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

Name: Fátima Mendes Costa
Supervisors: Prof. Dr. Frederico Castelo Alves Ferreira; Prof. Dr. Nuno Ricardo Torres Faria

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 modeling 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 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 investigation of metabolic responses to various environmental conditions.

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

1. Introduction

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, home detergents, and medications. Delivering a more environmentally friendly surfactant alternative to chemical surfactants is possible with microbial surfactants, classified as “biosurfactants”.[1] 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.[2] Nonetheless, due to their high production costs, they are hardly assessing the market. The knowledge gained from the metabolic engineering of microbial lipids for biofuel production can be used for effective synthesis, diminishing the production costs problem.

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 (Fig.1) 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 length of each fatty acid residue is influenced by the respective producer organism, the *Moesziomyces* yeast (previously referred to in the literature as *Pseudozyma*), of *Ustilaginaceae* family. [3,4] [5] 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.[4]

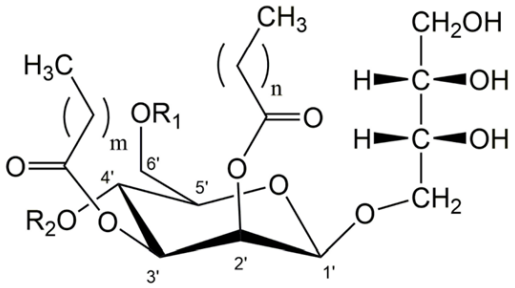

Figure 1 – General structure of di-acylated mannosylerythritol lipid. (MEL-A: R1=R2=Acetyl; MEL-B: R1=Acetyl ,R2=H; MEL-C: R1=H, R2=Acetyl, MEL-D: R1=R2=H). Variable chain-length and saturation of fatty acid side-chains at C2 (m= 2–16) and C3 (n= 2–10). [4]

MELs are generally produced 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.[6][7] 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. 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. [5]

The sugar core of MELs is produced by binding an erythritol molecule, through the pentose phosphate pathway, onto GDP-mannose, obtained via glycolysis, from several hydrophilic precursors. The hydrophobic tail is formed of fatty acids that can be added to the culture medium and integrated into the MEL or synthesized *de novo* by the microorganism. Partial peroxisomal *β*-oxidation, is another step of MEL metabolism, giving MEL's unique fatty acid patterns, since it provides fatty acids for MEL's acylation. [8] 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. A review of the most important metabolic pathways already stated was constructed using *U. maydis* as the model organism and compared to genomic data for various MEL-producing strains (Fig.2). [9] To forecast genetic modifications that rearrange the metabolism toward the generation of the compound of interest, metabolic modelling is required.

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. 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, by varying the limits on the exchange reactions that operate as sources of substrate and waste metabolites, different circumstances can be simulated as well. Flux balance analysis (FBA) can be used to analyse constraint-based models by the constraint-based reconstruction and analysis (COBRA) strategy. [10] [11]

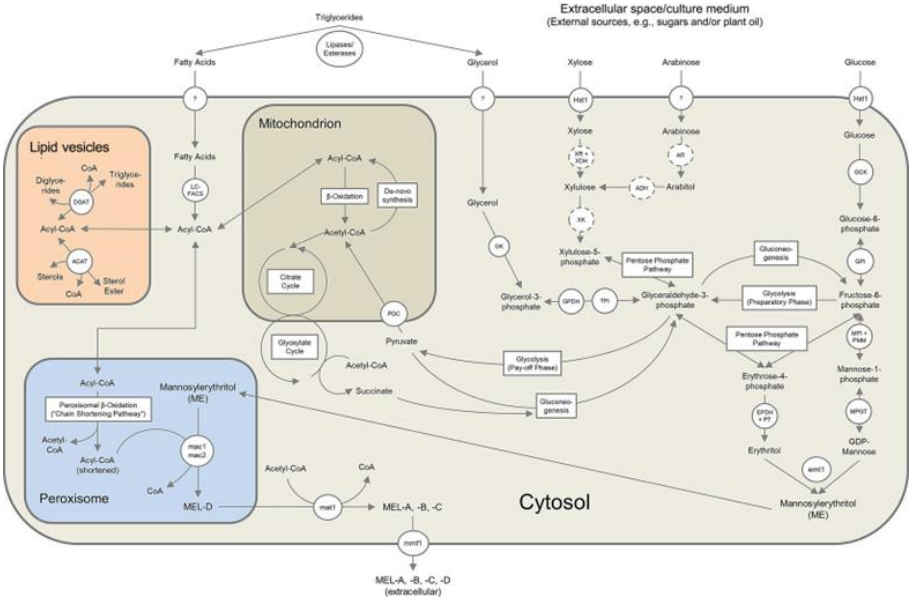

Figure 2 – Detailed metabolic pathways that lead to the formation of MEL from various substrates in *U.maydis*. [9]

2. Materials and Methods

2.1 MEL production

2.1.1. 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 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.

2.1.2. Media and cultivation conditions

An inoculum was prepared according to the procedure referred to in the literature. [12] The glycerol stocks of *M. antarcticus* was transferred to an erlenmeyer flask, with 50 mL of a mineral medium (3 g/L NaNO3, 0.3 g/L $MgSO_4.7H_2O$, 0.3 g/L KH_2PO_4, 1 g/L yeast extract) 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. Initially, 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. Growth rates and glucose uptake rates were estimated using experiments carried out for this study and data from the literature.[13,14]

2.1.3. Yeast cultivation parameters

2.1.3.1. 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 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.

2.1.3.2. 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). [15] The DCW conversion coefficient value used for growth in *M. antarcticus* was 0.30 g DCW/L/OD_{600}. [16] A linear relationship of the exponential growth phase was reached with the ln(OD) over time (Equation A1). The growth rate of each condition was obtained with the slope value of each condition. [17]

2.1.4 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, as described in the literature.[13] The substrate uptake rates were estimated with a linear equation and converted into mmol/gDCW/h (Equation A2). [17]

2.1.5. MEL and fatty acids quantification

The methyl ester derivatives were prepared to allow the quantification of MEL and fatty acids with feeding of different concentrations at day 4. 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 experimental protocol procedure was followed according literature [13] and MELs were analysed by Gas Chromatography (GC) system (Hewlett-Packard, HP5890). 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.

2.2. 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. [18] The metabolic model used was iUma22, the first genome-scale metabolic model (GSMM) of *Ustilago maydis*, reconstructed from sequencing and annotation for the simulation of metabolic activities and is characterized by 1855 reactions and 1233 species of metabolites. Through experiments with glucose growth, the model's quality was evaluated and the precision of rate predictions was examined. [19,20]. Until the moment there is not a described MEL metabolism model for *M. antarcticus*.

2.2.1. Pathways Reconstruction

The available model for *U. maydis*, retrieved from GitHub database, was modelled 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 1.3.). 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.

2.2.2. 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).[21] The versatile function `optimizeCbModel` can be used to compute FBA and many of its variants. [22] 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. Nonetheless, the objective of this work is to upgrade the available GSMM for MEL production by *Moesziomyces antarcticus*. 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 and consequently eliminate 11 metabolites related to the production of Ustilagic Acid (Fig.A1). This was possible using the command: `>> model = removeRxns(model, rxnRemoveList);`

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 (Fig.A2). 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}. 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. [10] 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, one of the four homologs of MEL:

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

3. Results and Discussion

3.1 *In silico* metabolism of *Ustilago Maydis* model

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. Pathways dynamic visualization makes it possible to see where the fluxes values retrieved by MATLAB are present on the cell. (Fig.A3)

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. Imposing different flux values for a substrate, 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. [23] The attributed values for fluxes will generate a solution, according to the premise that the objective function must be optimised [24]and the solution value should be a value of flux, that corresponds to the organism's exponential growth rate, when biomass reaction is the objective reaction.

3.1.1. Deleted and added reactions from *Ustilago maydis* model

The *Ustilago maydis* model, iUma22, to being upgraded to a *Moesziomyces Antarcticus* model, needed to suffer the deletions of reactions and metabolites that are present in *U. maydis*, but not in *M. antarcticus*. The reactions leading to the production of Ustilagic Acid were deleted, and the FBA solution values were obtained having biomass equation and MEL production as objectives. All the predictions were performed with the same constraint. The values obtained (Table A1) showed some flux values slightly different from the values obtained with iUma22 model, in the condition where biomass reaction was the objective reaction. The value of flux of NH_4, H_2O, malate, and the absence of entry of glycerol are the four fluxes that registered changes. With the maximization of MAC2 reaction, the reaction that leads to the production of MEL-D, the biomass reaction does not present any flux. The entry of oxygen (O_2) is 0.00 mmol/gDCW/h with MEL production as the objective, that also presented the production of CO_2. The glycerol flux value is also lower, which means that the flux of production is decreased.

The GSMM presented for *U. maydis* also lacks some reactions present in *M. antarcticus*. Since peroxisome is not present in this model, the main pathways that are lacking are related to lipids' metabolic reactions. The addition of reactions lead to minimal changes in the flux values, compared with values obtained to the iUma22 model without this modification. 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 2% of the value obtained in the original model.

3.2. The effect of glucose on growth and MEL production of *Moesziomyces antarcticus* - Batch cultures at different glucose concentrations

With the experiment described in the previous chapter, *M. antarcticus* strain growth rate values were assessed with 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 values were also calculated to have values from flux with glucose, possible to apply in the model as a constraint. These were achieved with the HPLC technique of sugar quantification and are represented at Table 1. The growth rate value for each condition was also obtained, and both values were compared with literature values reached with similar conditions, but using *U. maydis* as the organism, instead of *M. antarcticus*. 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. It was possible to compare the experimental and the *in silico* outputs and the values obtained were registered between 5% and 9% with *U.maydis*, 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. [19]

Table 1 – Growth rate (h^{-1}) values and substrate uptake values (mmol/gDCW/h) obtained with experiments described in the literature with *U. maydis* [19,25] 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)	Initial Concentration (g/L)	Substrate-uptake (mmol/gDCW/h)	Growth Rate (h^{-1})		Error (%)
			Experimental	Predicted	
M. antarcticus 5 g/L	4	2.05	0.16	0.16	3
M. antarcticus 20 g/L	15	2.30	0.20	0.18	11
U. maydis 20 g/L [19]	19	6.24	0.27	0.51	47
M. antarcticus 40 g/L	32	0.43	0.03	0.02	83
U. maydis 50 g/L [19]	50	2.20	0.18	0.17	7
U. maydis 50 g/L [25]	54	1.22	0.08	0.09	6
M. antarcticus 80 g/L	77	2.69	0.18	0.21	12
U. maydis 100 g/L [25]	106	0.67	0.04	0.04	5
M. antarcticus 120 g/L	121	0.37	0.06	0.01	390
U. maydis 130 g/L [19]	126	1.10	0.07	0.08	7
U. maydis 130 g/L [19]	132	0.74	0.04	0.04	9
U. maydis 200 g/L [19]	203	0.33	0.02	0.01	122
U. maydis 200 g/L [19]	216	0.36	0.02	0.01	122

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, as expected. 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. 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.[4,25] 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* and with 50 g/L of glucose with *U. maydis*.

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. 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*, 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 considered.

3.3. Fitting of experimental results with modelled results

With the values obtained experimentally with the *M.antarcticus* strain a fitting of the values to the iUma22 model was desired. 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.

3.3.1. *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. The experimental values obtained with *M. antarcticus* strain and the values with *U. maydis* retrieved from literature were compared (Fig.3). 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. 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 [13,14], 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 1), this value is double.

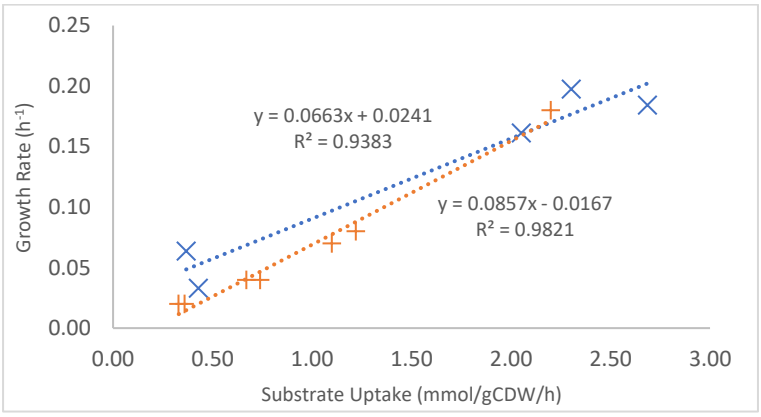

Figure 3 - Comparision of values of experimental rates. Growth rate (h^{-1}) and substrate uptake rate (mmol/gDCW/L) values were obtained experimentally with *U. maydis* (orange) and *M. antarcticus* (blue). *U. maydis* values were retrieved from the literature [19] and the linear regression line equation 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.

3.3.2. *Moesziomyces antarcticus* fitting model with MEL optimization

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, with the final goal of producing at least one MEL homolog. The first approach was to understand if the same values were obtained *in silico*, without having any change in the *U. maydis* model. This was done by maximizing the reaction and applying the flux values of glucose experimentally obtained to the reaction of glucose uptake, D-Glucose exchange reaction. In Table 2 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 are bigger than the values obtained from the comparison of growth values. Since the majority of the values were retrieved from work previously reported in the literature, it

was also possible to acquire 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 67%. With xylose fed-batch experimental values, the error presented was higher than 1000%.

Table 2 - 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, [14] comparing the experimental values with the value predicted retrieved through MATLAB. Glu.40:Glu.80, Glu.80:Glu.40 and Glu.80:Glu.80 – 40 g/L or 80 g/L of glucose at day 0 and addition of 40 g/L or 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.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

4. Discussion

The model iUma22 was constructed with *Ustilago maydis* as the organism of reference. All the principal metabolic routes that allow the production of MEL are present, however, is necessary to notice the absence of peroxisome in this model. Nevertheless, the organism of interest is not *U. maydis*, but rather the *Moesziomyces* specie. With the pathways highlighted and applying the constraint of the input flux value of glucose on the cell, was possible to obtain the predicted values of growth on this model, which were compared to growth rate values obtained to *U. maydis*. These only showed a difference between 5% and 9% of the experimental values, obtained with glucose concentrations between 50 g/L and 130 g/L. The values obtained for lower concentrations or high values of glucose showed a difference of more than 47%, meaning that the model has a good prediction for most of the inputs, but has to be well curated with lower and higher values of glucose.

To understand how the predictions of the model differ when employing values obtained with *M. antarcticus* and not with *U. maydis*, fermentations with glucose as a carbon source were realized, and after the results were compared with the predicted values of the model. 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 was already mentioned by the authors in the literature. [19]

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*, which was shown in Figure 3. 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 have a perfect fit from the experimental values. Nevertheless, as observed in Figure 4, this difference is very low, explained by the fact that the model used to do the predictions was not curated to be used with the organism of this work.

The values of MEL and lipids concentration previously obtained [14] were also analysed, to compare the flux production value of MEL with *in silico* approaches. Comparing the error values between the

values obtained experimentally and the values predicted, when MEL production was optimised, the values of error obtained are 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 explained with the model being recently released and optimized to have a fitting for growth, and not to specific metabolites.

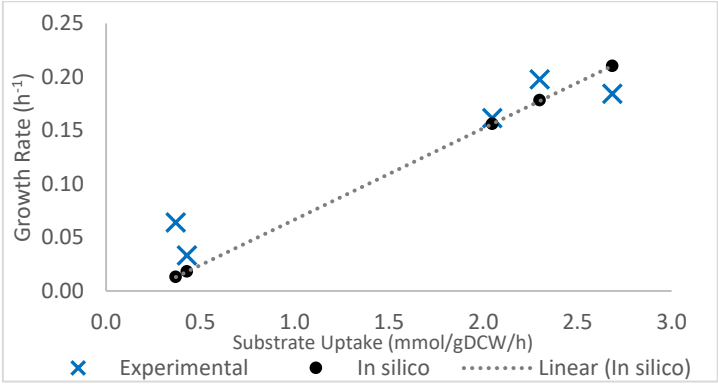

Figure 4 - Growth rate and substrate uptake rate experimental values obtained with M.antarcticus strain in batch experiments with different glucose (blue crosses) in contrast to the linearity obtained to growth rate values with the same substrate uptake rate value, predicted in silico with the iUma 22 model (black dots).

5. Conclusion and Future Prospects

The present model was not constructed to have predictions of very low or very high values of glucose concentration, since it is just possible to achieve some linearity in the data between 40 g/L and 130 g/Lf glucose. Even if the growth values and substrate uptake values obtained with *M. antarcticus* were not that different from the ones obtained in the literature with *U. maydis* [19] not having duplicates of any experiment could also influence in this difference. However, the predicted values of flux for MEL production presented a bigger discrepancy than the ones obtained experimentally. Thus, to have a higher level of confidence in that values, other analytical approaches could be used. Nuclear Magnetic Resonance (NMR) experiments can enable the establishment of the flux distribution by ^{13}C-NMR experiments. 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. By combining genomics, RNA-seq, and NMR data, as well as transcriptomics and metabolomics studies, it may be possible to better understand not only the MEL-cluster behavior but also the key genes involved in its regulation, [26] enabling the achievement of a more accurate metabolic model of *Moesziomyces*.

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. The final objective is to reach logical strategies to produce this biosurfactant, increasing substrate carbon use efficiency, and bringing biosurfactant economic costs to be competitive with fossil oil-driven surfactants. 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.

References

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

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

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

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

21. 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.

22. 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.

23. 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.

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

25. 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.

26. 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.

Appendix

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

Equation A1 – 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); μ= growth rate with units of 1/h or OD/h for exponential and linear growth, respectively.[17]

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

Equation A2 – 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.

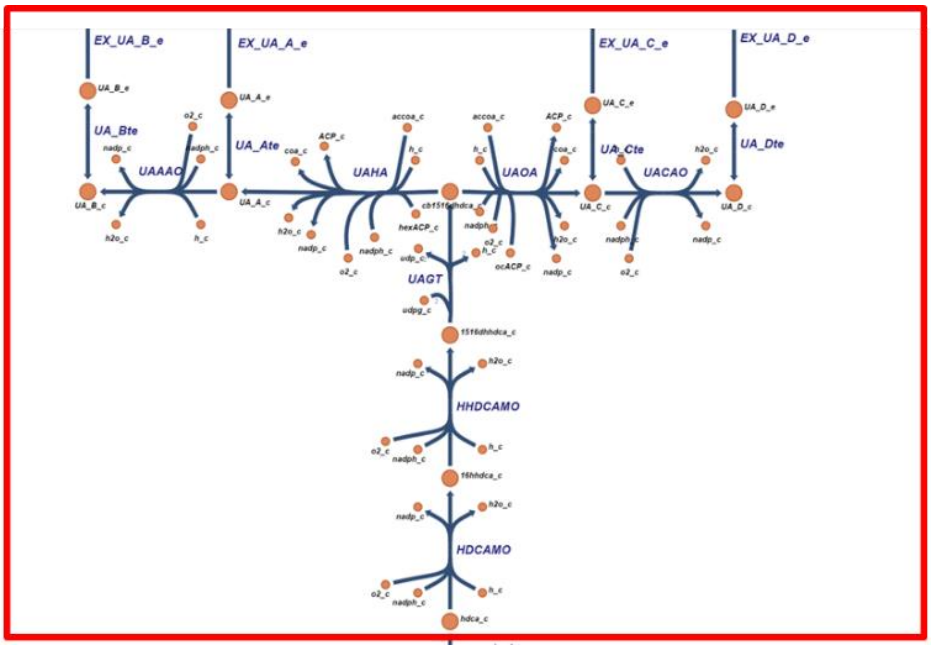

Figure A1 – Highlight of the deleted reactions from the iUma22 model, made with the Escher building application.

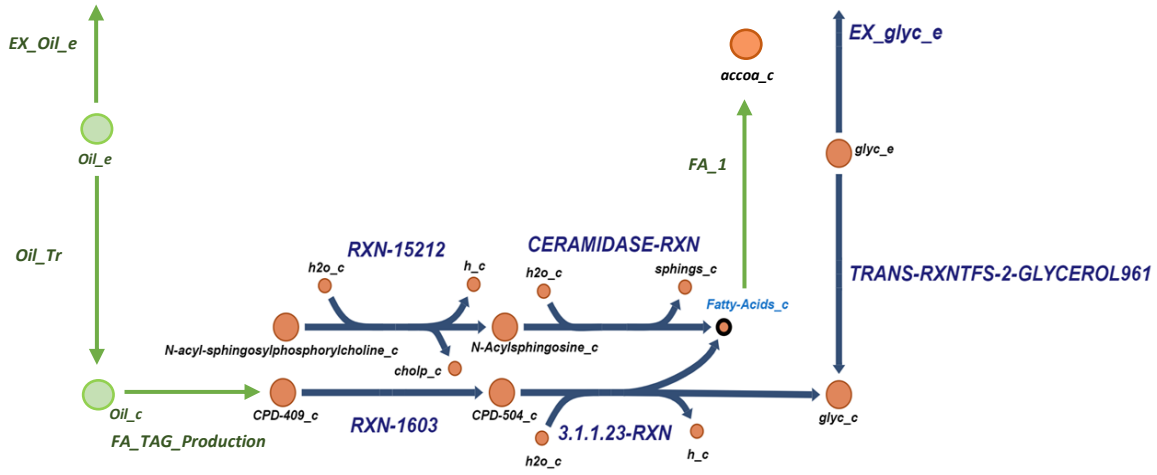

Figure A2 – Addition of reactions and metabolites to fix dead-end reactions. Added reactions and added metabolites (green), reactions (blue) and metabolites (orange) already present in the iUma22 model. The original reactions are visualized by Escher software with building option.

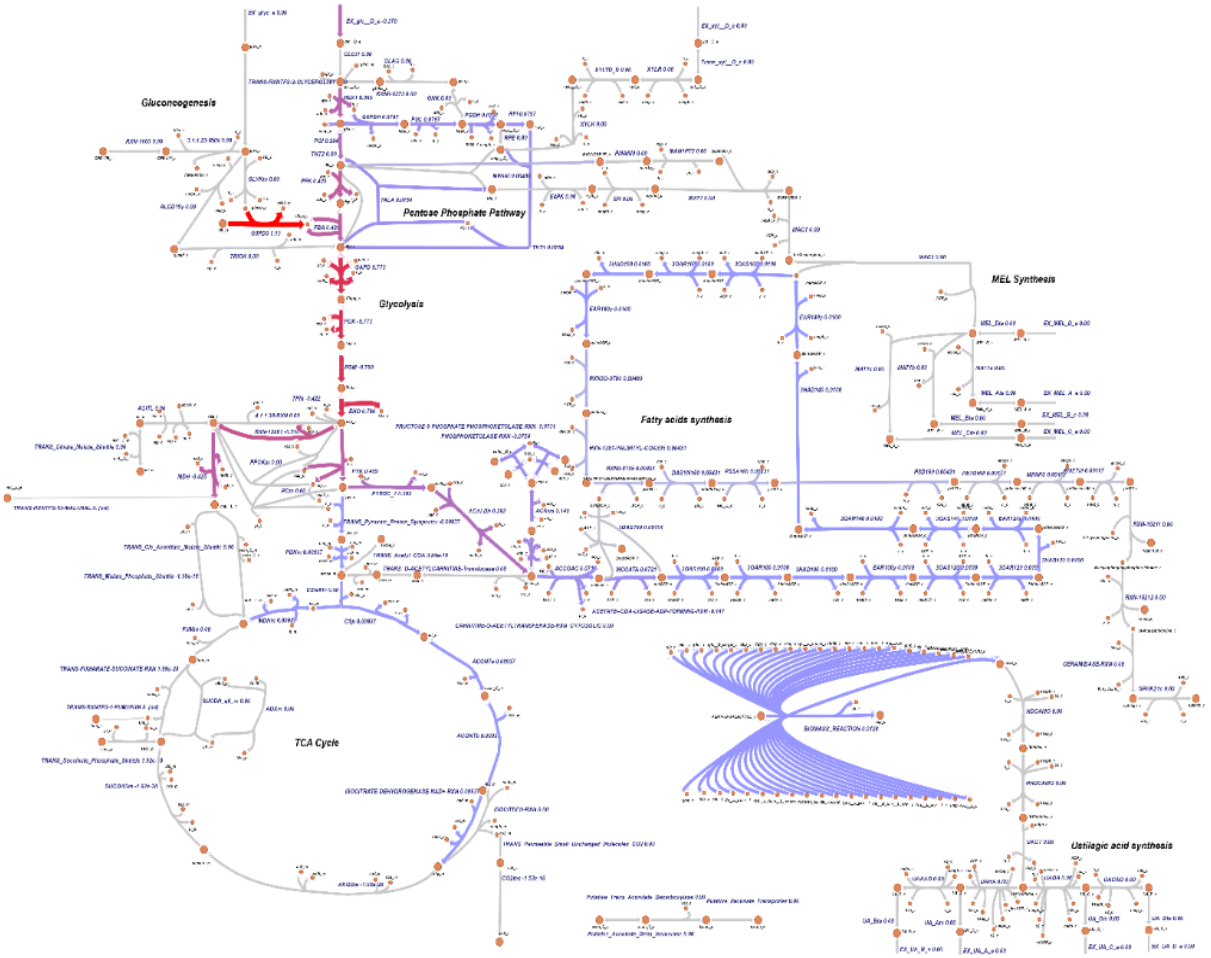

Figure A3 - Visualization of fluxes of the principal pathways that lead to MEL production in *U. maydis*, with maximization of biomass reaction, in the iUma22 model with Escher-FBA application.

Table A1 - *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 **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 **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	0.3171