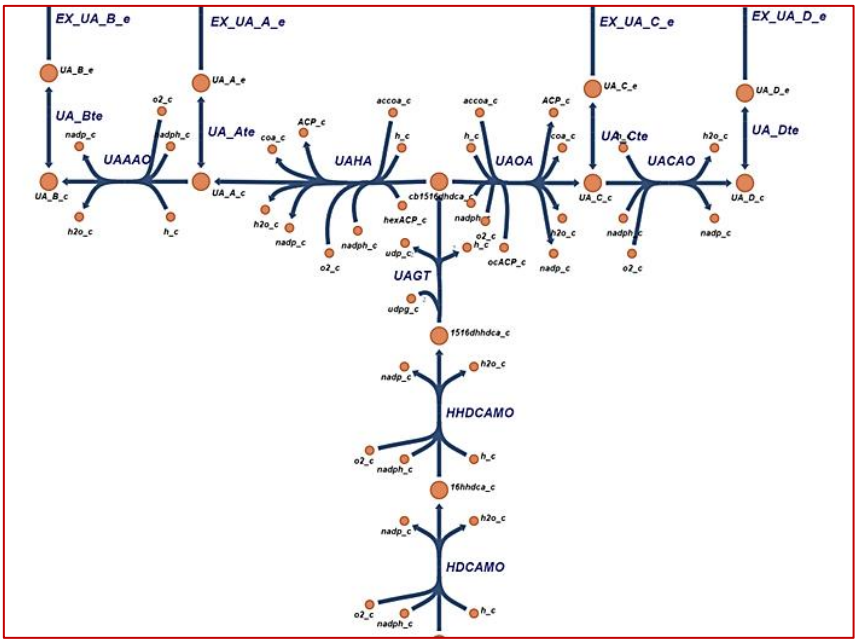



**Figure 26** – Highlight of the deleted reactions from the iUma22 model, made with the Escher building application.

The model contains three distinct compartments: the cytosol, the mitochondria, and the extracellular space. Another organelle existent in *M. antarcticus*, which may be important for the correct metabolism, and was already covered as being crucial to enable the production of Mannosylerythritol lipids by oils as substrate, is the peroxisome. In order to allow the synthesis of MELs by the stoichiometric model, it was necessary to add an exchange reaction for oils and to link the dead-end reaction present in the model, which produced fatty acids, to the lipids metabolism (Table A3.1). The command throughout is possible to add reactions in the COBRA toolbox is the following:

```
>>[model, rxnIDexists] = addReaction(model, 'rxnID', 'reactionFormula',
'Formula', varargin)
```

Simulation experiments were carried out with different values for glucose maximum uptake rate in units of mmol g $DCW^{-1}$ $h^{-1}$. Other additional constraints could be applied to the stoichiometric model and examples of commands to run in COBRA toolbox are presented in Table 2.

**Table 2** - Example of COBRA Toolbox commands to perform flux balance analysis. [2]

| Action | Command |
|---|---|
| Change bounds for anaerobic growth | model = changeRxnBounds(model,'EX_o2(e)',0,'l'); |
| Change bounds for aerobic growth | model = changeRxnBounds(model,'EX_o2(e)',-1000,'l'); |
| Change glucose uptake rate to 10 mmol $gDCW^{-1}$ $h^{-1}$ | model = changeRxnBounds(model,'EX_glc(e)',-10,'l'); |
| Simulate maximum growth by FBA | solution = optimizeCbModel(model); |
| Simulate maximum growth of regulated model | [FBAsols,DRgenes,constrainedRxns,cycleStart,states] = optimizeRegModel(model); |
| Change objective maximum ATP yield | model = changeObjective(model,'ATPM'); |

Other frameworks, like OptKnock or OptGene can be used to suggest possible gene knockout strategies that could improve MEL production in *U. maydis* or *M. antarcticus*. The production titre obtained upon the gene deletions suggested by these approaches can be further examined using the maximization of biomass per flux unit as the objective function. For example, anaerobic environments are modelled when the lower bound of the $O_2$ exchange reaction is constrained to zero flux value, enabling no $O_2$ to enter the system, which allows the identification of growth-coupled designs for numerous metabolites in the model under anaerobic conditions.

To simulate various media, all metabolites present in the media should have exchange reactions with lower bounds constrained to their intended uptake rate, while all the metabolites that are not present in the cell should have exchange reactions with values for lower bounds confined to zero. [2] Since in this case, the goal was to maximize the conversion of known metabolites such as erythritol and GDP-mannose in mannosylerythritol, the objective reaction of the model was changed with the following command, to retrieve an FBA solution value that represents the maximization of production of MEL D

```
>> model = changeObjective(model,'MAC2');
```

The corresponding FBA simulations of this work were automatically generated with MATLAB software, by constraining lower values of input of carbon sources. The corresponding script is described in the appendix section.

# 5. Results

## 5.1. *In silico* metabolism of *Ustilago maydis* model

### 5.1.1. Targeting and visualization

A metabolic flux model that describes the main metabolic pathway used for MEL production was designed, focusing on the production of mannosylerythritol (ME) and lipids. This model describes specific cell features concerning central carbon metabolism and cell bioenergetics, highlighting the pathways for the production and assembling of the MEL building blocks. To achieve a better understanding of the pathways necessary to produce MEL, a graphic pathway image was constructed (Figure 27) in *Excel*. The main routes, including glycolysis, the pentose phosphate pathway, the tricarboxylic acid cycle, gluconeogenesis, fatty acid synthesis, and *β*-oxidations (partial and complete) are highlighted.

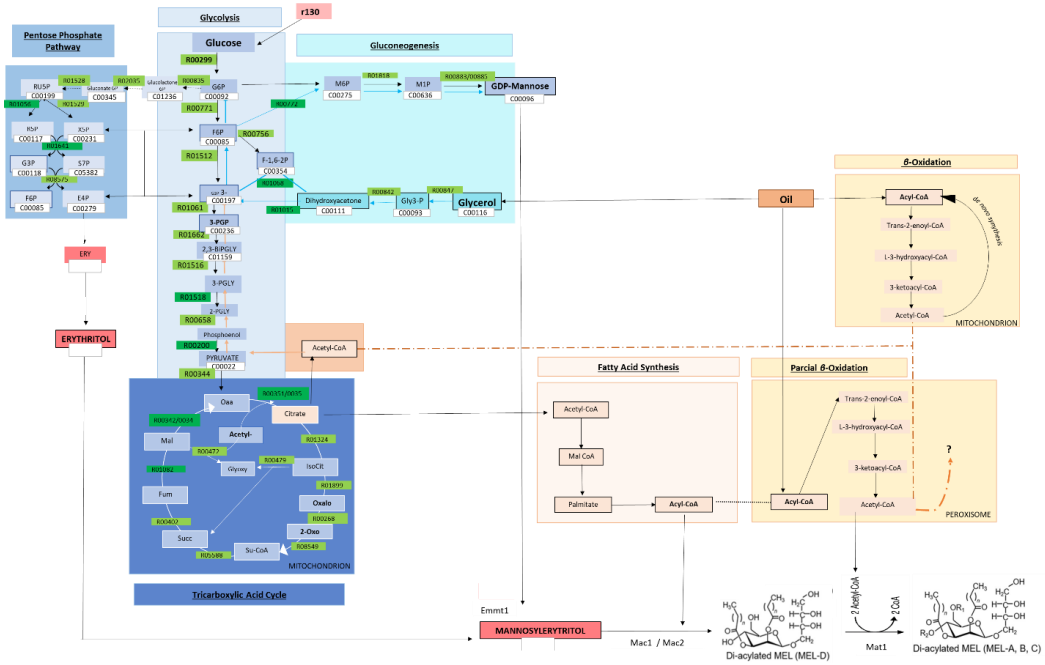



**Figure 27** - Graphic model of overall metabolism for MEL production, with glycolysis, tricarboxylic acid cycle, pentose phosphate pathway, gluconeogenesis, fatty acid synthesis, mitochondrial *β*-oxidation, and peroxisomal *β*-oxidation. The solid lines represent reactions between metabolites (adapted from [14,16,25,30] ). The KEGG IDs for each reaction are in green and the KEGG ID code for each metabolite is presented in white (also presented in Table A1).

Nevertheless, the dynamic visualization of the pathways was desired, with the purpose of a better and cleaner design, but also to make it possible to see where the fluxes values retrieved by MATLAB are present on the cell. This can be achieved with the Escher program and the script available by the authors of the paper. With the building option of this software, it was possible to add the desired reactions corresponding to pathways covered in Chapter 2.3, leading to the final map visualization represented in Figure 28.

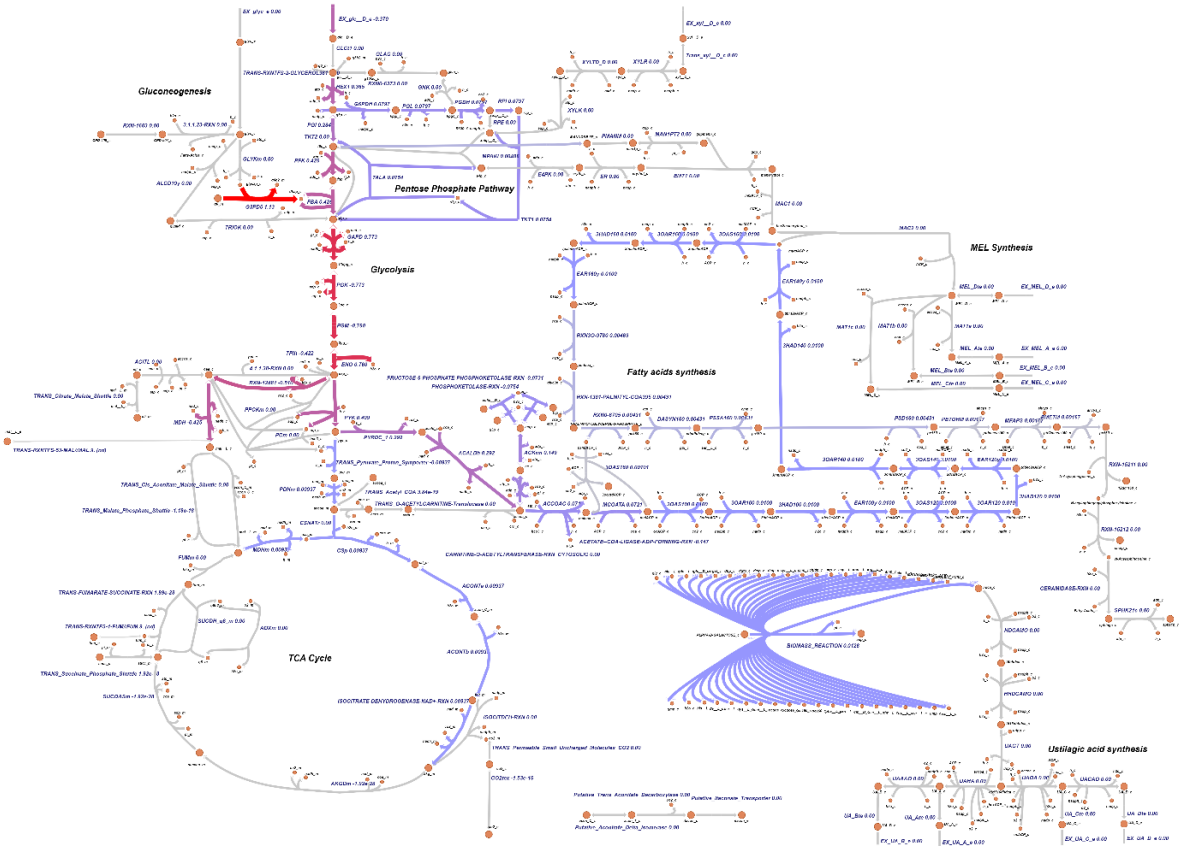



**Figure 28** - Visualization of fluxes of the principal pathways that lead to MEL production in *U. maydis*, with maximization of biomass reaction, in the iUma22 model with Escher-FBA application.

## 5.1.2. *Ustilago maydis* model parametrization

The model retrieved from GitHub, and also available in the Biomodels database [67] and constructed based on *Ustilago Maydis* metabolism, was analysed in MATLAB. As non-mathematical model organisms, *Escherichia coli* is the best prokaryotic model and *Saccharomyces cerevisiae* is the best eukaryotic model. In such models, according to the genotype, the objective is to model the phenotype. In mathematical models' systems, the phenotype, as the objective, is the cellular growth and the genotype that will be modelled is the metabolic network. Hence, the FBA approach was used to calculate the growth rates of *U. maydis* on glucose media, and flux values constraints were applied to understand how the reactions in the model change their flux values accordingly.

For reversible and irreversible reactions, the bounds set on reaction rates in a metabolic model typically vary from -1000 to 1000 and 0 to 1000, respectively. [47] The values of fluxes come in mmol/gDCW/h units, and to constrain a specific substrate flux, a value between the bounds values previously mentioned should be applied. Imposing different flux values for a substrate, even if just a small change in the input, should have a complex response in the metabolism. As so, at least one of the reactions in the model must have a constrained lower or upper bound, for the estimated fluxes to be

meaningful. [51] Given some network constraints, the flux variability analysis (FVA) approach, another possible analysis approach, estimates the minimum and maximum value of a reaction flux. [68]

In this case, values for glucose uptake were considered. The attributed values for fluxes will generate a solution, according to the premise that the objective function has to be optimised. An objective function is a network reaction, or a linear combination of network reactions, for which a linear programming problem is optimized. Usually, it is a given biological objective, commonly the maximization of biomass production. [68] The solution value should be a value of flux flowing through it, that corresponds to the organism's exponential growth rate. It is possible to compute a solution to the flux balance analysis problem using the following command in COBRA toolbox:

```
>> FBAsolution=optimizeCbModel(model);[1]
```

The value obtained from the previous command, coming as an f, is provided by applying the FBA approach. If f = 0 and `FBAsolution.stat = 0`, the model does not grow, with turns it into an inconsistent model. However, the solution obtained, applying different flux values for glucose, in most cases generated a value of f ≠ 0 and `FBAsolution.stat = 1`. According to what was previously described in Chapter 4.2.2 - Table 1, this `FBAsolution.stat` means that an optimal solution has been discovered and is returned.

Different initial concentrations of glucose were considered, and the respective substrate uptake rates were applied as glucose input constraints, to have the corresponding predicted growth values given by MATLAB (Table 3). With these values, it was possible to compare the experimental and the *in silico* outputs, which gave the difference between them as percentage values. The values obtained were registered between 5% and 9%, with the maximum glucose concentrations having values of 122% of error and the lower value registering 47%. The lower value of this experiment should not be considered, since the value obtained as substrate uptake rate is not realistic, since it is a very high value, and the biomass values obtained registered some outliers. [60]

**Table 3** – Substrate uptake rate values and growth rate values, with glucose as substrate, retrieved from the literature [60], and compared with the FBA solution given by MATLAB, with *U. maydis*.

| Experiment ID | Initial Concentration (g/L) | Substrate-uptake (mmol/gDCW/h) | Experimental Growth Rate ($h^{-1}$) | Predicted Growth Rate ($h^{-1}$) | Difference | Error (%) |
|---|---|---|---|---|---|---|
| **2196_glc** | 19 | 6.24 | 0.27 | 0.51 | 0.24 | 47 |
| **2229v1** | 50 | 2.20 | 0.18 | 0.17 | 0.01 | 7 |
| **Wierckx_50glc** | 54 | 1.22 | 0.08 | 0.09 | 0.01 | 6 |
| **Wierckx_100glc** | 106 | 0.67 | 0.04 | 0.04 | 0.00 | 5 |
| **LV3_130v1** | 126 | 1.10 | 0.07 | 0.08 | 0.01 | 7 |
| **LV3_130v2** | 132 | 0.74 | 0.04 | 0.04 | 0.00 | 9 |
| **LV3_200v1** | 203 | 0.33 | 0.02 | 0.01 | 0.01 | 122 |
| **LV3_200v2** | 216 | 0.36 | 0.02 | 0.01 | 0.01 | 122 |

As shown in Figure 29, or in the appendix section in Figure A1, growth with different carbon sources was simulated through Escher-FBA application when biomass growth is maximized. This allows an understanding of how the values from the fluxes changes, and which pathway is taken, when an arbitrary input value of 10 mmol/gDCW/h is given to each source (glucose, xylose or glycerol).

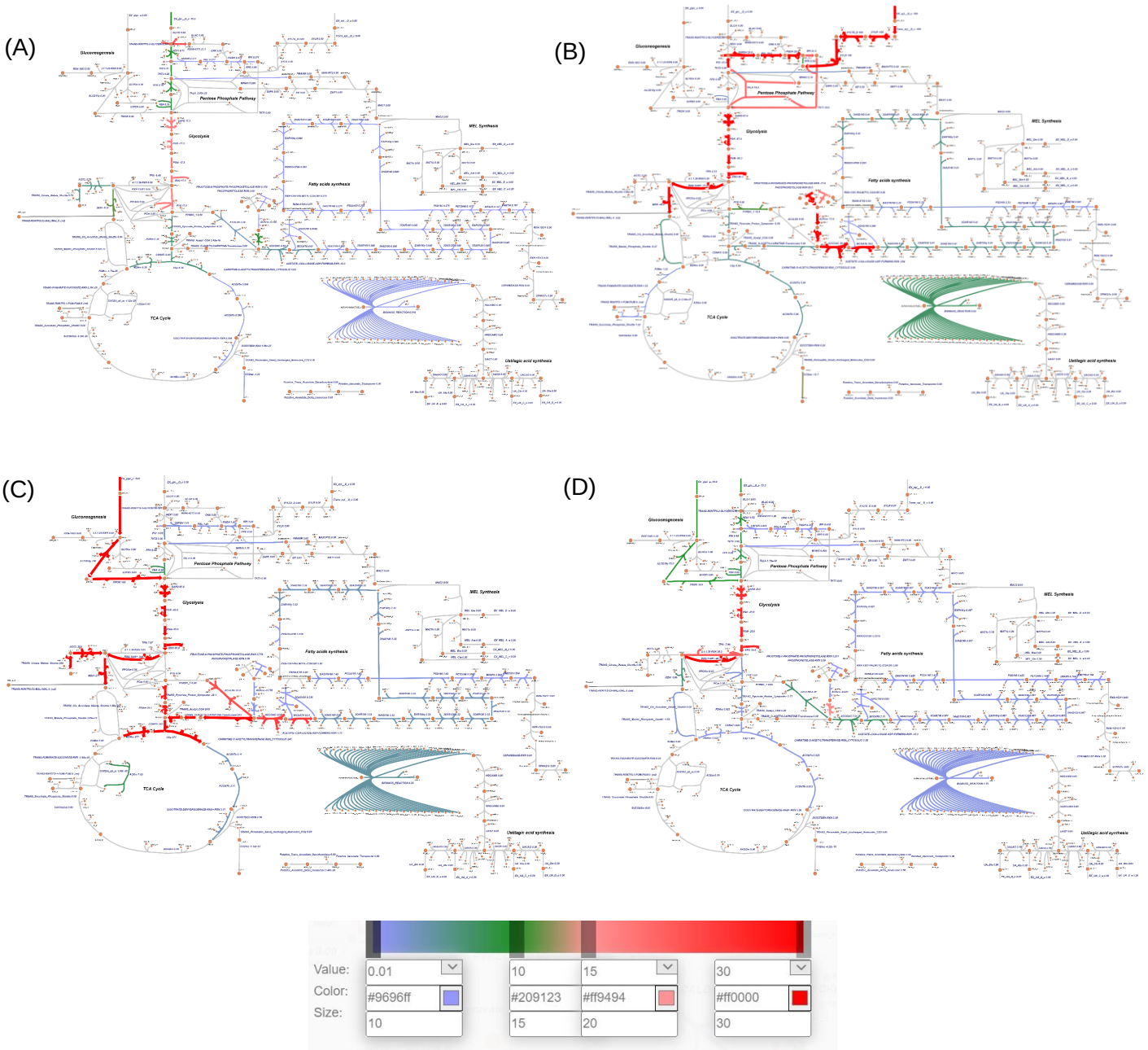



**Figure 29** - Examples of Escher-FBA simulations. (A) Simulated growth with glucose as a carbon source. Flux value of entry of glucose = 10 mmol/gDCW/h. (B) Simulated growth with xylose as a carbon source. Flux value of entry of xylose = 100 mmol/gDCW/h .(C) Simulated growth with glycerol as carbon source. Flux value of entry of glycerol = 100 mmol/gDCW/h. (D) Simulated growth with glucose and glycerol as carbon sources. Flux value of entry of glucose = 10 mmol/gDCW/h. Flux value of entry of glycerol = 10 mmol/gDCW/h. All fluxes are in units of mmol/gDCW/h and are shown by colors according to the value: 0 (grey); 0.1 – 2 (light purple); 2 – 14 (green); 14-20 (light red); Bigger than 20 (red).

### 5.1.2.1. Deleted reactions from *Ustilago maydis* model

The upgrade of the *Ustilago maydis* model, iUma22, to a *Moesziomyces antarcticus* model must include the deletions of reactions and metabolites that are present in *U. maydis*, but not in *M. antarcticus*. Although the two organisms share the same lineage, secondary metabolites like Ustilagic acid [69], is exclusively produced by *U. maydis*. Therefore, as already explained in Chapter 4.2.2., the reactions leading to the production of Ustilagic Acid were deleted, and the consequent FBA solution values were obtained, as shown in Table 4, having biomass equation and MEL production as objectives.

**Table 4** - *In silico* flux values calculated in COBRA toolbox in MATLAB. Predicted values with iUma22 model and model with deleted reactions of ustilagic acid production, with biomass reaction and MAC2 reaction optimized. The constraint applied was the input of the glucose flux of 2.20 mmol/gDCW/h (corresponding to 50 g/L of initial glucose).

| Conditions | Reactions | iUma22 model | iUma22 model with deletions |
|---|---|---|---|
| **For Biomass maximization**<br><br>**Glucose input (50g/L) = 2.20 mmol/gDCW/h** | FBA Solution | 0.169 | 0.169 |
| | BIOMASS_REACTION | 0.169 | 0.169 |
| | EX_o2_e | -0.011 | -0.011 |
| | EX_nh4_e | -1.113 | -0.796 |
| | EX_pi_e | 7.087 | 7.087 |
| | EX_so4_e | -0.014 | -0.014 |
| | EX_h2o_e | -3.830 | -3.988 |
| | EX_glc__D_e | -2.200 | -2.200 |
| | EX_gly_e | 0.317 | 0.000 |
| | EX_mal__L_e | 1.332 | 1.491 |
| **MAC2 reaction maximization**<br><br>**Glucose input (50g/L) = 2.20 mmol/gDCW/h** | FBA solution | 0.317 | 0.317 |
| | EX_co2_e | 2.366 | 2.366 |
| | EX_nh4_e | -0.106 | -0.106 |
| | EX_h2o_e | 3.383 | 3.383 |
| | EX_glc__D_e | -2.200 | -2.200 |
| | EX_mal__L_e | 0.277 | 0.277 |
| | EX_gly_e | 0.106 | 0.106 |
| | MAC2 | 0.317 | 0.317 |
| | EX_MEL_D_e | 0.317 | 0.317 |
| | E4PK; ER; EMT1 MAC1; MAC2; MEL_Dte | 0.317 | 0.317 |

The values obtained showed some flux values slightly different from the values previously obtained without the deletion of the 15 reactions that lead to the production of a product that is only present in *U.maydis,* only in the condition where biomass reaction was the objective reaction. All the predictions were performed with the same constraint, the flux value of glucose input equal to 2.20 mmol/gDCW/h, which corresponds to an initial experimental concentration of 50 g/L of D-glucose.

Even if the same FBA solution is obtained, the value of flux of $NH_4$, $H_2O$, malate, and the absence of entry of glycerol are the four fluxes that registered changes from the values obtained on iUma22 model, when biomass is the reaction optimized. The need for less $NH_4$ entering the cell was noticed, registered with a difference of 0.32 mmol/gDCW/h, oppositely to the demand of more $H_2O$, with more 0.16 mmol/gDCW/h then the value on the original model. Malate, contrarily, exited the cell with a higher flux value, with a difference of 0.16 mmol/gDCW/h, and glycerol had a flux value equal to 0.00 mmol/gDCW/h, whereas the value registered in the original model was 0.32 mmol/gDCW/h. Thus, it seems that the flux of $NH_4$ is directly related to the flux of glycerol, and the flux of $H_2O$ is related to the flux of malate since it presented the same flux values of difference.

With the maximization of MAC2 reaction (Table 4), the reaction that leads to the production of MEL-D, one of the four homologs of MEL, no relevant differences were detected between the original model and the model with deletions. Nevertheless, it is possible to compare the differences between the flux values of the exchange reactions that are active when the objective function is the biomass reaction. With the same flux value of glucose entering the cell, with the maximization of MEL production, the biomass reaction does not present any flux. The entry of oxygen ($O_2$) with a value of 0.01 mmol/gDCW/h when biomass was the objective reaction, is 0.00 mmol/gDCW/h with MEL production as the objective, that also presented the production of $CO_2$, with a flux value of 2.37 mmol/gDCW/h. The glycerol flux value is also lower, which means that the flux of production is decreased.

Regarding MEL biosynthesis reactions that are active when MAC2 reaction is optimized, it is possible to have the following reactions registering the same flux: E4PK - Erythrose-4-phosphate kinase; ER - Erythrose reductase; EMT1 - Glycosyltransferase (assembly of GDP-mannose and erythritol) MAC1 - Mannosylerithritol C2 Acyltransferase; MAC2 - Mannosylerithritol C3 Acyltransferase; MEL_Dte – MEL D Transporter and EX_MEL_D_e (Exchange of MEL D). The same outcome was observed when these reactions suffered a “knock-out” on the Escher-FBA application, which presented the same values of fluxes as the previous chapter.

#### 5.1.2.2. Added reactions to *Ustilago maydis* model

Along with the fact that the GSMM presented for *U. maydis* have specific reactions of that organism, that needed to be deleted, it also lacks some reactions present in *M. antarcticus*. As previously mentioned, since peroxisome is not present in this model, the main pathways that are lacking are related to lipids' metabolic reactions. It is possible to access that “Fatty-Acids_c[c]” metabolite, as the “Glycerol” metabolite, are present in this model at the cytoplasm. However, they are not derived from the same metabolite, since fatty acids come from a mono-glyceride “CPD-504_c[c]”, that do not come from any other reaction.

This means that this mono-glyceride is derived from a dead-end reaction, where the metabolite does not present any reaction that leads to its production. Having dead-end metabolites in the model, meaning that some reactions connecting specific metabolites are not present, will lead to the accumulation of certain metabolites in the cytosol and that some others appear in the cell without

explanation, contrary to the steady-state balance assumption that FBA approach relies on. At the same time, “Fatty acids” is a metabolite that do not have any connection to acetyl-CoA metabolite.

To fix the dead-end reaction, an extracellular oil metabolite and a cytoplasmatic oil metabolite needed to be added to the model, as well as the corresponding reactions of extracellular entry, transport, and conversion into the monoglyceride and the glycerol. To connect “Fatty acids” to “Acetyl-CoA”, a simple reaction was also added, since these two metabolites already existed on the model, but are not linked (Figure 30).

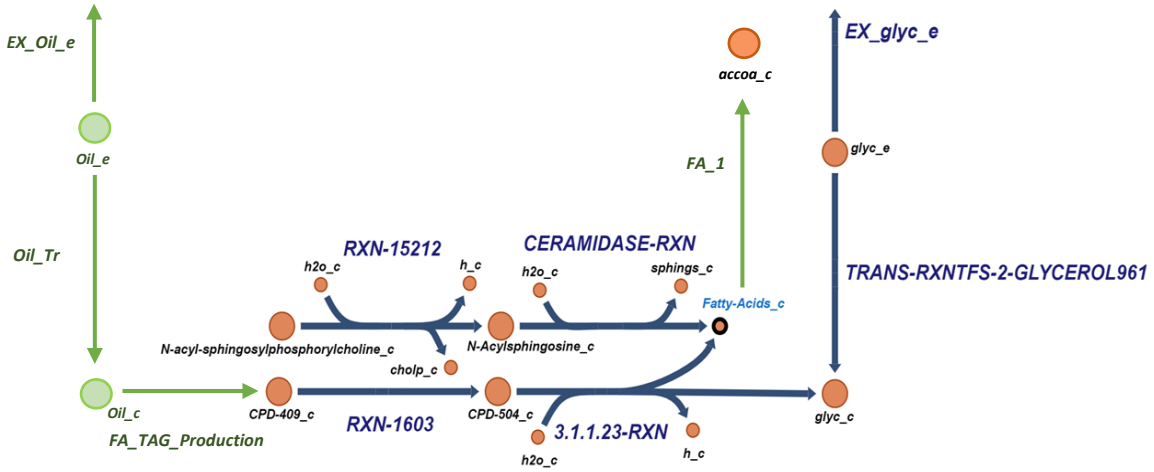



**Figure 30** – Addition of reactions and metabolites to fix dead-end reactions. Added reactions and added metabolites (green), reactions (blue) and metabolites (orange) already present in the iUma22 model. The original reactions are visualized by Escher software with building option.

The addition of this metabolite and these four reactions lead to minimal changes in the flux values, compared with values obtained to the iUma22 model without this modification (Table A3.2), having the input value for glucose of 2.20 mmol/gDCW/h. Compared with values obtained for biomass optimization, residual flux values were registered to the production of $Ca_2$, while values of production of malate diminished by 0.029 mmol/gDCW/h, which represents 2% of the value obtained in the original model. This error value was also registered on the production of fumarate, which was just active in the model with the added reactions.

Following this, the lower boundary flux values of all the reactions added were constrained to arbitrary values, to perceive if these constraints could originate different values of lipids production. The majority of these changes lead to an infeasible solution, indicating the death of the cell.

## 5.2. The effect of the carbon source on growth and MEL production of *Moesziomyces* yeasts

### 5.2.1. Glucose, Galactose, and Lactose assessed as a carbon source for *Moesziomyces* yeasts

The three conditions with different sugars (glucose, lactose, and galactose) in shake flasks, described in Chapter 4.1.2., were analysed by HPLC (Chapter 4.1.4) for each strain. The peak intensity of each sugar, possible to attribute according to the retention time, was registered to calculate the sugar concentration in each sample (Figure A2). With the use of the different sugars, sugar consumption and biomass growth values were evaluated and the values retrieved are presented in Table 5. The highest sugar consumption rate was reached when lactose was the carbon source, with 0.65 g/L.h and 0.57 g/L.h , in *M. antarcticus* and *M. bullatus*, respectively.

**Table 5** – Sugar consumption (g/L.h), maximum biomass concentration (g/L), and biomass growth rate ($h^{-1}$) values for glucose, galactose, and lactose as carbon sources in fermentations with *M. antarcticus* and *M. bullatus* strains.

| Sugar | Strain | Sugar consumption (g/L.h) | Maximum biomass production (g/L) | Biomass growth rate ($h^{-1}$) |
|---|---|---|---|---|
| Glucose | *M. antarcticus* | 0.31 ± 0.06 | 9.5 (Day 6) | 0.107 ± 0.003 |
| | *M. bullatus* | 0.35 ± 0.03 | 7.8 (Day 6) | 0.091 ± 0.046 |
| Galactose | *M. antarcticus* | 0.41 ± 0.06 | 6.5 (Day 3) | 0.074 ± 0.001 |
| | *M. bullatus* | 0.40 ± 0.02 | 7.9 (Day 6) | 0.079 ± 0.002 |
| Lactose | *M. antarcticus* | 0.65 ± 0.06 | 10.5 (Day 2) | 0.136 ± 0.001 |
| | *M. bullatus* | 0.57 ± 0.05 | 12.3 (Day 3) | 0.041 ± 0.001 |

Regarding biomass production, *M. antarcticus* has a higher value when D-glucose and D-lactose are used as carbon sources, exhibiting a value of biomass growth rate of 0.136 $h^{-1}$, three times higher than the value of *M. bullatus* of 0.041 $h^{-1}$, when D-lactose was used, Nonetheless, approximated biomass growth rates were obtained in *M. antarcticus* and *M. aphidis* cultivations, using D-galactose (0.074 $h^{-1}$ and 0.079 $h^{-1}$, respectively).

As expected, [16] the highest concentrations of biomass were obtained with *M. antarcticus* at the end of the fermentation, when using D-glucose and D-lactose, with a value of 9.5 g/L and 9.3 g/L, respectively (Figure A3). Using D-galactose, the maximum biomass of 7.9 g/L was observed at day 6 in *M. bullatus*, a higher value if compared with the value of 6.5 g/L of biomass produced by *M. antarcticus* at day 3.

From Table 5, the maximum value of the concentration of biomass in all conditions is obtained with *M. bullatus* cultivation using D-lactose, on day 3, with a value of 12.3 g/L of biomass. The average value of maximum biomass per strain that were calculated was 8.83 g/L with *M. antarcticus* and 9.33 g/L with *M. bullatus*.

### 5.2.2. The effect of glucose on growth and MEL production of *Moesziomyces antarcticus*

Two experiments were realized, with *M. antarcticus* strain, and exclusively having glucose as a carbon source. The objective of these experiments was to retrieve the yeast growth rate values, to compare with values from the literature obtained with *U. maydis*, and to understand if the model is predicting values close to the ones getting experimentally. The substrate uptake rate was also calculated in both experiments, to have a flux from glucose possible to apply in the model as a constraint, achieved with the HPLC technique of sugar quantification.

The difference between the two experiments was only related to the conditions, since in the first experiment five different initial concentrations of glucose were applied, and in the second experiment five different concentrations of glucose were added after day 4, when the yeast growth is theoretically stabilized and the secondary metabolites, such as MELs, are produced. Such experimental values can be compared with the flux value of MEL production given by the model.

#### 5.2.2.1. Batch cultures at different glucose concentrations

With the experiment described in the previous chapter, *M. antarcticus* growth rate values were assessed with glucose as a carbon source, as described in Chapter 4.1.2. The biomass values registered until day 7 (168 hours) are represented in Figure 31.

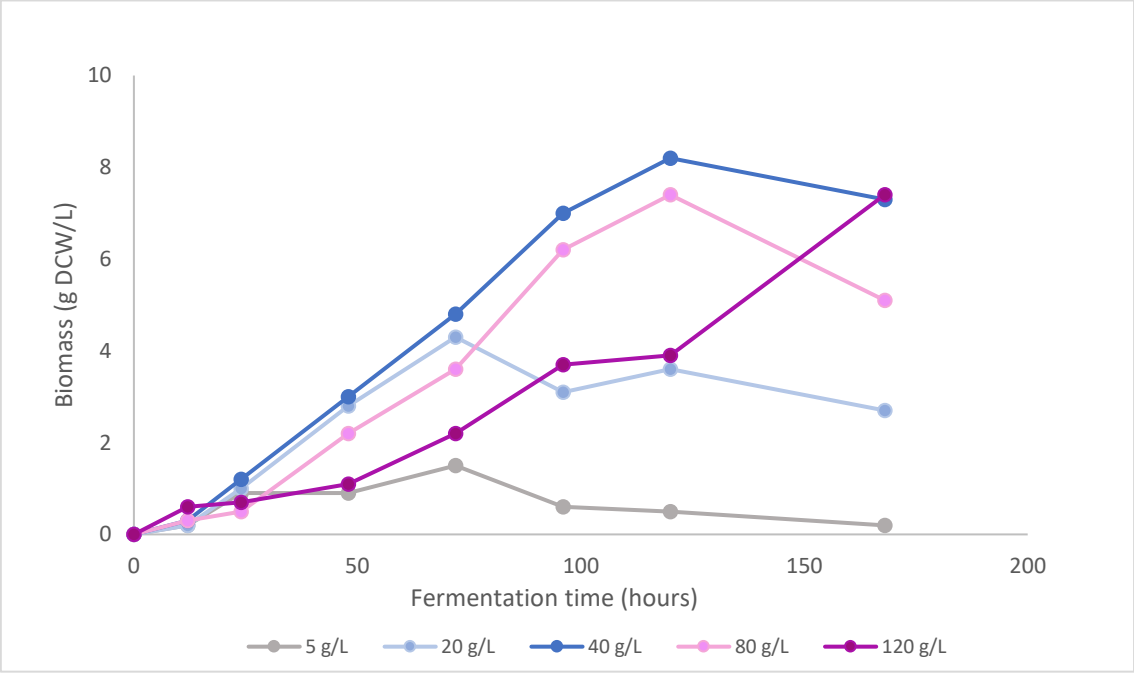



**Figure 31** – *M. antarcticus* growth registered as biomass values (g DCW/L) of different concentrations of glucose through time (hours). 5 g/L (grey); 20 g/L (light blue); 40 g/L (dark blue); 80 g/L (light pink); 120 g/L (dark pink).

In all the glucose concentrations, between 0 hours and 12 hours, the absorbance values registered showed an explicit augment (Figure A4). At 72 hours after the beginning of the experiment, the yeasts with 5 g/L and 20 g/L of glucose exhibited the highest value of absorbance within each record for the specific concentration. Subsequentially, the OD values measured for the flask with 5 g/L and 20 g/L of glucose started to decline, as well as the biomass value which also lowered after that point.

The other three concentrations, 40 g/L, 80 g/L, and 120 g/L of glucose, all presented constant growth until day 4 (96 hours). The lower concentration of this group, 40 g/L, presented values of 7.0 gDCW/L at 96 hours and 8.2 gDCW/L at 120 hours, which was the highest biomass value measured from all the glucose concentrations experimented, with a lower value of 7.3 gDCW/L at 168 hours. The concentration of 80 g/L of glucose exhibited a similar profile to the concentration below, reaching the highest value of 7.4 gDCW/L at 120 hours, showing a decline in the value, presenting just 5.1 gDCW/L at 168 hours. However, on day 7 (168 hours), the higher concentration of 120 g/L of glucose was still growing, with the last value registered of 7.4 gDCW/L, higher than the biomass value registered on day 5 with the same concentration.

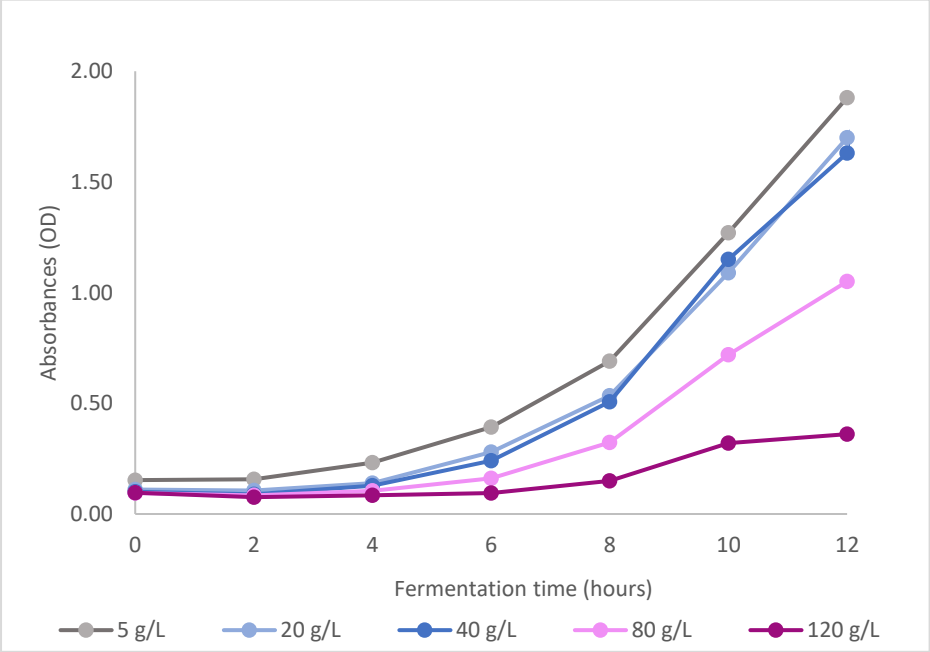



**Figure 32** - *M. antarcticus* growth values of absorbance (OD) through time (hours) with different glucose concentrations. 5 g/L (grey); 20 g/L (light blue); 40 g/L (dark blue); 80 g/L (light pink); 120 g/L (dark pink).

The HPLC values for sugar quantification were also retrieved and the substrate uptake rate was also calculated, as previously described (Chapter 4.1.4.1), and are represented in Table 6. With these values, and the OD values before stated, the growth rate value for each condition was obtained, and both were compared with literature values reached with similar conditions, but using *U. maydis* as the organism, instead of *M. antarcticus*.

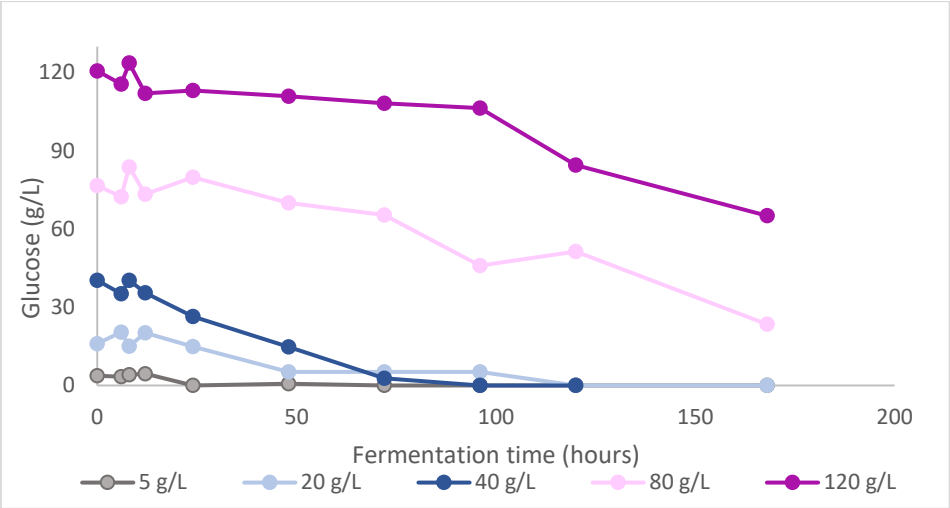



**Figure 33** - Glucose consumption of *M. antarcticus* with different initial glucose concentrations through time (hours). 5 g/L (grey); 20 g/L (light blue); 40 g/L (dark blue); 80 g/L (light pink); 120 g/L (dark pink).

As it is possible to realize from Figure 33, the glucose consumption presented a similar profile in the three lower concentrations, reaching 0 g/L on day 1 (24 hours) with the lowest concentration of 5 g/L of glucose, on day 3 (72 hours) with 20 g/L of glucose, and only at day 5 (120 hours) with 40 g/L of glucose. The two higher concentrations showed a decreasing concentration of sugar, as expected, with a linear decrease only from 72 hours to 80 g/L of glucose and from 96 hours to 120 g/L of glucose. None of the two concentrations reached 0 g/L of sugar on the medium, however with 80 g/L of initial glucose, the concentration value surpassed half of the initial concentration, reaching 23.48 g/L on day 7, and with 120 g/L of glucose reaching the value of 65.11 g/L in the medium.

**Table 6** – Growth rate ($h^{-1}$) values and substrate uptake values (mmol/gDCW/h) obtained with experiments described in the literature with *U. maydis* [60,70] and obtained in this work with *M. antarcticus* having glucose as the carbon source. The values are presented in order of increasing initial glucose concentration.

| Experiment (glucose g/L) | Hours | Initial Concentration (g/L) | Growth Rate ($h^{-1}$) | Substrate-uptake (mmol/gDCW/h) |
|---|---|---|---|---|
| *M. antarcticus* 5 g/L | 0-168 | 4 | 0.16 | 2.05 |
| *M. antarcticus* 20 g/L | 0-168 | 15 | 0.20 | 2.30 |
| *U. maydis* 20 g/L [60] | 0-64 | 19 | 0.27 | 6.24 |
| *M. antarcticus* 40 g/L | 0-168 | 32 | 0.03 | 0.43 |
| *U. maydis* 50 g/L [60] | 0-50 | 50 | 0.18 | 2.20 |
| *U. maydis* 50 g/L [70] | 0-120 | 54 | 0.08 | 1.22 |
| *M. antarcticus* 80 g/L | 0-168 | 77 | 0.18 | 2.69 |
| *U. maydis* 100 g/L [70] | 0-120 | 106 | 0.04 | 0.67 |
| *M. antarcticus* 120 g/L | 0-168 | 121 | 0.06 | 0.37 |
| *U. maydis* 130 g/L [60] | 0-264 | 126 | 0.07 | 1.10 |
| *U. maydis* 130 g/L [60] | 0-264 | 132 | 0.04 | 0.74 |
| *U. maydis* 200 g/L [60] | 0-264 | 203 | 0.02 | 0.33 |
| *U. maydis* 200 g/L [60] | 0-264 | 216 | 0.02 | 0.36 |

On Table 6 are represented the values of the growth rate and substrate uptake rate obtained, with an initial concentration of glucose ranging from 5 g/L to 200 g/L. As it is possible to conclude, the highest values of both growth and substrate rate are obtained with the lowest concentrations of glucose in the experiment here described with 4 g/L of glucose. This was expected since the values obtained in the literature using 19 g/L of glucose were 6.24 mmol/gDCW/h for substrate uptake rate and 0.27 $h^{-1}$ of growth rate. In fact, in the literature, the value obtained with *U. maydis* using 19 g/L of glucose was not further considered as a value to be given in the model. This assumes that the iUma22 model was not constructed to predict conditions of low glucose concentrations, which is in accordance with the unrealistic high substrate-uptake values obtained with *M. antarcticus* in this work. Therefore, the three concentrations of 4 g/L, 15 g/L, and 19 g/L will not be considered for the following description of results.

The highest growth rate value, of 0.18 h-1, was reached with the concentration of 77 g/L of glucose in the experiment before described with *M. antarcticus*, and with 50 g/L of glucose with *U. maydis*. The lowest growth rate value of 0.02 $h^{-1}$ was obtained when 200 g/L of glucose is given to *U. maydis*. With *M. antarcticus* the lower value of growth rate to be considered must be 0.06 $h^{-1}$, achieved when 120 g/L of glucose is present on the medium. These two concentration values are the highest presented, for each strain, in the range of all the concentrations presented. The growth rate value of 0.03 $h^{-1}$, obtained when 40 g/L of glucose is used to feed *M. antarcticus* strain, is not realistic, as the excepted value should be in the range of 0.16 $h^{-1}$ – 0.25 $h^{-1}$, as previously described in the literature.[11,70]

Regarding the substrate uptake values, the highest values reached with each strain, are obtained at the same conditions that the higher values of growth rate, with 77 g/L of glucose with *M. antarcticus* reaching 2.69 mmol/gDCW/h and with 50 g/L of glucose with *U. maydis* having a substrate-uptake value of 2.20 mmol/gDCW/h. The two lower values of uptake of glucose were accomplished with the higher substrate value given *to M. antarcticus* and with the higher substrate value given to *U. maydis*. As so, just 0.37 mmol/gDCW/h was registered given 120 g/L of glucose to *M. antarcticus*, and only 0.33 mmol/gDCW/h was calculated with *U. maydis* strain with 200 g/L of initial glucose.

#### 5.2.2.2. Fed-Batch cultures at different glucose concentrations

Afterward, a fed-batch experiment was performed, with 40 g/L of initial glucose in 5 different conditions, and with more glucose added on day 4, performing a final concentration of 40 g/L (40 g/L+0 g/L), 60 g/L (40 g/L+20 g/L), 80 g/L (40 g/L+40 g/L), 120 g/L (40 g/L+80 g/L), and 200 g/L (40 g/L+160 g/L). The biomass values shown in Figure 34, represent the growth after the addition of glucose on day 4, and the correspondent OD values (Figure A5) were taken into account to calculate the growth rate values for each condition.

From day 4 (96 hours) until day 10 (240 hours), the maximum biomass value achieved for all the concentrations of glucose, except for the higher concentration of 40 g/L+160g/L, was obtained on day 7. At this point, 176 hours after the begging of the fermentation, a value of 11.1,13.1, 13.1 and 12.9 was achieved with concentration values of 40 g/L, 60 g/L, 80 g/L and 120 g/L of total glucose,

respectively. After day 7, the values of biomass of these four conditions started to decline, with a clear decrease of the lowest concentration, which did not suffer any addition of glucose on day 4. The values of 60 g/L and 80 g/L presented an almost equal representation, which was according to the excepted since the two concentrations' values are almost equal.

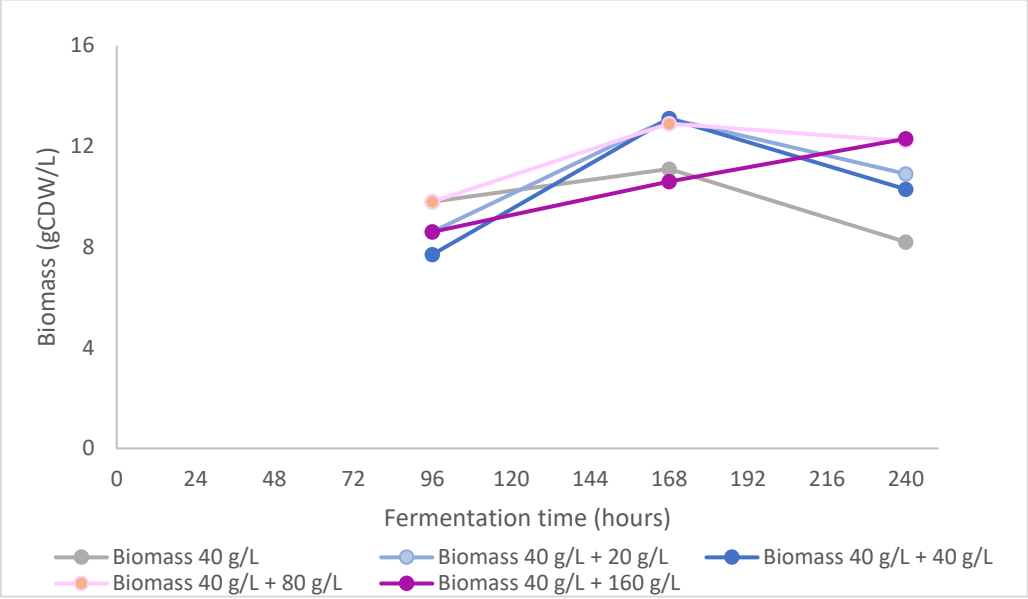



**Figure 34** - Values of biomass (g/L) through time (hours) with different glucose concentrations added from day 4 (96 hours). 40 g/L (grey); 40 g/L + 20 g/L (light blue); 40 g/L + 40 g/L (dark blue); 40 g/L + 80 g/L (light pink); 40 g/L + 160 g/L (dark pink).

At the highest concentration used, 200 g/L of total glucose, a maximum of 12.3 g/L of biomass was determined at day 10. As shown in Table 7, and comparing with the experiment later reported, without the addition of more glucose at day 4, the experiment with these additions benefited from more biomass production, with values from 11 to 13 g/L of biomass, while, when comparing with the same range of concentrations of the last experiment, values of just 7 to 8 g/L of biomass were obtained.

**Table 7** – Comparison between batch experiment with glucose as carbon source with 5 g/L, 20 g/L, 40 g/L, 80 g/L, 120 g/L of initial concentration and with the fed-batch experiment with 40 g/L of initial glucose and 20 g/L, 40 g/L, 80 g/L and 160 g/L of glucose added at day 4. Both experiments were realized with the *M. antarcticus* strain.

| Carbon Source | Concentration (g $L^{-1}$) | Sugar consumption (g/L.h) | Maximum biomass production (g DCW/L) |
|---|---|---|---|
| **Glucose** | 5 | 0.16 | 1.5 (Day 3) |
| | 20 | 0.34 | 4.3 (Day 3) |
| | 40 | 0.39 | 8.2 (Day 5) |
| | 80 | 0.34 | 7.4 (Day 5) |
| | 120 | 0.31 | 7.4 (Day 7) |
| **Glucose added on day 4** | 40 | 0.30 | 11.1 (Day 7) |
| | 40 + 20 (60) | 0.21 | 13.1 (Day 7) |
| | 40 + 40 (80) | 0.09 | 13.1 (Day 7) |
| | 40 + 80 (120) | 0.34 | 12.9 (Day 7) |
| | 40 + 160 (200) | 0.33 | 12.3 (Day 10) |

Sugar consumption values showed similar values to the ones obtained without feeding glucose on day 4, with values between 0.20 g/L.h and 0.34 g/L.h, while the values previously registered were between 0.16 g/L.h and 0.39 g/L.h. To note that, when 40 g/L of glucose was added on day 4 to 40 g/L of glucose, with a total of 80 g/L of glucose, a value of 0.09 g/L.h was obtained.

In this experiment, the MEL and lipids quantification was also accessed, possible by GC analysis (Chapter 4.1.5). The maximum values of MEL and lipids concentration were obtained at day 7 with 40 g/L + 40 g/L of glucose, with 8.4 g/L of MEL production and 10.2 g/L of lipids production (Figure A6).

Similar profiles of MEL production were achieved in three conditions, 40 g/L, 40 g/L + 20 g/L and 40 g/L + 80g/L, with the highest value of MEL production on day 7 with 40 g/L of glucose of 3.98 g/L of MEL, with 40 g/L +80 g/L of 3.24 g/L of MEL, and on day 10 with 40 g/L + 80 g/L of glucose of 4.42 g/L of MEL. With the other two concentrations of glucose, the addition at day 4 of 40 g/L and of 160 g/L to the initial 40 g/L, the profile of MEL production presented a growing production, with the highest values of lipids being reached in both conditions at day 7. The other three conditions of different glucose concentrations showed an equal pattern, with constant values of lipids production between day 2 to day 10.

These values are not expected since it is common to have a growing increase of lipids and MEL concentrations through time. For this reason, to have accurate modelling, and derived from not having duplicate values from this experiment, values obtained in the same conditions were taken from literature to perform the *in silico* predictions.

## 5.3. Fitting of experimental results with model

With the values obtained experimentally with the *M. antarcticus* strain described in chapter 5.2. a fitting of the values to the iUma22 model was necessary. With this fitting, the parameters that differ between the *Ustilago Maydis* strain and *Moesziomyces antarcticus* strain should be adjusted, turning the model presented in one closer to the *Moesziomyces* species.

### 5.3.1. *Moesziomyces antarcticus* fitting model for growth

The first comparison, already made on the values obtained to biomass growth, with glucose only added at day 0, was previously described in Chapter 5.1.2. – Table 3. In this chapter, the values of substrate uptake rate (mmol/gDCW/h) and growth rate ($h^{-1}$) are highlighted in Table 8. Below, in figure 35, the experimental values obtained with *M. antarcticus* strain and the values with *U. maydis* retrieved from literature are compared. The two strains exhibited a good value of $R^2$, even if a lower value was obtained with *M. antarcticus* strain, as shown in the linear equations.

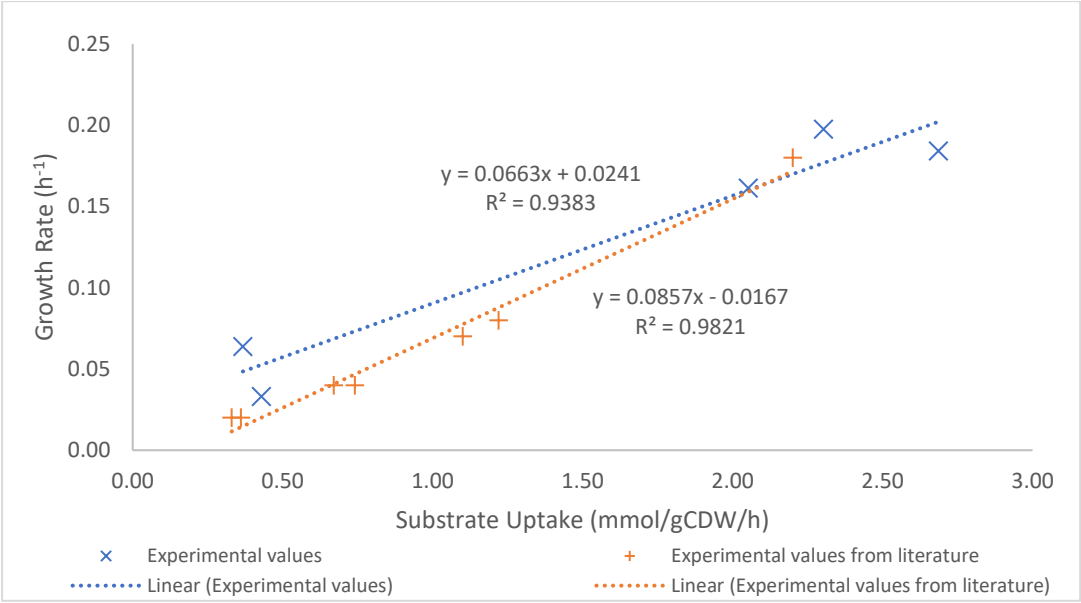



**Figure 35** - Comparision of values of experimental rates. Growth rate ($h^{-1}$) and substrate uptake rate (mmol/gDCW/L) values were obtained experimentally with *U. maydis* (orange) and *M. antarcticus* (blue). *U. maydis* values were retrieved from the literature [60] and the linear regression line equation is y=0.09x-0.017 with an $R^2$=0.98, while *M. antarcticus* values presented an equation equal to y=0.07x+0.024 with an $R^2$=0.94.

After the experimental comparison with data obtained for *U. maydis* strain, it was necessary to obtain the predicted growth values from MATLAB with the iUma22 model to the glucose flux values obtained experimentally with the *M. antarcticus* strain. As it is shown in Table 8, the difference between experimental values to predicted values was in the range between 0.00 and 0.05, while this difference was only between 0.00 and 0.01 with values obtained in *U. maydis* (Table 3).

**Table 8** - Applying experimental values obtained in batch cultures with different initial glucose concentrations and comparing the solution predicted by MATLAB.

| Experiment ID | Initial Concentration (g/L) | Substrate-uptake (mmol/gDCW/h) | Experimental Growth-Rate ($h^{-1}$) | Predicted Growth-Rate ($h^{-1}$) | Difference | Error (%) |
|---|---|---|---|---|---|---|
| *M. antarcticus* 5 g/L | 4 | 2.05 | 0.16 | 0.16 | 0.00 | 3 |
| *M. antarcticus* 20 g/L | 15 | 2.30 | 0.20 | 0.18 | 0.02 | 11 |
| *M. antarcticus* 40 g/L | 32 | 0.43 | 0.03 | 0.02 | 0.01 | 83 |
| *M. antarcticus* 80 g/L | 77 | 2.69 | 0.18 | 0.21 | 0.03 | 12 |
| *M. antarcticus* 120 g/L | 121 | 0.37 | 0.06 | 0.01 | 0.05 | 390 |

Another approach was to use different data sets from the ones used until now, employing a different carbon source, but still with *M. antarcticus* strain, and see if the model can predict growth rate values as well as it predicted when glucose was used as a carbon source. Xylose experimental results

were taken from the literature [3,71], and knowing that the same strain was utilised in those experiments, FBA approach was again employed to predict the values of biomass growth.

The error obtained in an experiment with 40 g/L xylose added on day 0 was 170 %. As it is possible to retrieve, compared to the error values obtained with glucose with the same strain (Table 8), this value is double. Even if the error was 390% when 120 g/L of glucose was used, the other error values were between 3% and 83%, which can be compared with the values obtained for *U. maydis* (Table 3), which were between 5% and 47%. In both ranges of values, the higher concentration values, 120 g/L in *M. antarcticus* and 200 g/L in *U. maydis*, were not taken into account because the error value obtained was not realistic when compared to the other concentrations.

### 5.3.2. *Moesziomyces antarcticus* fitting model with MEL optimization

With the growth values obtained with the model to *M. antarcticus,* it was then necessary to understand if the value of fluxes for the metabolite of interest was predicted accordingly. The objective function of the model was changed from biomass reaction to MAC2 reaction, which converts hexoyl-2-mannosylerythritol into mannosylerythritol lipids type D, with the final goal of producing at least one MEL homolog.

Already knowing values from MEL production with different glucose uptake values, it was possible to convert those values from concentration units to flux units. With this conversion, and having MEL D production as the objective function of the iUma22 model, the first approach was to understand if the same values were obtained *in silico*, without having any change in the *Ustilago maydis* model. This was done by maximizing the MAC2 reaction in Escher-FBA and applying the flux values of glucose experimentally obtained to the reaction of glucose uptake, D-Glucose exchange reaction.

In Table 9 is possible to compare the flux values obtained experimentally and compare those with the predicted values given by the *in silico* approach employed. The difference between the values ranges between 0.05 and 0.10 mmol/gDCW/h, which is bigger than the values obtained from the comparison of growth values. Since most of the values were retrieved from work previously reported in the literature, it was also possible to calculate the error generated between the values predicted with xylose.

The values of error in percentage surpass 50% with values from fed-batch glucose experiments, between 56% and 82%. The highest value from the table, 82%, was obtained with the only experimental value done on this work that, as already mentioned, should not be accounted as an accurate value. With xylose fed-batch experimental values, the error presented was higher than 1000%. Escher-FBA was also applied to see what were the changes in the value of fluxes through the pathways, comparing an input constraint of 10 mmol/gDCW/h of glucose or xylose (Figure 36).

**Table 9** - Substrate uptake rate values (mmol/gDCW/h) with glucose and xylose as carbon source, with values obtained experimentally through fed-batch on this work (Glu. Ant. 40g/L + 40 g/L) and retrieved from the literature, [3] comparing the experimental values with the value predicted retrieved through MATLAB. Glu.40:Glu.80 – 40 g/L of glucose at day 0 and addition of 80 g/L of glucose at day 4; Glu.80:Glu.40 - 80 g/L of glucose at day 0 and addition of 40 g/L of glucose at day 4; Glu.80:Glu.80 - 80 g/L of glucose at day 0 and addition of 80 g/L of glucose at day 4. Xyl.40:Xyl.40 and Xyl.40:Xyl.80 - 40 g/L of xylose at day 0 and the addition of 40 g/L and 80 g/L of xylose at day 4, respectively.

| Experiment | Substrate-uptake (mmol/gDCW/h) | Experimental MEL production flux (mmol/gDCW/h) | Predicted MEL production flux (mmol/gDCW/h) | Difference | Error (%) |
|---|---|---|---|---|---|
| Glu. Ant. 40g/L + 40 g/L | 0.770 | 0.048 | 0.094 | 0.05 | 82 |
| Glu.40:Glu.80 | 1.049 | 0.045 | 0.138 | 0.09 | 67 |
| Glu.80:Glu.40 | 1.015 | 0.052 | 0.132 | 0.08 | 61 |
| Glu.80:Glu.80 | 0.800 | 0.044 | 0.099 | 0.05 | 56 |
| Xyl.40:Xyl.40 | 0.307 | 0.108 | 0.007 | 0.10 | 1440 |
| Xyl.40:Xyl.80 | 0.300 | 0.083 | 0.006 | 0.08 | 1275 |

From the visualization of these pathways, it can be concluded that most of the flux goes through the same pathways, leading to MEL D production, with activated fluxes in glycolysis, the pentose phosphate pathway, and MEL synthesis. As such, the principal difference observed is when glucose is given as input to the cell, rather than xylose, where higher fluxes of glycolysis are seen through the color difference, and with the activation of TCA, which is not active when xylose input flux is active.

Then, the modifications described in chapter 5.1.2 were applied, deleting the Ustilagic acid production reactions, and adding the reactions of oil exchange and the conversion of the metabolite “Fatty acids” to “Acetyl-CoA”. The values obtained with this approach were only possible to obtain with MATLAB, applying the experimental values to the modified model previously described, and seeing how much the flux values from MEL biosynthesis pathways change (Table A4). Again, the flux values that were registered do not show relevant modifications from the ones obtained with the original iUma22 model.

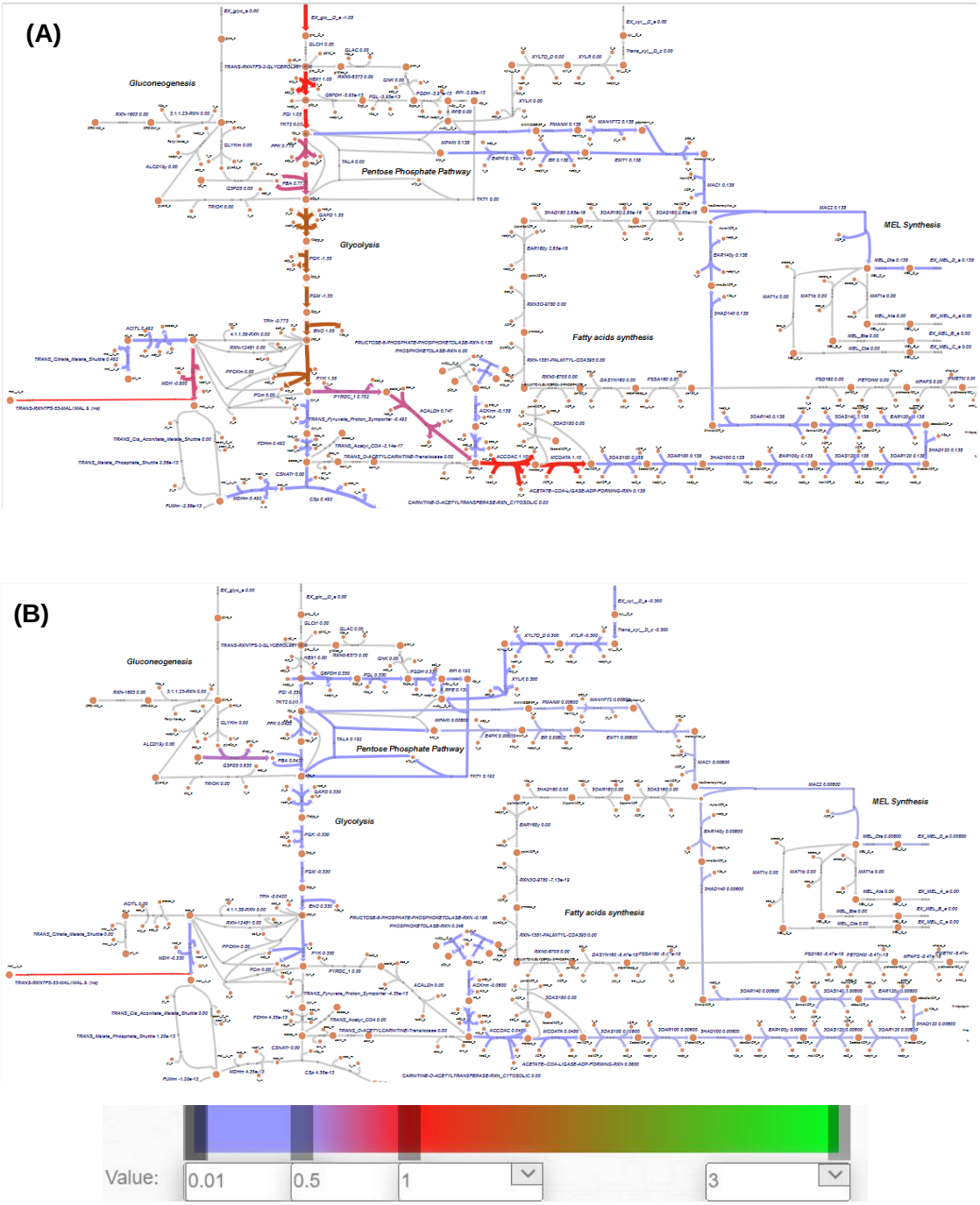



**Figure 36** - Escher-FBA fluxes visualizations with MEL production optimized, by maximizing MAC2 reaction. (A) Glucose entry, mimicking the experiment with initial glucose (day 0) of 40 g/L and addition of 80 g/L of glucose at day 4. Glucose uptake rate = 1.049 mmol/gDCW/h. (B) Xylose entry, mimicking the experiment with initial xylose (day 0) of 40 g/L and addition of 80 g/L of xylose at day 4. Xylose uptake rate = 0.300 mmol/gDCW/h. All fluxes are in units of mmol/gDCW/h and are shown by colours according to the value: 0 (grey); 0.01–0.50 (purple); 0.50–1.00 (red); 1.00–3.00 (green).

# 6. Discussion

The model iUma22 was constructed with *Ustilago maydis* as the reference organism. All the principal metabolic routes that allow the production of the metabolite of interest in this study, mannosylerythritol lipids, are present, and each metabolite is identified as the corresponding KEGG ID code. However, it is necessary to notice the absence of peroxisomes in this model, and the associated pathways that take place on the organelle, especially partial *β*-oxidation. Nevertheless, the organism of interest is not *U. maydis*, but rather *Moesziomyces* species which, cv until now, has not been described in the KEGG database.

The first approach was to perceive which of the 1855 reactions are relevant to the production of the metabolite of interest and how it is possible to visualize these reactions. By highlighting these pathways and applying the constraint of the input flux value of glucose on the cell, it was possible to obtain the predicted values of growth in this model. These values were compared with the growth rate values obtained for *U. maydis* by the authors of the model, which only showed a difference between 5% and 9% of the experimental values, obtained with glucose concentrations between 50 g/L and 130 g/L (Table 3 – Chapter 5.1.2). The values obtained for lower concentrations (20 g/L) or higher values of glucose (200 g/L) showed a difference of more than 47%. This could mean that the model has a good prediction for the majority of the inputs but has to be well curated with lower and higher values of glucose concentrations.

To obtain a better prediction of flux values when employing experimental values obtained for *Moesziomyces* organisms, some reactions were deleted from the model. The flux balance analysis approach was used to determine how much the predicted values, to the same glucose input value, changed from the original model, without modifications. The same FBA solution was obtained, and most of the flux values were maintained for the majority of reactions (Table 4 – Chapter 5.1.2.1). However, there was less need for $NH_4$, which was directly linked to the retention of glycerol in the cell, at the same time that $H_2O$ was required, with the same flux difference of the higher production of malate. Thus, these two metabolite reactions were correlated.

This analysis was performed under biomass optimization conditions. With the maximization of the MAC2 reaction, no flux value appeared to be affected by the deletion of reactions leading to the production of Ustilagic acid. Even so, comparing the flux values with the values obtained with the maximization of biomass, the flux of $O_2$ entry is 0, and the flux value for the production of $CO_2$ is different from 0. The glycerol flux value was also lower with the activation of pathways of MEL biosynthesis, which means that glycerol could be implicated in the production of MEL.

The lipid reactions were added to the original iUma22 model. In this case, only residual flux values were registered, with $Ca_2$, malate, and fumarate exchange reactions being the only exchange reactions that changed from the original value, indicating that this addition affected the values obtained *in silico*.

With the model explored and the necessary changes made to upgrade it, experiments were conducted with the objective of choosing the strain with the highest growth rate from the *Moesziomyces* species as the organism of interest to compare predicted values with experimental ones. With the knowledge gained through the first shake flask experiment, with three different sugars as a carbon substrate for the growth of *Moesziomyces* yeasts, *M. antarcticus* was the strain with a higher biomass growth rate in MEL production, reaching higher values of biomass produced and with more sugar consumption when compared with the *M. bullatus* strain (Table 5 – Chapter 5.2.1). In addition to the HPLC technique, which allowed quantification of the amount of substrate present in the yeast culture medium, qualitative techniques could also be performed.

Based on these results, *Moesziomyces antarcticus* was used in subsequent experiments. To understand how the predictions of the model differ when employing values obtained with this strain and not with *U. maydis*, fermentation with glucose as a carbon source was performed. This sugar was chosen because, in an article describing the iUma22 model [60], the growth characteristics were evaluated with different initial concentrations of glucose, and the results were compared with the predicted values of the model. To compare the experimental values obtained in this work with the values reported by the authors of the paper, it was necessary to have the most similar experimental protocol, with the only difference being the organism used.

Knowing that the two organisms have a genome with a high synteny value [12], the comparison between the growth rate values and the substrate uptake values showed that the growth of both organisms on glucose is substrate inhibited, as higher values of initial glucose concentration present lower values of both growth and substrate uptake rate. This has already been mentioned by the authors of the literature. [60] Also it is necessary to note that values obtained in both organisms demonstrated how different substrate uptake rate and growth rate values can result from similar initial substrate concentrations.

Based on these comparison results, it was expected that the predicted growth rates obtained with MATLAB to experimental growth values with *M. antarcticus* presented a slightly higher difference than with *U. maydis*, since the model needs to be adapted to the *Moesziomyces* strain. However, as shown in Figure 35 (Chapter 5.3.1), the difference between the linear regression lines was not large, therefore, the difference between the experimental and predicted growth values was also not very large.

A graph with the experimental values of growth rate and substrate uptake rate, in contrast to the values of these rates predicted *in silico*, showed that the model needs more curation to obtain a perfect fit from the experimental values. Nevertheless, as observed in Figure 37, this difference is very low and could only be due to the fact that the model used to make the predictions was not curated to be used with the organism in this work.

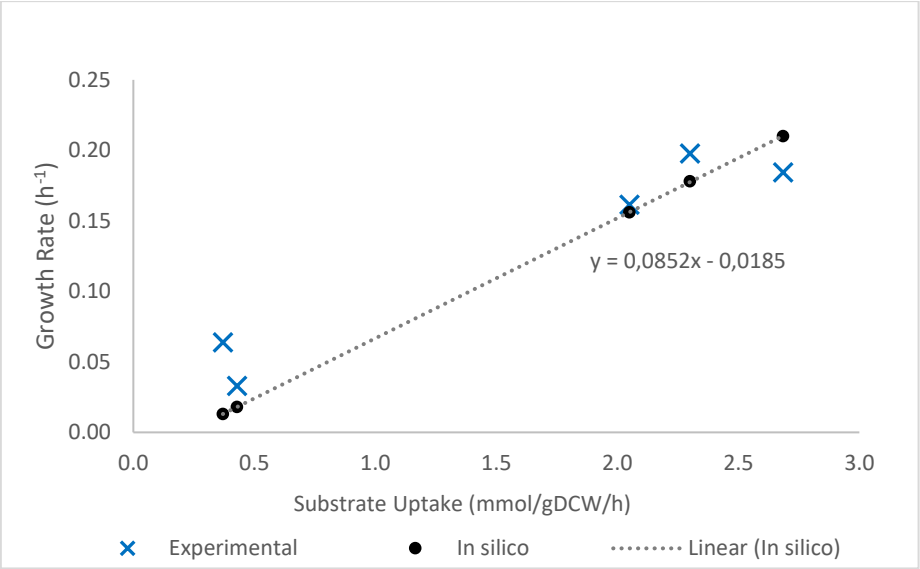



**Figure 37** - Growth rate and substrate uptake rate experimental values obtained with *M.antarcticus* strain in batch experiments with different glucose (blue crosses) in contrast to the linearity obtained to growth rate values with the same substrate uptake rate value, predicted in silico with the iUma 22 model (black dots).

Afterwards, the values from the fed-batch experiment with glucose, with equal initial concentration and different concentrations added on day 4, were also assessed to obtain the substrate uptake and biomass values. The biomass concentration values obtained were higher than those recorded in the previous experiment. The values of MEL and lipids concentration were also analysed, to compare the flux production value of MEL using *in silico* approaches. The experimental results were unexpected, since MEL and lipid concentrations did not increase over time in most of the conditions. Nevertheless, it was possible to compare these results with previously obtained results [3] under the same conditions and with the same organism. Substrate uptake values were converted to flux units and used to obtain the corresponding *in silico* MEL production flux values.

Comparing the error values between the values obtained experimentally and the values predicted from the model, when MEL production was optimised, the values of error obtained were higher than the errors with biomass optimization. This means that the predictions are more accurate when the growth of the cell is the objective, and not when MEL production is maximized. This could be due to the fact that the model was recently released and was optimized to have a fitting for growth, and not to specific metabolites.

The modified metabolic flux model, designed in MATLAB, which has the deleted equations from Ustilagic Acid production and added reactions to avoid dead-end metabolites which generate accumulation in the cell, was also implemented to predict flux values and compare them with the experimental values obtained for biomass growth and MEL production. In both conditions, the predictions were only perceived in the compounds already mentioned: $O_2$, $CO_2$ and Glycerol, when the biomass reaction was the objective, and $Ca_2$, malate, and fumarate, when MEL production was optimized, with MAC2 reaction as the objective function.

# 7. Conclusion and Future Prospects

The first possible conclusion from this work is that the present model was not constructed to predict very low or very high values of glucose concentrations, that were converted into substrate uptake flux values for the cell model. This is perceived because it is just possible to achieve some linearity in the data between 40 g/L and 130 g/L of glucose. With less than 40 g/L of glucose or more than 150 g/L of glucose, the flux input value given to the model is lower than 0.20 mmol/gDCW/h, which is the value where the model predicts that a value of that dimensions will kill the cell.

Even if the growth values and the substrate uptake values obtained with *M. antarcticus* were not that different from the ones obtained in the literature with *U. maydis* [60], not having duplicates of any experiment implies a less rigorous dataset, which could also influence the difference between the values obtained to the two organisms. This is also applied to the experiment where the MEL production values were assessed since that experiment did not present the usual patterns of production of MEL and lipids frequently obtained.[71]

The predicted values of flux for MEL production, however, presented a larger discrepancy than the values obtained experimentally when compared to the values for biomass growth in the batch experiment. This could be due to not just the fact that it is necessary to repeat the experiment and have higher insights of the iUma22 model, but also possibly because of the lack of all MEL homologs and major metabolites on the production of this biosurfactant coded on available databases. This means that even if the model predicts a value of flux in the pathways leading to its production, it is not as accurate as when the metabolites are well reported and described in databases.

Nevertheless, the original model should undergo a deeper analysis, with all the equations from the model revised to understand if there are any metabolic gaps or other reactions/metabolites that are not present in *Moesziomyces* organisms, that were not identified in this work. To allow a more accurate model, other carbon sources, such as other sugar substrates or lipidic substrates (for example fried oil) should be assessed, and the values of the growth rate and MEL production could be used to have a better fit of the upgrade model to *M. antarcticus*. To have a higher level of confidence in that values, other analytical approaches could be used.

For structural qualification, Nuclear Magnetic Resonance (NMR) experiments make it possible to understand the structural composition of the sample under analysis. This technique can enable the establishment of the flux distribution by $^{13}$C-NMR experiments, which give the carbon flow from hydrophilic (hexoses, like glucose or pentoses) and hydrophobic (free fatty acids, different types of oils) labeled $^{13}$C substrates, that lead to the production of the MEL building blocks. Furthermore, because the reconstructed networks are carbon mapped, they may be used in $^{13}$C flux research, which will be valuable to understand which ratio of carbon follows which route and if the model is predicting accordingly with that flow. Another approach, easier to perform and that can give, for example, information related to the type of MELs present in each sample is TLC. In this case, TLC will allow quick detection of the composition of the samples and whether they contain MEL or not.

By combining genomics, RNA-seq, and NMR data from producing and non-producing situations, it may be possible to better understand not only the MEL-cluster behavior but also the key genes involved in its regulation, showing inflection points in terms of gene expression and yield. [23] As so, having studies on *M. antarcticus* genome, as well as performing transcriptomic and metabolomics studies, should also be executed. Data collected from all of that approaches should also be considered to achieve a more accurate metabolic model of *Moesziomyces*.

The fitting of the experimental data to the metabolic network model should also have followed another approach that allowed calibration. The three different models obtained for *Moesziomyces* species, the one with deleted reactions, the one with added reactions, and a third with both modifications, needed to use the Mean Relative Error (MRE) scores measurement. With this and the auto-scaled transformation of estimated datasets, the model version with the best fit to the experimental data could be identified *in silico*. The MRE measurement, described in the literature, states that the model calibration requires comparing experimental data with the model's estimated values to continuously improve the model's structure until there are minimal discrepancies between the datasets. Comparing experimental growth rates observed with various organic carbon sources with the values of projected growth rates is a general step in this calibration. Because the values in the dataset can differ on orders of magnitude and dimensions, $\log_{10}$ or auto-scale transformation can also be used. [72]

In conclusion, with the *in silico* model working properly and having the structural analysis obtained by $^{13}$C-NMR experiments, current bottlenecks in MEL production using *Moesziomyces* strains could be identified, and the final objective is to reach logical strategies for the production of this biosurfactant. These strategies have the goal to increase substrate carbon use efficiency, aiming to bring biosurfactant economic costs to be competitive with fossil oil-driven surfactants. The recognition of the difficulties in MEL production, with acknowledgment of the pathways that do not contribute to that production and where the flux values are higher, should conduce to a solution for the economy of the production of this biosurfactant.

Bioinformatics is thus the answer to many of the laboratory's arduous and time-consuming experiments. Machine learning and data science are growing methods in the bioprocesses industry, and utilising well-curated cell models could help to predict accurate experimental values without the necessity of dispendious work. Models can mimic the conditions experienced by the cell, having an equal profile as the experimental one, likewise leading to data that need less processing, making it easier to analyse and to find solutions more readily.

# 8. References


1. Heirendt, L.; Arreckx, S.; Pfau, T.; Mendoza, S.N.; Richelle, A.; Heinken, A.; Haraldsdóttir, H.S.; Wachowiak, J.; Keating, S.M.; Vlasov, V.; et al. Creation and Analysis of Biochemical Constraint-Based Models Using the COBRA Toolbox v.3.0. *Nature Protocols 2019 14:3* **2019**, *14*, 639–702, doi:10.1038/s41596-018-0098-2.

2. Orth, J.D.; Fleming, R.M.T.; Palsson, B.Ø. Reconstruction and Use of Microbial Metabolic Networks: The Core Escherichia Coli Metabolic Model as an Educational Guide. *EcoSal Plus* **2010**, *4*, doi:10.1128/ECOSALPLUS.10.2.1/ASSET/AF2F042E-12AF-48A1-967D-9627A8D81014/ASSETS/GRAPHIC/10.2.1_FIG_022.GIF.

3. Maria Esteves Silva, M.; Castelo Alves Ferreira Nuno Ricardo Torres Faria, F.; Ricardo Torres Faria, N. Mannosylerythritol Lipids: Searching for Production and Downstream Routes, Universidade de Lisboa - Instituto Superior Técnico, 2017.

4. Kitamoto, D.; Isoda, H.; Nakahara, T. Functions and Potential Applications of Glycolipid Biosurfactants — from Energy-Saving Materials to Gene Delivery Carriers —. *J Biosci Bioeng* **2002**, *94*, 187–201, doi:10.1016/S1389-1723(02)80149-9.

5. Twigg, M.S.; Baccile, N.; Banat, I.M.; Déziel, E.; Marchant, R.; Roelants, S.; van Bogaert, I.N.A. Microbial Biosurfactant Research: Time to Improve the Rigour in the Reporting of Synthesis, Functional Characterization and Process Development. *Microb Biotechnol* **2021**, *14*, 147–170, doi:10.1111/1751-7915.13704.

6. Banat, I.M.; Joshi, S.J. Biosurfactants: Production and Potential Applications in Microbial Enhanced Oil Recovery (MEOR). *Biocatal Agric Biotechnol* **2018**, *14*, 23–32, doi:10.1016/j.bcab.2018.01.010.

7. Saika, A.; Koike, H.; Yamamoto, S.; Kishimoto, T.; Morita, T. Enhanced Production of a Diastereomer Type of Mannosylerythritol Lipid-B by the Basidiomycetous Yeast Pseudozyma Tsukubaensis Expressing Lipase Genes from Pseudozyma Antarctica. *Appl Microbiol Biotechnol* **2017**, *101*, 8345–8352, doi:10.1007/S00253-017-8589-6/FIGURES/6.

8. Freitag, J.; Ast, J.; Linne, U.; Stehlik, T.; Martorana, D.; Bölker, M.; Sandrock, B. Peroxisomes Contribute to Biosynthesis of Extracellular Glycolipids in Fungi. *Mol Microbiol* **2014**, *93*, 24–36, doi:10.1111/MMI.12642/FORMAT/PDF.

9. Abdel-Mawgoud, A.M.; Stephanopoulos, G. Simple Glycolipids of Microbes: Chemistry, Biological Activity and Metabolic Engineering. *Synth Syst Biotechnol* **2018**, *3*, 3–19, doi:10.1016/J.SYNBIO.2017.12.001.

10. Coelho, A.L.S.; Feuser, P.E.; Carciofi, B.A.M.; de Oliveira, D.; de Andrade, C.J. Biological Activity of Mannosylerythritol Lipids on the Mammalian Cells. *Applied Microbiology and Biotechnology 2020 104:20* **2020**, *104*, 8595–8605, doi:10.1007/S00253-020-10857-9.

11. Beck, A.; Zibek, S. Growth Behavior of Selected Ustilaginaceae Fungi Used for Mannosylerythritol Lipid (MEL) Biosurfactant Production – Evaluation of a Defined Culture Medium. *Front Bioeng Biotechnol* **2020**, *8*, doi:10.3389/fbioe.2020.555280.

12. Morita, T.; Koike, H.; Hagiwara, H.; Ito, E.; Machida, M.; Sato, S.; Habe, H.; Kitamoto, D. Genome and Transcriptome Analysis of the Basidiomycetous Yeast Pseudozyma Antarctica Producing Extracellular Glycolipids, Mannosylerythritol Lipids. *PLoS One* **2014**, *9*, doi:10.1371/journal.pone.0086490.

13. Shu, Q.; Wei, T.; Lu, H.; Niu, Y.; Chen, Q. Mannosylerythritol Lipids: Dual Inhibitory Modes against Staphylococcus Aureus through Membrane-Mediated Apoptosis and Biofilm Disruption. *Appl Microbiol Biotechnol* **2020**, *104*, 5053–5064, doi:10.1007/S00253-020-10561-8/FIGURES/6.

14. Beck, A.; Haitz, F.; Thier, I.; Siems, K.; Jakupovic, S.; Rupp, S.; Zibek, S. Novel Mannosylerythritol Lipid Biosurfactant Structures from Castor Oil Revealed by Advanced Structure Analysis. *J Ind Microbiol Biotechnol* **2021**, *48*, doi:10.1093/jimb/kuab042.

15. Niu, Y.; Wu, J.; Wang, W.; Chen, Q. Production and Characterization of a New Glycolipid, Mannosylerythritol Lipid, from Waste Cooking Oil Biotransformation by Pseudozyma Aphidis ZJUDM34. *Food Sci Nutr* **2019**, *7*, 937–948, doi:10.1002/FSN3.880.

16. Beck, A.; Werner, N.; Zibek, S. Mannosylerythritol Lipids: Biosynthesis, Genetics, and Production Strategies. *Biobased Surfactants* **2019**, 121–167, doi:10.1016/B978-0-12-812705-6.00004-6.

17. Fan, L.; Li, H.; Niu, Y.; Chen, Q. Characterization and Inducing Melanoma Cell Apoptosis Activity of Mannosylerythritol Lipids-A Produced from Pseudozyma Aphidis. *PLoS One* **2016**, *11*, doi:10.1371/journal.pone.0148198.

18. Paulino, B.N.; Pessôa, M.G.; Molina, G.; Kaupert Neto, A.A.; Oliveira, J.V.C.; Mano, M.C.R.; Pastore, G.M. Biotechnological Production of Value-Added Compounds by Ustilaginomycetous Yeasts. *Appl Microbiol Biotechnol* **2017**, *101*, 7789–7809, doi:10.1007/S00253-017-8516-X.

19. Faria, N.T.; Santos, M.; Ferreira, C.; Marques, S.; Ferreira, F.C.; Fonseca, C. Conversion of Cellulosic Materials into Glycolipid Biosurfactants, Mannosylerythritol Lipids, by Pseudozyma Spp. under SHF and SSF Processes. *Microb Cell Fact* **2014**, *13*, doi:10.1186/S12934-014-0155-7.

20. Morita, T.; Fukuoka, T.; Imura, T.; Kitamoto, D. Mannosylerythritol Lipids: Production and Applications. *J. Oleo Sci* **2015**, *64*, 133–141, doi:10.5650/jos.ess14185.

21. Saavedra, E.; Ramos-Casillas, L.E.; Marín-Hernández, A.; Moreno-Sánchez, R.; Guerra-Sánchez, G. Glycolysis in Ustilago Maydis. *FEMS Yeast Res* **2008**, *8*, 1313–1323, doi:10.1111/J.1567-1364.2008.00437.X.

22. Morita, T.; Ogura, Y.; Takashima, M.; Hirose, N.; Fukuoka, T.; Imura, T.; Kondo, Y.; Kitamoto, D. Isolation of Pseudozyma Churashimaensis Sp. Nov., a Novel Ustilaginomycetous Yeast Species as a Producer of Glycolipid Biosurfactants, Mannosylerythritol Lipids. *J Biosci Bioeng* **2011**, *112*, 137–144, doi:10.1016/J.JBIOSC.2011.04.008.

23. Solano-Gonzá lez, S.; Solano-Campos, F.; Solano-González, S. Production of Mannosylerythritol Lipids: Biosynthesis, Multi-Omics Approaches, and Commercial Exploitation. *Mol Omics* **2022**, doi:10.1039/D2MO00150K.

24. Morita, T.; Koike, H.; Koyama, Y.; Hagiwara, H.; Ito, E.; Fukuoka, T.; Imura, T.; Machida, M.; Kitamoto, D. Genome Sequence of the Basidiomycetous Yeast Pseudozyma Antarctica T-34, a

Producer of the Glycolipid Biosurfactants Mannosylerythritol Lipids. **2013**, doi:10.1128/genomeA.00064-13.

25. Wada, K.; Koike, H.; Fujii, T.; Morita, T. Targeted Transcriptomic Study of the Implication of Central Metabolic Pathways in Mannosylerythritol Lipids Biosynthesis in Pseudozyma Antarctica T-34. *PLoS One* **2020**, *15*, doi:10.1371/journal.pone.0227295.

26. Liu, Y.; Zou, Z.; Hu, Z.; Wang, W.; Xiong, J. Morphology and Molecular Analysis of Moesziomyces Antarcticus Isolated from the Blood Samples of a Chinese Patient. *Front Microbiol* **2019**, *10*, doi:10.3389/FMICB.2019.00254/FULL.

27. Tanaka Motoo Koitabashi Hiroko Kitamoto, E. A Teleomorph of the Ustilaginalean Yeast Moesziomyces Antarcticus on Barnyardgrass in Japan Provides Bioresources That Degrade Biodegradable Plastics. *Antonie Van Leeuwenhoek* **2019**, doi:10.1007/s10482-018-1190-x.

28. Fan, L.L.; Dong, Y.C.; Fan, Y.F.; Zhang, J.; Chen, Q.H. Production and Identification of Mannosylerythritol Lipid-A Homologs from the Ustilaginomycetous Yeast Pseudozyma Aphidis ZJUDM34. *Carbohydr Res* **2014**, *392*, 1–6, doi:10.1016/J.CARRES.2014.04.013.

29. Lorenz, S.; Guenther, M.; Grumaz, C.; Rupp, S.; Zibek, S.; Sohn, K. Genome Sequence of the Basidiomycetous Fungus Pseudozyma Aphidis DSM70725, an Efficient Producer of Biosurfactant Mannosylerythritol Lipids. *Genome Announc* **2014**, *2*, doi:10.1128/GENOMEA.00053-14.

30. Akkermans, V.; Verstraete, R.; Braem, C.; D'Aes, J.; Dries, J. Mannosylerythritol Lipid Production from Oleaginous Yeast Cell Lysate by Moesziomyces Aphidis. *Industrial Biotechnology* **2020**, *16*, 222–232, doi:10.1089/ind.2019.0040.

31. Hewald, S.; Josephs, K.; Bölker, M. Genetic Analysis of Biosurfactant Production in Ustilago Maydis. *Appl Environ Microbiol* **2005**, *71*, 3033, doi:10.1128/AEM.71.6.3033-3040.2005.

32. Hewald, S.; Linne, U.; Scherer, M.; Marahiel, M.A.; Kämper, J.; Bölker, M. Identification of a Gene Cluster for Biosynthesis of Mannosylerythritol Lipids in the Basidiomycetous Fungus Ustilago Maydis. *Appl Environ Microbiol* **2006**, *72*, 5469–5477, doi:10.1128/AEM.00506-06.

33. Walker, V. Neonatal and Paediatric Biochemistry. *Scientific Foundations of Biochemistry in Clinical Practice* **1994**, 292–315, doi:10.1016/B978-0-7506-0167-2.50022-4.

34. Ullmann, L.; Phan, A.N.T.; Kaplan, D.K.P.; Blank, L.M.; Phan, L.; Kaplan, A.N.T.; Blank, D.K.P.;; Perlin, M.H. Ustilaginaceae Biocatalyst for Co-Metabolism of CO 2-Derived Substrates toward Carbon-Neutral Itaconate Production Citation: Ullmann. **2021**, doi:10.3390/jof.

35. Jerry Kaneko, J. Carbohydrate Metabolism and Its Diseases. *Clinical Biochemistry of Domestic Animals* **2008**, 45–80, doi:10.1016/B978-0-12-370491-7.00003-9.

36. Donzella, S.; Cucchetti, D.; Capusoni, C.; Rizzi, A.; Galafassi, S.; Chiara, G.; Compagno, C. Engineering Cytoplasmic Acetyl-CoA Synthesis Decouples Lipid Production from Nitrogen Starvation in the Oleaginous Yeast Rhodosporidium Azoricum. *Microb Cell Fact* **2019**, *18*, 199–199, doi:10.1186/S12934-019-1250-6.

37. Kim, Y.; Kim, E.Y.; Seo, Y.M.; Yoon, T.K.; Lee, W.S.; Lee, K.A. Function of the Pentose Phosphate Pathway and Its Key Enzyme, Transketolase, in the Regulation of the Meiotic Cell Cycle in Oocytes. *Clin Exp Reprod Med* **2012**, *39*, 58–67, doi:10.5653/CERM.2012.39.2.58.

38. de Mendoza, D.; Schujman, G.E. Lipid Biosynthesis. *Encyclopedia of Microbiology* **2009**, 219–228, doi:10.1016/B978-012373944-5.00077-8.

39. Kádár, Z.; Fonseca, C. Bio-Products from Sugar-Based Fermentation Processes. In *Biorefinery*; Springer International Publishing, 2019; pp. 281–312.

40. Kitamoto, ı; Yanagishita, H.; Haraya, K.; Kitamoto, H.K. *Contribution of a Chain-Shortening Pathway to the Biosynthesis of the Fatty Acids of Mannosylerythritol Lipid (Biosurfactant) in the Yeast Candida Antarctica: Effect of-Oxidation Inhibitors on Biosurfactant Synthesis*; 1998;

41. Yu, M.; Liu, Z.; Zeng, G.; Zhong, H.; Liu, Y.; Jiang, Y.; Li, M.; He, X.; He, Y. Characteristics of Mannosylerythritol Lipids and Their Environmental Potential. *Carbohydr Res* 2015, *407*, 63–72.

42. Wanders, R.J.A.; Waterham, H.R.; Ferdinandusse, S. Metabolic Interplay between Peroxisomes and Other Subcellular Organelles Including Mitochondria and the Endoplasmic Reticulum. *Front Cell Dev Biol* **2016**, *0*, 83, doi:10.3389/FCELL.2015.00083.

43. Jezierska, S.; Claus, S.; van Bogaert, I. Yeast Glycolipid Biosurfactants. *FEBS Lett* **2018**, *592*, 1312–1329, doi:10.1002/1873-3468.12888.

44. Assal, D.C. el; Fleming, R.M.T.; Richelle, A.; Heinken, A. *Adding Biological Constraints to a Flux Balance Model Reviewers*;

45. Raia, V.; Schilling, M.; Böhm, M.; Hahn, B.; Kowarsch, A.; Raue, A.; Sticht, C.; Bohl, S.; Saile, M.; Möller, P.; et al. Dynamic Mathematical Modeling of IL13-Induced Signaling in Hodgkin and Primary Mediastinal B-Cell Lymphoma Allows Prediction of Therapeutic Targets. *Cancer Res* **2010**, *71*, 693–704, doi:10.1158/0008-5472.CAN-10-2987.

46. Becker, S.A.; Palsson, B.O. Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Comput Biol* **2008**, *4*, e1000082, doi:10.1371/JOURNAL.PCBI.1000082.

47. Thiele, I.; Palsson, B. A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction. *Nature Protocols 2010 5:1* **2010**, *5*, 93–121, doi:10.1038/nprot.2009.203.

48. Song, H.S.; Cannon, W.R.; Beliaev, A.S.; Konopka, A. Mathematical Modeling of Microbial Community Dynamics: A Methodological Review. *Processes 2014, Vol. 2, Pages 711-752* **2014**, *2*, 711–752, doi:10.3390/PR2040711.

49. Kauffman, K.J.; Prakash, P.; Edwards, J.S. Advances in Flux Balance Analysis. *Curr Opin Biotechnol* **2003**, *14*, 491–496, doi:10.1016/J.COPBIO.2003.08.001.

50. Bordbar, A.; Monk, J.M.; King, Z.A.; Palsson, B.O. Constraint-Based Models Predict Metabolic and Associated Cellular Functions. *Nat Rev Genet* **2014**, *15*, 107–120, doi:10.1038/nrg3643.

51. Becker, S.A.; Feist, A.M.; Mo, M.L.; Hannum, G.; Palsson, B.; Herrgard, M.J. Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox. *Nat Protoc* **2007**, *2*, 727–738, doi:10.1038/NPROT.2007.99.

52. Schmidt, H.; Jirstrand, M. Systems Biology Toolbox for MATLAB: A Computational Platform for Research in Systems Biology. *Bioinformatics* **2006**, *22*, 514–515, doi:10.1093/BIOINFORMATICS/BTI799.

53. Gombert, A.K.; Nielsen, J. Mathematical Modelling of Metabolism. *Curr Opin Biotechnol* **2000**, *11*, 180–186, doi:10.1016/S0958-1669(00)00079-3.

54. King, Z.A.; Dräger, A.; Ebrahim, A.; Sonnenschein, N.; Lewis, N.E.; Palsson, B.O. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Comput Biol* **2015**, *11*, 1004321, doi:10.1371/journal.pcbi.1004321.

55. Rowe, E.; Palsson, B.O.; King, Z.A. Escher-FBA: A Web Application for Interactive Flux Balance Analysis. *BMC Syst Biol* **2018**, *12*, 84, doi:10.1186/S12918-018-0607-5/FIGURES/3.

56. Pitkänen, E.; Jouhten, P.; Hou, J.; Syed, M.F.; Blomberg, P. Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. *PLoS Comput Biol* **2014**, *10*, 1003465, doi:10.1371/journal.pcbi.1003465.

57. Malik-Sheriff, R.S.; Glont, M.; Nguyen, T.V.N.; Tiwari, K.; Roberts, M.G.; Xavier, A.; Vu, M.T.; Men, J.; Maire, M.; Kananathan, S.; et al. BioModels-15 Years of Sharing Computational Models in Life Science. *Nucleic Acids Res* **2019**, *48*, 407–415, doi:10.1093/nar/gkz1055.

58. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **1999**, *27*, 29–34, doi:10.1093/NAR/27.1.29.

59. Kanehisa, M.; Furumichi, M.; Sato, Y.; Kawashima, M.; Ishiguro-Watanabe, M. KEGG for Taxonomy-Based Analysis of Pathways and Genomes. *Nucleic Acids Res* **2022**, doi:10.1093/NAR/GKAC963.

60. Liebal, U.W.; Ullmann, L.; Lieven, C.; Kohl, P.; Wibberg, D.; Zambanini, T.; Blank, L.M. A Genome-Scale Metabolic Model for the Smut-Fungus Ustilago Maydis. **2022**, doi:10.1101/2022.03.03.482780.

61. Faria, N.T.; Santos, M. v.; Fernandes, P.; Fonseca, L.L.; Fonseca, C.; Ferreira, F.C. Production of Glycolipid Biosurfactants, Mannosylerythritol Lipids, from Pentoses and d-Glucose/d-Xylose Mixtures by Pseudozyma Yeast Strains. *Process Biochemistry* **2014**, *49*, 1790–1799, doi:10.1016/J.PROCBIO.2014.08.004.

62. Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M. BioNumbers--the Database of Key Numbers in Molecular and Cell Biology. *Nucleic Acids Res* **2010**, *38*, doi:10.1093/NAR/GKP889.

63. Wada, K.; Saika, A.; Ushimaru, K.; Sato, S.; Fukuoka, T.; Morita, T. Metabolomic Evaluation of the Central Metabolic Pathways of Mannosylerythritol Lipid Biosynthesis in Moesziomyces Antarcticus T-34. *J Oleo Sci* **2022**, *71*, 119–125, doi:10.5650/JOS.ESS21229.

64. Sato, T.K.; Tremaine, M.; Parreiras, L.S.; Hebert, A.S.; Myers, K.S.; Higbee, A.J.; Sardi, M.; McIlwain, S.J.; Ong, I.M.; Breuer, R.J.; et al. Directed Evolution Reveals Unexpected Epistatic Interactions That Alter Metabolic Regulation and Enable Anaerobic Xylose Use by Saccharomyces Cerevisiae. *PLoS Genet* **2016**, *12*, doi:10.1371/JOURNAL.PGEN.1006372.

65. Lieven, C. Generation of a Genome-Scale Metabolic Model of Ustilago Maydis, Causative Agent of Corn Smut, RWTH Aachen University , 2015.

66. Feist, A.M.; Palsson, B.O. The Biomass Objective Function. *Curr Opin Microbiol* **2010**, *13*, 344–349, doi:10.1016/J.MIB.2010.03.003.

67. Li, C.; Donizelli, M.; Rodriguez, N.; Dharuri, H.; Endler, L.; Chelliah, V.; Li, L.; He, E.; Henry, A.; Stefan, M.I.; et al. BioModels Database: An Enhanced, Curated and Annotated Resource for Published Quantitative Kinetic Models. *BMC Syst Biol* **2010**, *4*, doi:10.1186/1752-0509-4-92.

68. Orth, J.D.; Thiele, I.; Palsson, B.O. What Is Flux Balance Analysis? *Nat Biotechnol* 2010, *28*, 245–248.

69. Teichmann, B.; Liu, L.; Schink, K.O.; Bölker, M. Activation of the Ustilagic Acid Biosynthesis Gene Cluster in Ustilago Maydis by the C2H2 Zinc Finger Transcription Factor Rua1. *Appl Environ Microbiol* **2010**, *76*, 2633, doi:10.1128/AEM.02211-09.

70. Becker, J.; Tehrani, H.H.; Ernst, P.; Blank, L.M.; Wierckx, N. An Optimized Ustilago Maydis for Itaconic Acid Production at Maximal Theoretical Yield. *Journal of Fungi 2021, Vol. 7, Page 20* **2020**, *7*, 20, doi:10.3390/JOF7010020.

71. Santos, M.; Ferreira, F.; Fonseca, C. Mannosylerythritol Lipids Bioproduction by Moesziomyces Antarcticus: Sequential Stages Approach, Universidade de Lisboa - Instituto Superior Técnico: Lisboa, 2017.

72. Perez-Garcia, O.; Villas-Boas, S.; Singhal, N. A Method to Calibrate Metabolic Network Models with Experimental Datasets. *Advances in Intelligent Systems and Computing* **2014**, *294*, 183–190, doi:10.1007/978-3-319-07581-5_22.

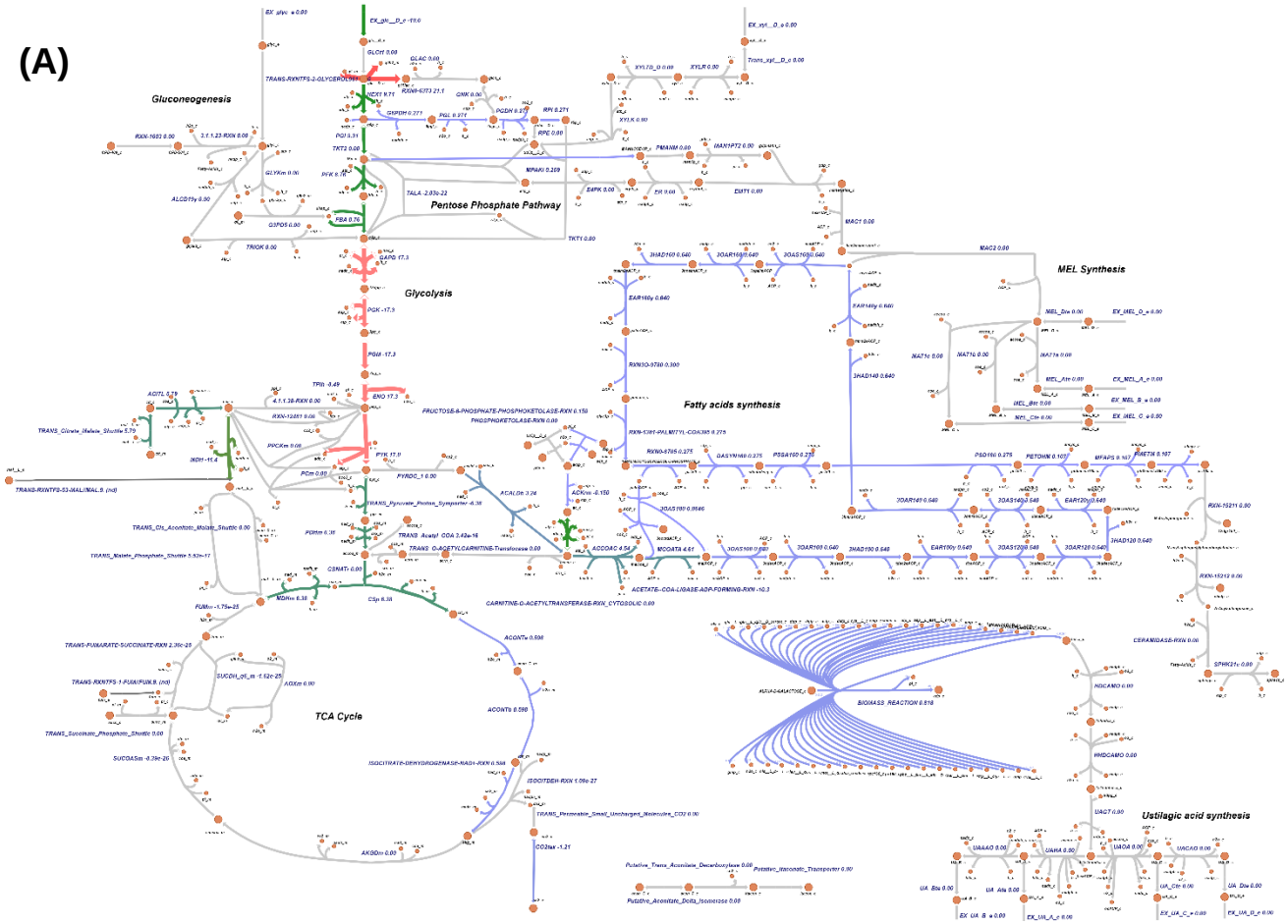
# 9. Appendix

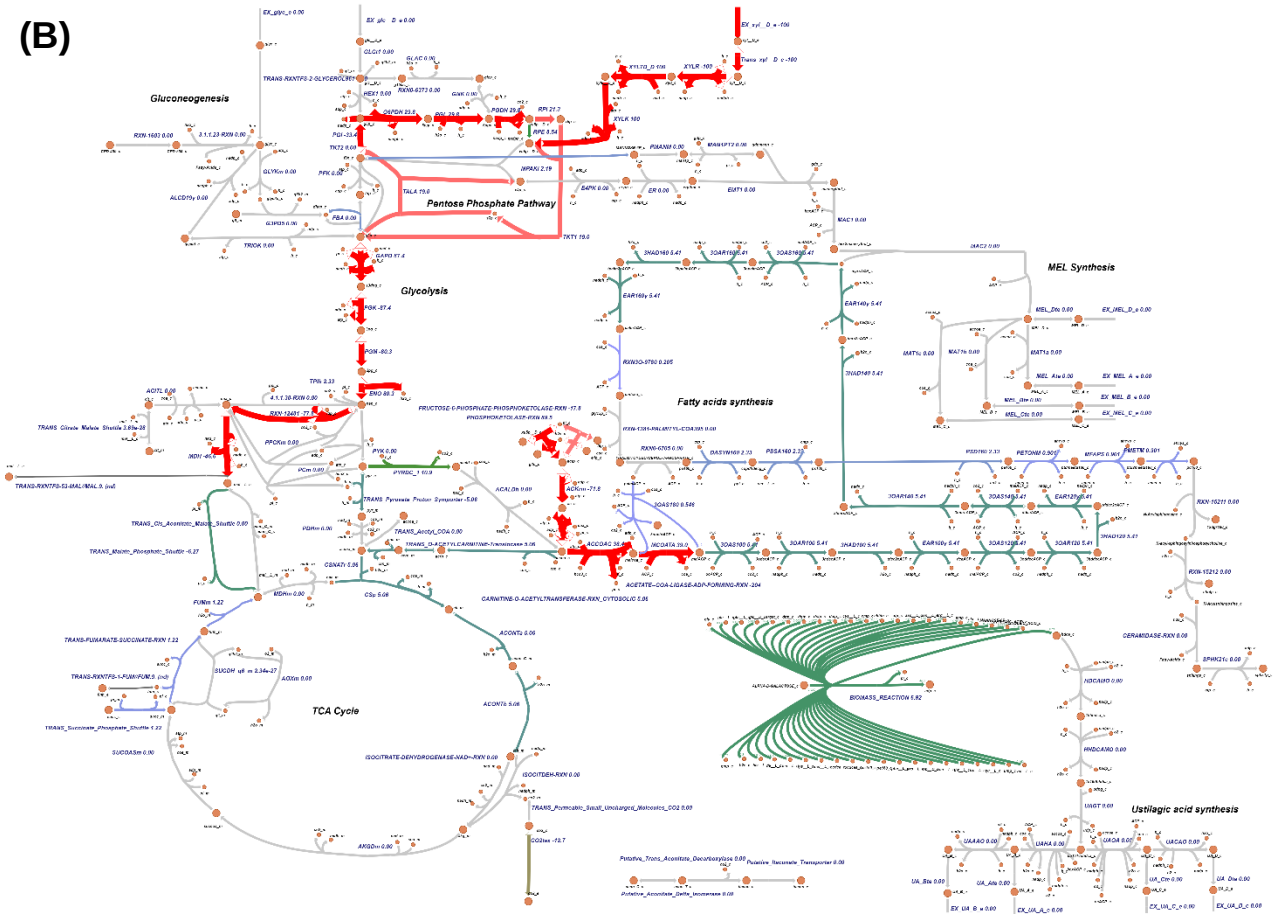
**Table A1** – KEGG ID and names of metabolites, according to each metabolic pathway, from iUma22 model.

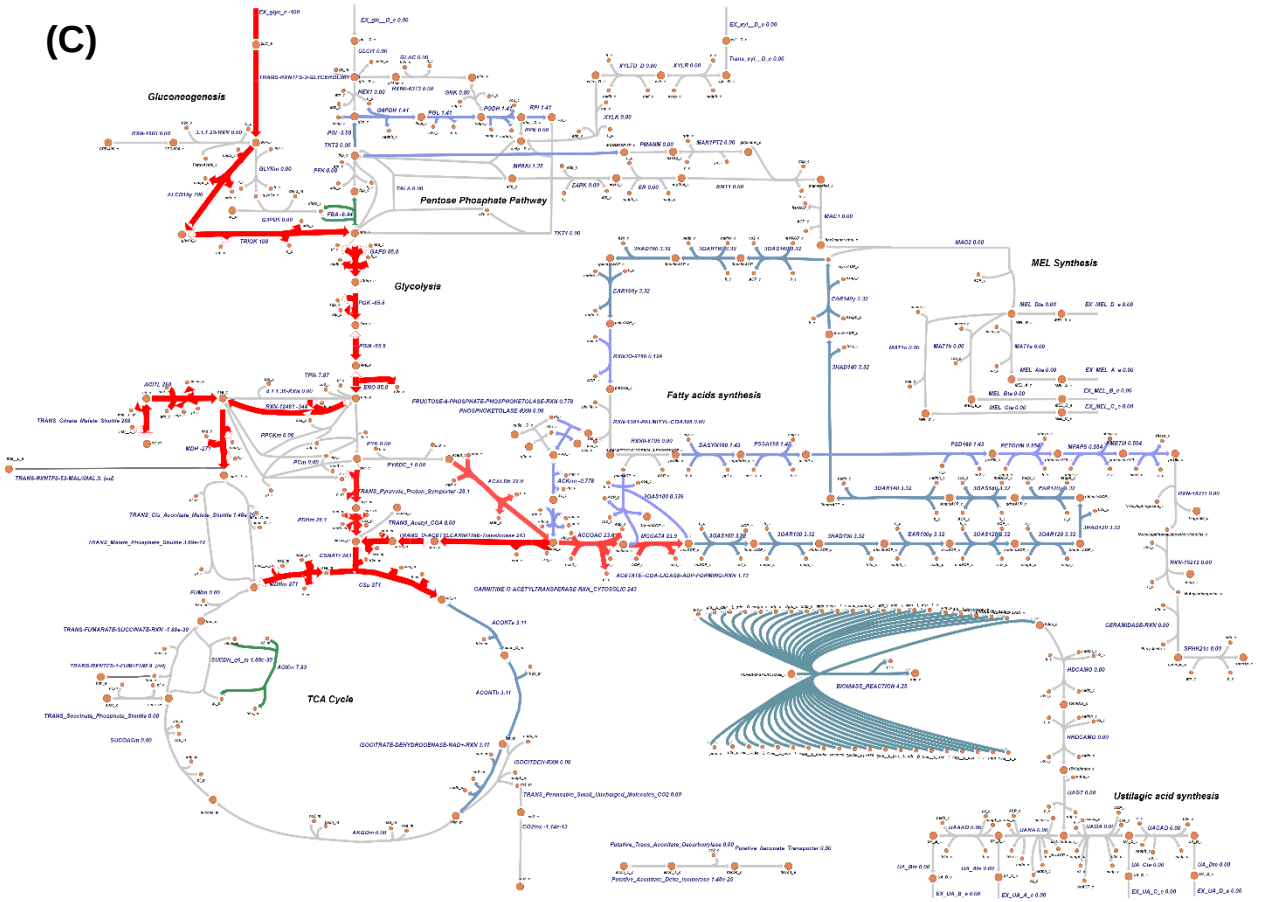
| Pathway | Abbreviation | KEGG ID [compartment] | Description |
|---|---|---|---|
| Glycolysis | pyr_c[c] | C00022[Cytoplasm] | Pyruvate |
| | glc__D_c[c] | C00031[Cytoplasm] | D-Glucose |
| | f6p_c[c] | C00085[Cytoplasm] | D-Fructose 6-phosphate |
| | g6p_c[c] | C00668[Cytoplasm] | D-Glucose 6-phosphate |
| | 3pg_c[c] | C00197[Cytoplasm] | 3-Phospho-D-glycerate |
| | fdp_c[c] | C00354[Cytoplasm] | Fructose 1,6-bisphosphate |
| Pentose Phosphate Pathway | r5p_c[c] | C00117[Cytoplasm] | D-Ribose 5-phosphate |
| | ru5p__D_c[c] | C00199[Cytoplasm] | D-Ribulose 5-phosphate |
| | xu5p__D_c[c] | C00231[Cytoplasm] | D-Xylulose 5-phosphate |
| | 6pgc_c[c] | C00345[Cytoplasm] | D-gluconate 6-phosphate |
| | 6pgl_c[c] | C01236[Cytoplasm] | 6-phospho D-glucono-1,5-lactone |
| | s7p_c[c] | C05382[Cytoplasm] | Sedoheptulose 7-phosphate |
| Gluconeogenesis | glyc3p_c[c] | C00093[Cytoplasm] | sn-Glycerol 3-phosphate |
| | gdpmann_c[c] | C00096[Cytoplasm] | GDP-alpha-D-mannose |
| | dhap_c[c] | C00111[Cytoplasm] | Dihydroxyacetone phosphate |
| | glyc_c[c] | C00116[Cytoplasm] | Glycerol |
| | g3p_c[c] | C00118[Cytoplasm] | D-Glyceraldehyde 3-phosphate |
| | MANNOSE-6P_c[c] | C00275[Cytoplasm] | alpha-D-mannose 6-phosphate |
| | man1p_c[c] | C00636[Cytoplasm] | Alpha-D-mannose 1-phosphate |
| Tricarboxylic acid cycle | pyr_m[m] | C00022[Mitochondrial Lumen] | Pyruvate |
| | akg_m[m] | C00026[Mitochondrial Lumen] | 2-Oxoglutarate |
| | oaa_m[m] | C00036[Mitochondrial Lumen] | Oxaloacetate |
| | succ_m[m] | C00042[Mitochondrial Lumen] | Succinate |
| | glx_m[m] | C00048[Mitochondrial Lumen] | Glyoxylate |
| | succoa_m[m] | C00091[Mitochondrial Lumen] | Succinyl-CoA |
| | fum_m[m] | C00122[Mitochondrial Lumen] | Fumarate |
| | mal__L_m[m] | C00149[Mitochondrial Lumen] | (S)-malate |
| | cit_m[m] | C00158[Mitochondrial Lumen] | Citrate |
| | OXALO-SUCCINATE_c[c] | C05379[Cytoplasm] | Oxalosuccinate |
| Other Metabolites | h2o_c[c] | C00001[Cytoplasm] | H2O |
| | atp_e[e]; atp_c[c] | C00002[Extracellular space]; [Cytoplasm] | ATP |
| | nadph_c[c] | C00005[Cytoplasm] | NADPH |
| | nadp_c[c] | C00006[Cytoplasm] | NADP+ |
| | o2_m[m] ; o2_e[e]; o2_c[c] | C00007[Mitochondrial Lumen];[Extracellular space]; [Cytoplasm] | Oxygen |
| | adp_c[c] | C00008[Cytoplasm] | ADP |
| | pi_c[c] | C00009[Cytoplasm] | Phosphate |

| | | | |
|---|---|---|---|
| | coa_c[c]<br>coa_m[m] | C00010[Mitochondrial Lumen]; [Cytoplasm] | Coenzyme A |
| | ppi_c[c] | C00013[Cytoplasm] | Diphosphate |
| | h_c[c] | C00080[Cytoplasm] | H+ |
| | datp_c[c] | C00131[Cytoplasm] | dATP |
| | dgtp_c[c] | C00286[Cytoplasm] | dGTP |

**(A)**




(B)
Gluconeogenesis
Pentose Phosphate Pathway
Glycolysis
Fatty acids synthesis
MEL Synthesis
TCA Cycle
Ustilagic acid synthesis
BIOMASS_REACTION 0.92


(C)
Gluconeogenesis
Pentose Phosphate Pathway
Glycolysis
Fatty acids synthesis
MEL Synthesis
TCA Cycle
Ustilagic acid synthesis
BIOMASS_REACTION 4.20

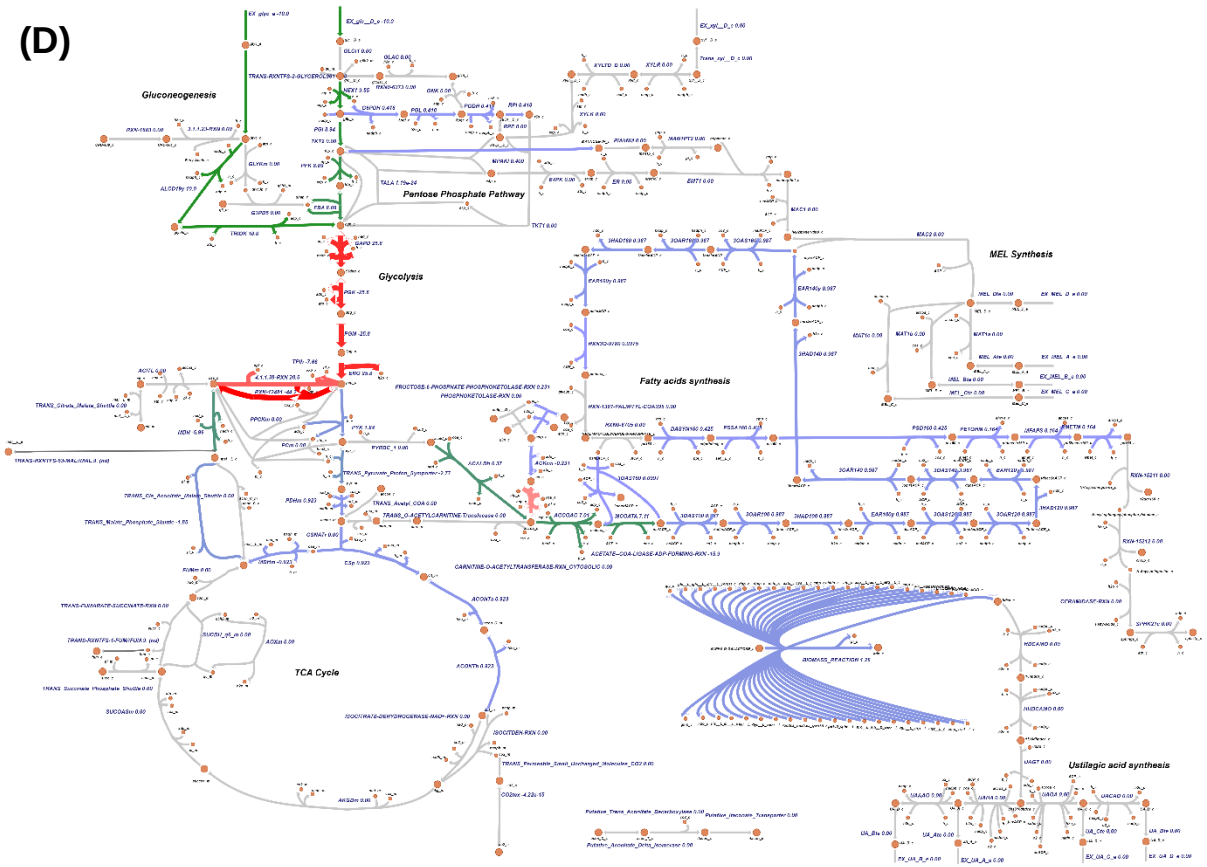



**Figure A1** – Examples of Escher-FBA simulations. (A) Simulated growth with glucose as a carbon source. Flux value of entry of glucose = 10 mmol/gDCW/h. (B) Simulated growth with xylose as a carbon source. Flux value of entry of xylose = 100 mmol/gDCW/h. (C) Simulated growth with glycerol as carbon source. Flux value of entry of glycerol = 100 mmol/gDCW/h. (D) Simulated growth with glucose and glycerol as carbon sources. Flux value of entry of glucose = 10 mmol/gDCW/h. Flux value of entry of glycerol = 10 mmol/gDCW/h. All fluxes are in units of mmol/gDCW/h and are shown by colors according to the value: 0 (grey); 0.1 – 2 (light purple); 2 – 14 (green); 14-20 (light red); Bigger than 20 (red).

**Table A2.1 -** Deleted reactions from iUma22 model, leading to the production of homologs of ustilagic acid.

| Number of the reaction | Abbreviation | Description | Reaction |
|---|---|---|---|
| **1692** | HDCAMO | Hexadecanoate Monooxigenase - CYP1 | h_c[c] + hdca_c[c] + nadph_c[c] + o2_c[c] -> nadp_c[c] + h2o_c[c] + 16hhdca_c[c] |
| **1693** | HHDCAMO | 16-Hydroxy Hexadecanoate Monooxigenase - CYP2 | h_c[c] + nadph_c[c] + o2_c[c] + 16hhdca_c[c] -> nadp_c[c] + h2o_c[c] + 1516dhhdca_c[c] |
| **1694** | UAGT | Ustilagic Acid glycosyl transferase - UGT1 | 2 udpg_c[c] + 1516dhhdca_c[c] -> 2 h_c[c] + 2 udp_c[c] + cb1516dhdca_c[c] |
| **1695** | UAHA | Ustilagic Acid hexyl acylase and acetylase - UAT1,FAS2,UHD1 | h_c[c] + hexACP_c[c] + accoa_c[c] + nadph_c[c] + o2_c[c] + cb1516dhdca_c[c] -> ACP_c[c] + coa_c[c] + nadp_c[c] + h2o_c[c] + UA_A_c[c] |
| **1696** | UAAAO | UA A alpha oxidase | h_c[c] + nadph_c[c] + o2_c[c] + UA_A_c[c] -> nadp_c[c] + h2o_c[c] + UA_B_c[c] |

| | | | |
|---|---|---|---|
| **1697** | UAOA | Ustilagic Acid octyl acylase and acetylase - UAT1,FAS2,UHD1 | h_c[c] + ocACP_c[c] + accoa_c[c] + nadph_c[c] + o2_c[c] + cb1516dhdca_c[c] -> ACP_c[c] + coa_c[c] + nadp_c[c] + h2o_c[c] + UA_C_c[c] |
| **1698** | UACAO | UA C alpha oxidase | h_c[c] + nadph_c[c] + o2_c[c] + UA_C_c[c] -> nadp_c[c] + h2o_c[c] + UA_D_c[c] |
| **1699** | UA_Ate | UA A Transporter | UA_A_c[c] <=> UA_A_e[e] |
| **1700** | UA_Bte | UA B Transporter | UA_B_c[c] <=> UA_B_e[e] |
| **1701** | UA_Cte | UA C Transporter | UA_C_c[c] <=> UA_C_e[e] |
| **1702** | UA_Dte | UA D Transporter | UA_D_c[c] <=> UA_D_e[e] |
| **1703** | EX_UA_A_e | Exchange Hexoyl Ustilagic Acid - UA A | UA_A_e[e] -> |
| **1704** | EX_UA_B_e | Exchange Hexoyl alpha-Hydroxy Ustilagic Acid - UA B | UA_B_e[e] -> |
| **1705** | EX_UA_C_e | Exchange Octyl Ustilagic Acid - UA C | UA_C_e[e] -> |
| **1706** | EX_UA_D_e | Exchange Octyl alpha-Hydroxy Ustilagic Acid - UA D | UA_D_e[e] -> |

**Table A2.2 -** Deleted metabolites from iUma22 model, leading to the production of homologs of ustilagic acid.

| **Abbreviation** | **KEGG ID [compartment]** | **Description** |
|---|---|---|
| UA_A_c[c] | No KEGG ID [Cytoplasm] | Hexoyl Ustilagic Acid - UA A |
| UA_B_c[c] | No KEGG ID [Cytoplasm] | Hexoyl alpha-Hydroxy Ustilagic Acid - UA B |
| UA_C_c[c] | No KEGG ID [Cytoplasm] | Octyl Ustilagic Acid - UA C |
| UA_D_c[c] | No KEGG ID [Cytoplasm] | Octyl alpha-Hydroxy Ustilagic Acid - UA D |
| UA_A_e[e] | No KEGG ID [Extracellular space] | Hexoyl Ustilagic Acid - UA A |
| UA_B_e[e] | No KEGG ID [Extracellular space] | Hexoyl alpha-Hydroxy Ustilagic Acid - UA B |
| UA_C_e[e] | No KEGG ID [Extracellular space] | Octyl Ustilagic Acid - UA C |
| UA_D_e[e] | No KEGG ID [Extracellular space] | Octyl alpha-Hydroxy Ustilagic Acid - UA D |
| 16hhdca_c[c] | No KEGG ID [Cytoplasm] | 16-Hydroxy Hexadecanoate |
| 1516dhhdca_c[c] | No KEGG ID [Cytoplasm] | 15,16-dihydroxy Hexadecanoate |
| cb1516dhdca_c[c] | No KEGG ID [Cytoplasm] | Cellobiosyl 15,16-dihydroxy Hexadecanoate |

**Table A3.1 -** Added metabolites and reactions to iUma22 model. The corresponding command to each addition is described after the tables.

| **Abbreviation** | **KEGG ID [compartment]** | **Description** |
|---|---|---|
| Oil_e[e] | No KEGG ID [Extracellular space] | Oil |
| Oil_c[c] | No KEGG ID [Cytoplasm] | Oil |

| **Number of the reaction** | **Abbreviation** | **Description** | **Reaction** |
|---|---|---|---|
| **1857** | EX_Oil_e | Exchange Oil (triacilglyceride) | Oil_e[e] -> |
| **1858** | Oil_Tr | Oil Transporter | Oil_e[e] <=> Oil_c[c] |
| **1859** | FA_TAG_Production | Oil ? | Oil_c[c] <=> CPD-409_c[c] |
| **1860** | FA_1 | Fatty acids? | Fatty-Acids_c[c] -> **accoa_c[c]** |

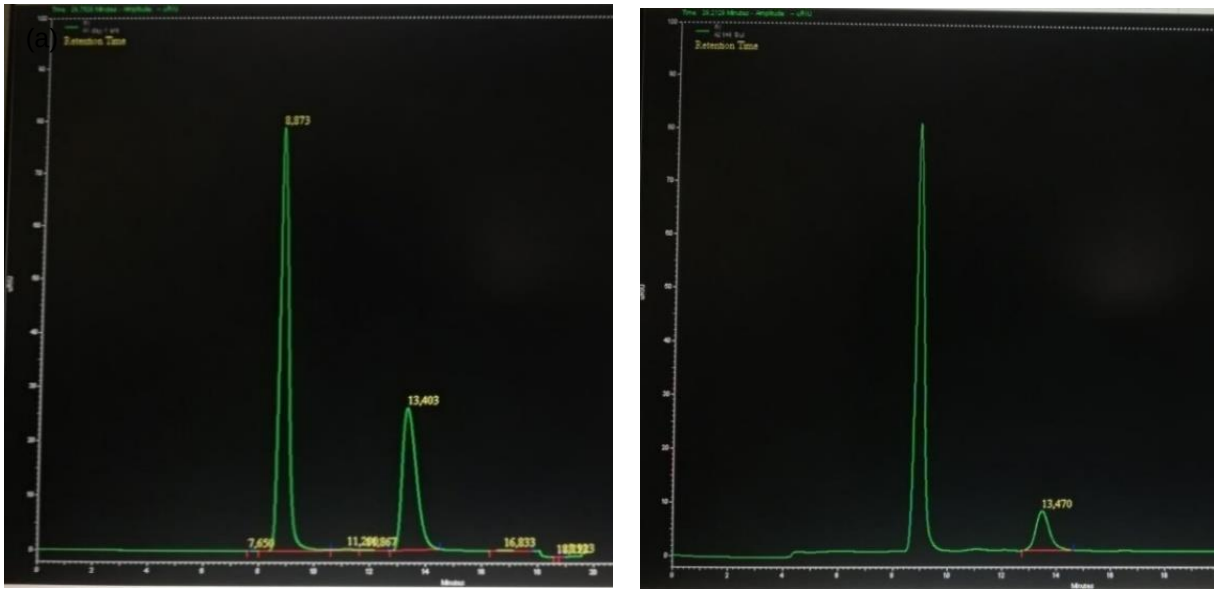
**Comands of the added metabolites and added reactions:**

```
      model = addMetabolite(model,'Oil_e[e]','Oil_e')
      model = addMetabolite(model,'Oil_c[c]','Oil_c')

[model, rxnIDexists] =
addReaction(model,'EX_Oil_e','reactionFormula','Oil_e[e]  -> ');

[model, rxnIDexists] =
addReaction(model,'Oil_Tr','reactionFormula','Oil_e[e] <=>
Oil_c[c]');
[model, rxnIDexists] =
addReaction(model,'FA_TAG_Production','reactionFormula','Oil_c[c]
<=> CPD-409_c[c] ');
[model, rxnIDexists] =
addReaction(model,'FA_1','reactionFormula','Fatty-Acids_c[c] ->
accoa_c[c] ');

Adds a reaction to the model or modify an existing reaction

  USAGE:

     [model, rxnIDexists] = addReaction(model, rxnID, varargin)

  INPUTS:
     model: COBRA model structure
     rxnID: Reaction name abbreviation (i.e. 'ACALD')

Authors:
        - Markus Herrgard 1/12/07
       - Richard Que 11/13/2008 Modified the check to see if duplicate reaction
already is in model by using S matrix coefficients to be able to handle larger
matricies
       - Ines Thiele 08/03/2015, made rxnGeneMat optional
       - Thomas Pfau May 2017  Change To parameter Value pairs
```

**Table A3.2 –** Flux values obtained with the iUma22 model and the model with the added metabolites and reactions.Differences obtained when utilising the modified model, with comparison with the original (iUma22) model are highlighted in the last column.

| Conditions | Reactions | iUma22 model | iUma22 model with additions | Difference from the "original" model |
|---|---|---|---|---|
| **For Biomass maximization**<br><br>**Glucose input (50g/L) = 2.20 mmol/gDCW/h** | FBA Solution | 0.169 | 0.169 | --- |
| | BIOMASS_REACTION | 0.169 | 0.169 | --- |
| | EX_o2_e | -0.011 | -0.011 | --- |
| | EX_nh4_e | -1.113 | -1.113 | --- |
| | EX_pi_e | 7.087 | 7.087 | --- |
| | EX_so4_e | -0.014 | -0.014 | --- |
| | EX_h2o_e | -3.830 | -3.800 | **0.78 %** |
| | EX_ca2_e | 0.000 | 1.598e-13 | Active |
| | EX_glc__D_e | -2.200 | -2.200 | --- |
| | EX_gly_e | 0.317 | 0.317 | --- |
| | EX_mal__L_e | 1.332 | 1.303 | **2.18%** |
| | EX_fum_e | 0.000 | 0.030 | Active |

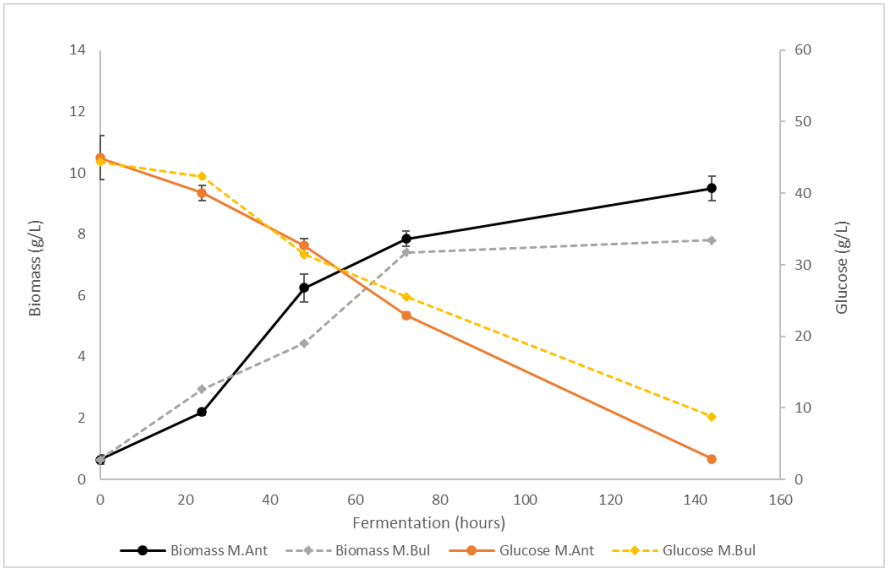
| | | | | |
|---|---|---|---|---|
| **MAC2 reaction maximization**<br><br>**Glucose input (50g/L) = 2.20 mmol/gDCW/h** | FBA solution | 0.317 | 0.317 | --- |
| | EX_co2_e | 2.366 | 2.366 | --- |
| | EX_nh4_e | -0.106 | -0.106 | --- |
| | EX_pi_e | 0.000 | -6.897e-14 | Active |
| | EX_h2o_e | 3.383 | 3.383 | --- |
| | EX_ca2_e | 0.000 | 1.598e-13 | Active |
| | EX_glc__D_e | -2.200 | -2.200 | --- |
| | EX_mal__L_e | 0.277 | 0.277 | --- |
| | EX_gly_e | 0.106 | 0.106 | --- |
| | EX_cys__L_e | 0.000 | 2.700e-16 | Active |
| | EX_met__L_e | 0.000 | -2.700e-16 | Active |
| | MAC2 | 0.317 | 0.317 | --- |
| | EX_MEL_D_e | 0.317 | 0.317 | --- |
| | E4PK; ER; EMT1 MAC1; MAC2; MEL_Dte | 0.317 | 0.317 | --- |



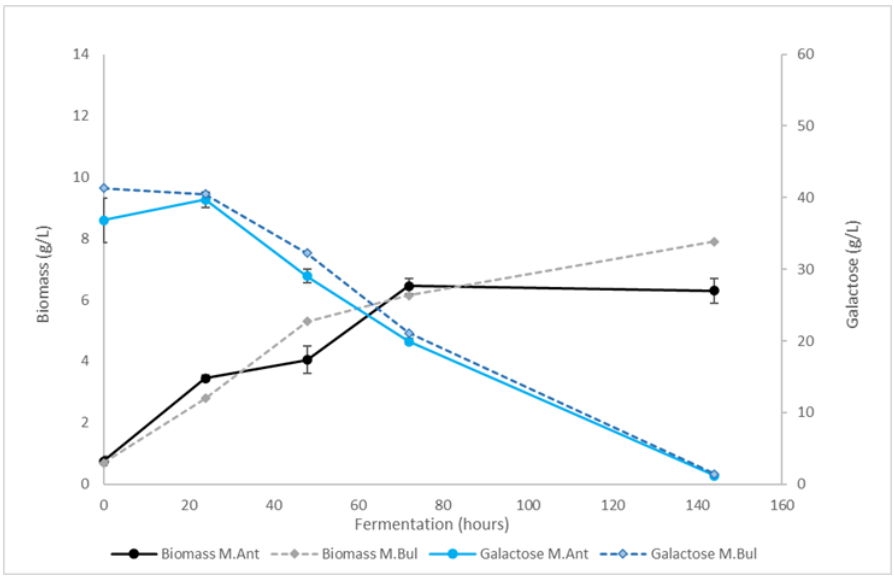

**(A)** **(B)**

**Figure A2** – HPLC peaks of glucose from inoculum used in the fermentation. (A) Glucose at day 1 with *M. antarcticus* strain. A peak with a value at 13.403 minutes is registered. (B) Glucose at day 6 with *M. bullatus* strain. A peak with a value at 13.470 minutes is registered.

**(A)**




**(B)**




**(C)**

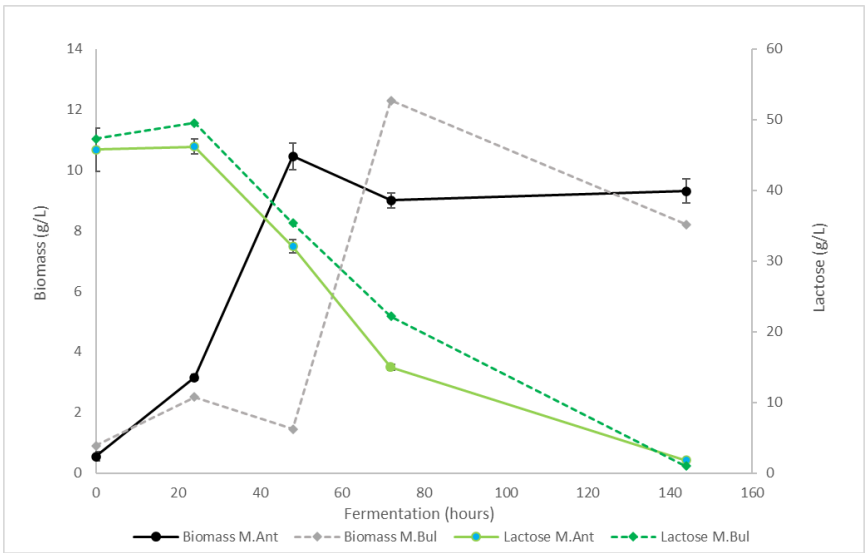



**Figure A3** - Biomass (g/L) production and sugars (g/L) consumption values at day 0 (0 hours), day 1 (24 hours), day 2 (48 hours), day 3 (72 hours) and day 6 (144 hours) of fermentation with *M. antarcticus* (filled line) and *M. bullatus* (dashed line) strains. The sugar used as carbon source is glucose (A), galactose (B) and lactose (C).

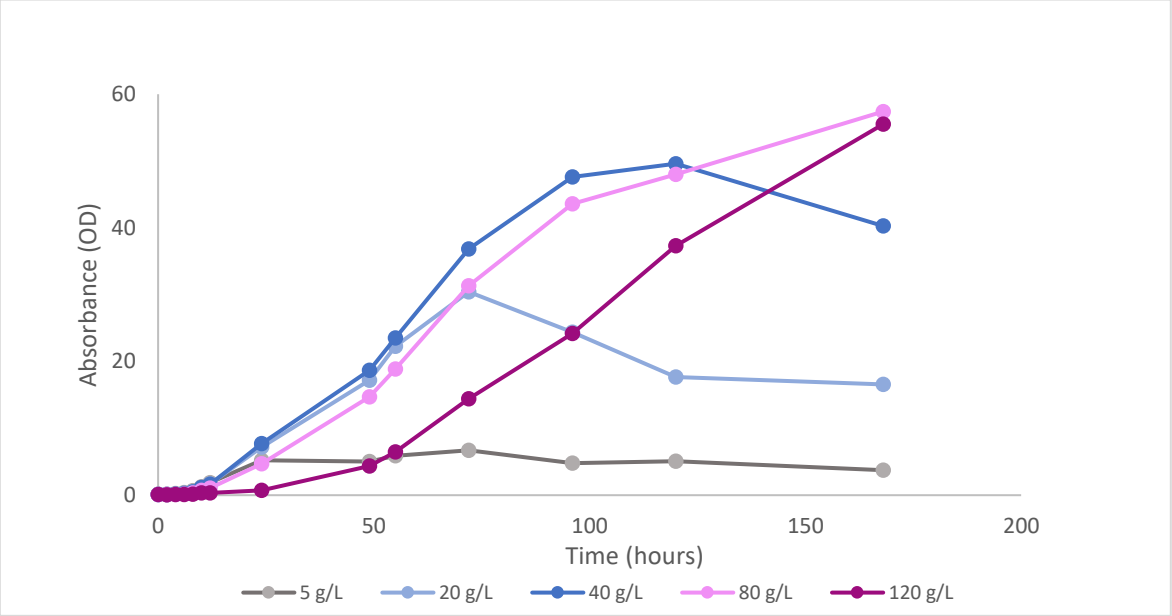



**Figure A4** - *M. antarcticus* growth values of absorbance (OD) through time (hours) with different glucose concentrations. 5 g/L (grey); 20 g/L (light blue); 40 g/L (dark blue); 80 g/L (light pink); 120 g/L (dark pink).

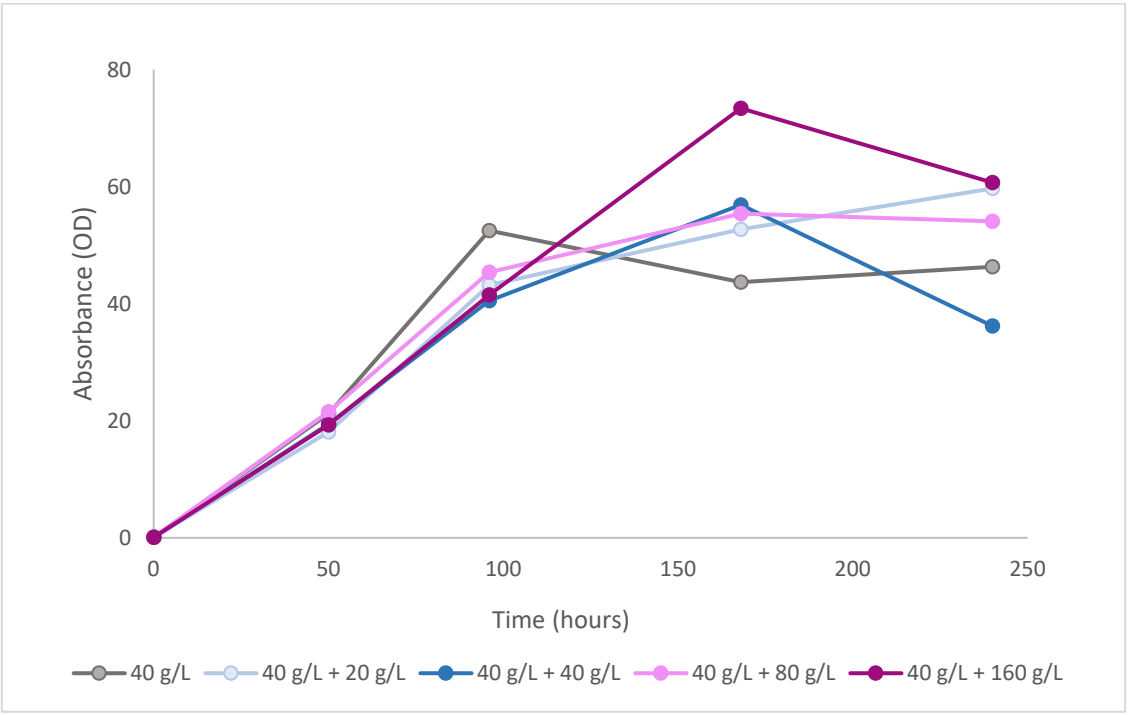
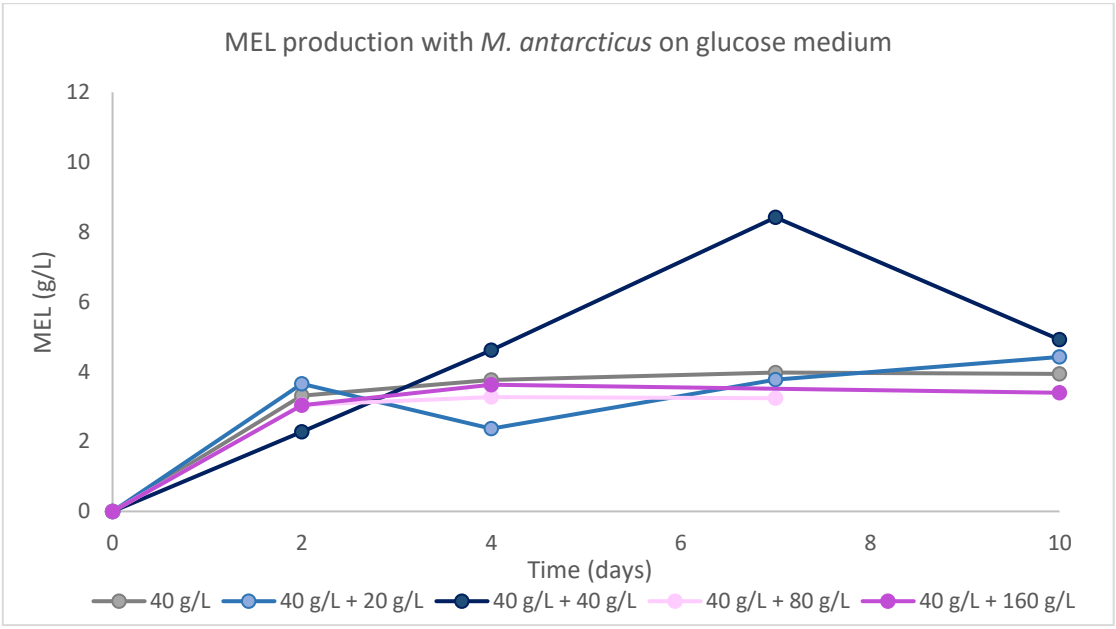



**Figure A5** - *M. antarcticus* growth values of absorbance (OD) through time (hours) with different glucose concentrations. 40 g/L (grey); 40 g/L + 20 g/L (light blue); 40 g/L + 40 g/L (dark blue); 40 g/L + 80 g/L (light pink); 40 g/L + 160 g/L (dark pink).

(A)




(B)

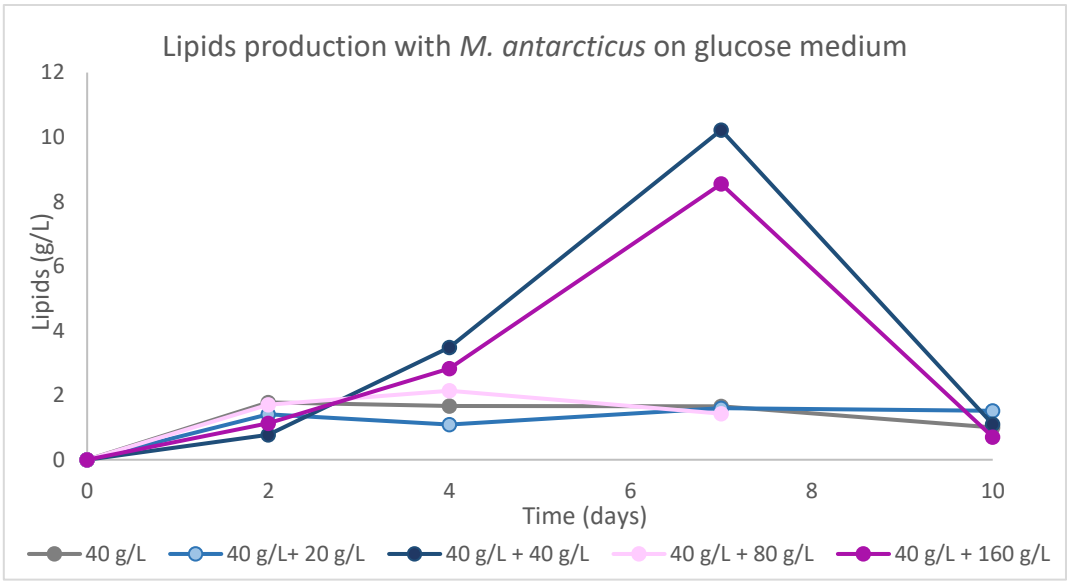



**Figure A6** – Values of MEL and lipids production (g/L) through time (days) with different concentrations of glucose. 40 g/L (grey); 40 g/L + 20g/L (light blue); 40 g/L + 40 g/L (dark blue); 40 g/L + 80 g/L (light pink); 40 g/L + 160 g/L (dark pink). (A) Values of MEL production, conserving GC values for C8, C10, C12 and C14. (B) Values of Lipids production, considering GC values for C16 and C18.

**Table A4** – Values from iUma22 model updated to *M. antarcticus*, with deletion and addition of specific reactions, optimized to MEL production.

| Conditions | Reactions | iUma22 model | iUma22 model with deletions and addition | Difference from the "original" model |
|---|---|---|---|---|
| **MAC2 reaction maximization**<br><br>**Glucose input (50g/L) = 2.20 mmol/gDCW/h** | FBA solution | 0.317 | **0.317** | **---** |
| | EX_co2_e | 2.366 | 2.366 | **---** |
| | EX_nh4_e | -0.106 | -0.106 | **---** |
| | EX_pi_e | 0.000 | 1.202e-12 | **Active** |
| | EX_h2o_e | 3.383 | 3.383 | **---** |
| | EX_ca2_e | 0.000 | 1.598e-13 | **Active** |
| | EX_glc__D_e | -2.200 | -2.200 | **---** |
| | EX_mal__L_e | 0.277 | 0.277 | **---** |
| | EX_gly_e | 0.106 | 0.106 | **---** |
| | EX_cys__L_e | 0.000 | 9.081e-14 | **Active** |
| | EX_met__L_e | 0.000 | -9.081e-14 | **Active** |
| | MAC2 | 0.317 | 0.3171 | **---** |
| | EX_MEL_D_e | 0.317 | 0.3171 | **---** |
| | E4PK; ER; EMT1 MAC1; MAC2; MEL_Dte | 0.317 | 0.317 | **---** |