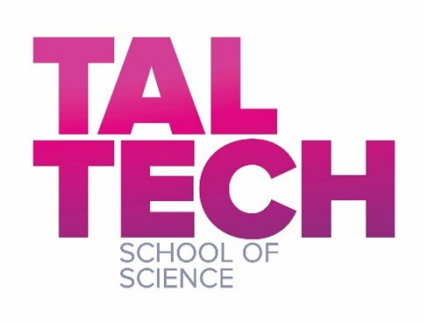

Comparison of genome-scale metabolic models for investigating lipogenesis metabolism in *Rhodotorula toruloides*

Bachelor thesis

Student: Maive Hanni
Supervisor: Alina Rekena,
Early Stage Researcher,
Department of Chemistry and Biotechnology
Study program: Applied Chemistry and Gene Technology

Tallinn 2024

Ülegenoomsete metaboolsete mudelite võrdlus *Rhodotorula toruloides* lipogeneesi uurimiseks

Bakaluareusetöö

Üliõpilane: Maive Hanni
Juhendaja: Alina Rekena, Doktorant-nooremteadur, Keemia ja biotehnoloogia instituut
Õppekava: Rakenduskeemia ja geenitehnoloogia

Tallinn 2024

Declaration

Hereby I declare that I have compiled the paper independently and all works, important standpoints and data by other authors have been properly referenced and the same paper has not been previously been presented for grading.

Author: Maive Hanni
Signature, date

The paper conforms to requirements in force.
Supervisor: Alina Rekena
Signature, date

Permitted to the defence. Chairman of the Defence Committee: Indrek Koppel
Signature, date

Contents

Abstract

Annotatsioon

Abbreviations

Introduction

1. Theoretical background

1.1. Need for new technologies to reduce reliance on fossil-based resources

Many countries globally are developing a bio-based economy to fight climate change and to lower the reliance on fossil-based resources [1]. In Europe, the Bio-Economy Strategy was developed to steer Europe towards a sustainable bio-based economy and it was reinforced in the European Green Deal aiming for climate neutrality by 2050 [2]. Bio-based products may enhance environmental sustainability compared to fossil based equivalents [1].

The shift towards a bioeconomy needs novel processes for production of chemicals, materials, and liquid fuels from sustainable substrates, that offer improved life cycle assessments, and use less energy to produce. Recent advancements have highlighted the potential of chemicals derived from plant oils and animal fats as alternative feedstocks to the petrochemical industry.

These feedstocks comprise a broad spectrum of molecules that can be utilized in various applications such as biofuels, cosmetics, plastics, surface coatings, surfactants, lubricants, paints, etc. [3]. Biodiesel is synthesized through the transesterification of triacylglycerols with short-chain alcohols (primarily methanol or ethanol) to yield monoalkyl esters, specifically fatty acid methyl esters (FAMEs) and fatty acid ethyl esters (FAEEs) [4].

The growth of worldwide biodiesel production drives the need to increase the production of fatty acid methyl esters [5]. But the production of biodiesel from oilseeds and waste oils doen not sustain the global demand [4] and the recent food crisis has shown the need for the development of second-generation biofuels derived from lignocellulosic raw materials and industrial waste streams [6].

Microbial lipids are another source of fatty acids considered as a potential feedstock for oleochemical production [7]. Research has focused on the development of biodiesel production from single cell oil (SCO) that are produced via fermentation using oleaginous microorganisms (microorganisms capable of accumulating lipids at more than 20% of the total cellular dry weight).

Biodiesel production from SCO relies on the utilization of low-value waste streams or residues, thus presenting a sustainable alternative for biofuel production. Moreover, the production of SCOs does not require land or other resources that are typically used for food production and it is not influenced by season or climate. [4]

1.2. *Rhodotorula toruloides*

Rhodotorula toruloides (previously *Rhodosporidium toruloides*) is an oleaginous yeast which can accumulate lipids up to 76.1% of cell dry weight [8]. What is more, *R. toruloides* has good tolerance to inhibitory compounds that are naturally found in biomass hydrolysates [9]. It is one of the most promising yeasts for sustainable production of chemicals and fuels [10].

The majority of the lipids produced by *Rhodotorula toruloides* are triacylglycerol (TAG) contained long-chain fatty acids (C16:0 (palmitic acid), C16:1 (palmitoleic acid), C18:0 (stearic acid), C18:1(oleic acid), and C18:2 (linoleic acid)) and they are comparable to vegetable oils [8], [11].

R. toruloides occurs naturally in leaves, soil, sea water, etc. It has a broad substrate range, which has made this yeast a popular for producing biological oils from inedible substrates such as pentose sugars and crude glycerol. [12] *R. toruloides* lipid fraction contains also carotenoid pigments, omega-3 linolenic acid and heptadecenoic acid, which makes it a promising organism for production of pharma- and nutraceuticals [13].

One of the major determinants of the oleaginous phenotype of *R. toruloides* is its capacity for acetyl-coenzyme A (acetyl-CoA) production. *R. toruloides* possesses the enzyme ATP-citrate lyase (ACL), which has been suggested to be the main source of acetyl-CoA for lipid synthesis in oleaginous species, as it has not been found to be present in non-oleaginous yeasts [14]. Mitochondrial beta-oxidation pathway provides additional source of acetyl-CoA in *R. toruloides*. [12] Synthesis of acetyl-CoA from xylulose 5-phosphate by phosphoketolase (XPK), and the conversion of S-malate into pyruvate by malic enzyme (ME) that provides for nicotinamide adenine dinucleotide phosphate (NADPH) enzymatic pathways, also differ from the model yeast *Saccharomyces cerevisiae* and which specifically facilitate the generation of lipid precursors.[10] Lipid biosynthetic reactions downstream of acetyl-CoA synthesis do not differ between oleaginous and non-oleaginous yeast species [12].

In *R. toruloides* metabolic pathways producing acetyl-CoA and a cofactor NADPH have been the main focus of metabolic studies due to their central role in lipid biosynthesis. Fatty acids mainly accumulate as TAGs, and they are produced via four enzymatic reactions that require 1 ATP and 2 NADPH molecules for adding 1 acetyl-CoA to the fatty acid chain [15]. Proteomics analysis has suggested that NADPH is primarily regenerated through the pentose phosphate pathway, when grown on xylose and glucose, but the role of malic enzyme is not clearly understood. The role of XPK in the generation of acetyl-CoA has not been acknowledged previously, whereas ACL has been demonstrated to be upregulated during lipid accumulation, especially in presence of xylose.[10]

1.3. Overview of cellular growth laws

Lipid accumulation in oleaginous microorganisms has long been known to be triggered by a nutrient imbalance in the culture medium. When cells run out of a key nutrient, usually nitrogen, excess carbon substrate continues to be assimilated by the cells and converted into storage fat. Cells assimilate carbon quicker than they can convert it into new cells so mechanism for storage the excess carbon is then found by converting it into lipid. Lipid accumulation requires a slow growth rate of the cells to allow the excess carbon to be assimilated faster than it can be converted into biomass so that the surplus carbon is channeled into lipid. With oleaginous yeast the process of lipid accumulation can also be achieved in continuous culture, where is it necessary to grow the cells at a sufficiently low dilution rate (= growth rate) to allow the cells to assimilate the glucose. The results from continuous cultivation studies clearly indicate that the rate of lipid synthesis is slower than the maximum growth rate. [16]

Lipid accumulation occurs when oleaginous microorganisms are cultivated in a medium with an excess of carbon where other nutrients, particularly nitrogen, is limiting their growth. Therefore, the carbon-to-nitrogen ratio (C/N) plays an important role in triggering lipid accumulation. [3]

What is the biochemical difference between these two groups of very distinct microorganisms? The first major biochemical difference to be identified between the oleaginous and nonoleaginous yeast species was the presence of ATP-citrate lyase in oleaginous yeast during lipid accumulation. ACL has proved to be one of the key enzymes that must be present in a eukaryotic microbial cell for it to be able to accumulate substantial amounts of triacylglycerol lipids. In those yeasts without ACL, lipid contents of the cell were invariably low. However, some yeasts had ACL activity but did not accumulate lipid, thus indicating that other enzyme activities were needed to ensure lipid accumulation. In summary, while the possession of ACL activity will not automatically engender lipid accumulation in a microorganism, if the enzyme is absent the cells will be unable to accumulate lipid and will therefore be nonoleaginous. Clearly, other enzyme activities must be in place to determine the extent to which lipids may accumulate in individual organisms. [16]

Malic enzyme has found to be another important enzyme in lipogenesis. It generates NADPH, which is used by fatty acid synthetase (NADPH is generated also by glucose-6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase, and NADP-dependent isocitrate dehydrogenase). When malic enzymes activity was blocked by using selective inhibitors (sesamol), he content of lipid in the cells was decreased by almost 90%, from 24% of the cell biomass to 2% but — and this was of major significance — without causing any major effect on growth. Thus, as only malic enzyme activity appeared to have been affected by the inhibitor, it was concluded that sesamol was specifically inhibiting both the cytoplamic and membrane-bound malic enzymes, and that without malic enzyme the cell was unable to accumulate lipid or to carry out desaturations of it. The essentiality of malic enzyme for lipid biosynthesis appeared to be established. [16]

Wynn and Ratledge (1997) went on to show that in a mutant of *Aspergillus Nidulans* lacking malic enzyme activity, only half the lipid (12% of the cell dry weight) that had been produced by the competent strain under nitrogen-limited growth conditions was now produced. [16] Fatty acid biosynthesis per se is still functional and phospholipids can be produced. Thus, the cell can manage without malic enzyme—it is not absolutely vital—but the cell cannot produce storage triacylglycerols in any abundance. Without malic enzyme activity the flux of carbon, from glucose to lipid, is considerably diminished and only essential lipids are produced—presumably by using other sources of NADPH. [16]

1.4. Overview of microbial cultivation methods

1.5. Genome-scale metabolic modeling

Genome-scale metabolic models (GEMs) are comprehensive descriptions of all known metabolic reactions of a organism, which are derived from the organism's available genome sequence [17]. GEMs account for genes, proteins, and biochemical reactions, which enable systematic analysis of metabolism, where typically the objective is to achieve a global picture of possible flux patterns [17][18].

A microorganism's phenotype can be described by its pattern of metabolic fluxes [17]. Metabolic flux is the rate of turnover of molecules through a metabolic pathway. Flux is regulated by the enzymes involved in this pathway. Within cells, regulation of flux is vital for all metabolic pathways to regulate the pathway's activity under different conditions. [19] In biotechnology the aim is often to increase the capacity of specific fluxes. For this, metabolic engineering methods have been developed and many of these rely on balancing intracellular metabolites using GEMs that in combination with appropriate objective functions and constraints can be used to predict potential gene targets for obtaining a preferred flux distribution. [17]

GEMs can be a powerful and helpful tool in metabolic studies, if their predictive power is good [10]. The use of GEMs enables the integration of omics data and experimental metabolic fluxes to generate a holistic view of metabolism in different physiological states, allowing a greater understanding of cellular physiology and providing valuable information for metabolic engineering and the development of microbial factories as a whole. [20] The number of organisms for which metabolic reconstructions have been created is increasing at a pace similar to whole genome sequencing. [21]

The metabolic reconstruction process usually is very labor- and time consuming. For for well-studied, medium genome sized bacteria it can take around six months to reconstruct the model. The metabolic reconstruction of human metabolism can take up to two years and six people. The reconstruction process is often iterative, for example the reconstruction of metabolic network of *Escherichia coli* has been expanded and refined throughout the last 19 years. Despite growing experience and knowledge, it is still not possible to completely

automatically reconstruct high-quality metabolic networks which can be used as reliable predictive models. [21] (See appendix A for more detailed reconstruction process.)

Constraint-based modeling

Genome-scale metabolic models are widely used to calculate metabolic phenotypes, but as metabolism operates under countless of constraints, they rely on defining a set of constraints. They include constraints that remain unchanged as they come by the laws of physics (e.g. conservation of mass and energy, thermodynamics), and constraints that can vary by organism, environmental condition, the state of the cell (e.g. nutrient uptake rate, biomass composition), and constraints that can also change through evolution or changes in gene expression. [22]

A major limitation in the predictive power of conventional GEMs is that enzyme abundances and kinetics, which act as limitations on metabolic fluxes, are not taken into account. When considering the production of a metabolite of interest, these models typically make the assumption that the uptake rate of the carbon source limits production. This may be an oversimplification, as metabolic fluxes are limited by their corresponding enzyme levels. [23] The synthesis of enzymes is resource- and energy-expensive, their catalytic capacities are limited by their kinetics, what is more, the quantity of enzymes is space-constraint [22].

This means that an increase in the requirement of an enzyme or a pathway would be a trade-off for other functions. Experimental evidence indicates that for various organisms resource re-allocation could be an effective strategy in response to nutrient and growth shift, which demonstrates the biological significance of proteome constraints. [18] Applying such constraints in a metabolism model reduces simulated flux distributions to those that are most economic and also limits the phenotypes that the model can simulate. These both contribute to more realistic results. Such models have already found numerous applications in e.g. unraveling the underlying mechanisms for observed metabolic phenotypes and the prediction of strain optimization strategies. [22] This suggests that proteome constraints could be a valuable addition to GEMs to improve model predictions [18].

Integration of enzyme constraints and proteomics data into GEMs was first enabled by the GECKO toolbox, allowing the study of phenotypes constrained by protein limitations. GECKO is a method for enhancement of GEMs with Enzymatic Constraints using Kinetic and Omics data, developed in 2017. This method extends the classical FBA approach by incorporating a detailed description of the enzyme demands for the metabolic reactions in a network, accounting for all types of enzyme-reaction relations, including isoenzymes, promiscuous enzymes and enzymatic complexes. Moreover, GECKO enables direct integration of proteomics abundance data, if available, as constraints for individual protein demands, represented as enzyme usage pseudo-reactions, whilst all the unmeasured enzymes in the network are constrained by a pool of remaining protein mass. [24]

Every metabolic reaction flux has a biological constraint that is equal to the enzyme's concentration multiplied by its turnover number (k_{cat}). Enzyme constraint is defined as the maximum rate of enzymatic reaction (v_{max}) that the metabolic flux cannot exceed. Enzyme-

constrained GEMs thereby ensure that each metabolic flux does not exceed its biological maximum capacity, equal to the product of the enzyme's abundance and turnover number [23].

Phenomenological constraint is imposed on metabolic flux (v; mmol/gDCW/h), formulated as enzyme kinetics: $v <= E \cdot k_{cat}$, where E is protein abundance (mmol/gDCW) and k_{cat} is the enzyme's turnover number (1/s), provided with an upper limit on individual or total protein abundances. The integration of enzymatic constraints in *S. cerevisiae* has significantly improved phenotype prediction. [23]

The reconstructed genome-scale networks can be converted into mathematical stoichiometric matrices, where rows represent metabolites and columns represent individual reactions. This matrix will determine the solution space of the metabolic network. GEMs are constrained by (1) the stoichiometry of the network; (2) preset upper and lower boundaries for selected reactions; and (3) the assumption of a steady state. [17] Overview of a reconstruction of a GEMs is shown in the figure 1.1.

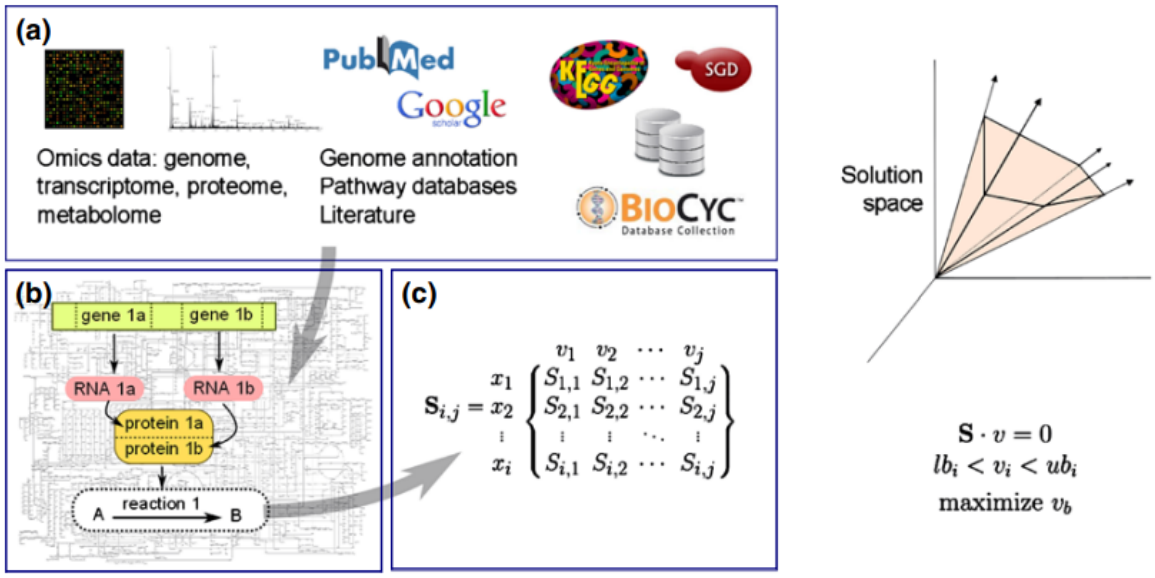

Figure 1.1: Reconstruction of a GEM. (a) The genome annotation is used to reconstruct the draft. (b) Gene-protein-reaction relationships are defined for the metabolic model. (c). A solution space is defined from the constraints applied to the model. Figure is from article [17].

Flux balance analysis

Flux balance analysis (FBA) is a mathematical approach for analyzing the flow of metabolites through a metabolic network. It is a widely used approach for studying biochemical networks, in particular the genome-scale metabolic network reconstructions. FBA calculates the flow of metabolites through this metabolic network, thereby making it possible to predict the growth rate of an organism or the rate of production of a biotechnologically important metabolite. [25]

The first step in FBA is to mathematically represent metabolic reactions. The core feature of this representation is a tabulation, in the form of a numerical matrix ($S = m \cdot n$), of the

stoichiometric coefficients of each reaction. These stoichiometries impose constraints on the flow of metabolites through the network. Constraints such as these lie at the heart of FBA, differentiating the approach from theory-based models based on biophysical equations that require many difficult-to-measure kinetic parameters. Every row of this matrix represents one unique compound (for a system with m compounds) and every column represents one reaction (n reactions).

The entries in each column are the stoichiometric coefficients of the metabolites participating in a reaction. There is a negative coefficient for every metabolite consumed, and a positive coefficient for every metabolite that is produced. A stoichiometric coefficient of zero is used for every metabolite that does not participate in a particular reaction. S is a sparse matrix since most biochemical reactions involve only a few different metabolites. The flux through all of the reactions in a network is represented by the vector v, which has a length of n. The concentrations of all metabolites are represented by the vector x, with length m. The system of mass balance equations at steady state ($dx/dt = 0$) is: $Sv = 0$. [25]

Steady state assumption presumes that all the flux rates and metabolite concentrations are constant over time. In experiments this state can be reached in chemostat cultivation and, in theory, during controlled batch logarithmic growth phase. [17]

Any v that satisfies this equation is said to be in the null space of S. In any realistic large-scale metabolic model, there are more reactions than there are compounds ($n > m$). In other words, there are more unknown variables than equations, so there is no unique solution to this system of equations. [25]

Although constraints define a range of solutions, it is still possible to identify and analyze single points within the solution space. For example, we may be interested in identifying which point corresponds to the maximum growth rate or to maximum ATP production of an organism, given its particular set of constraints. FBA is one method for identifying such optimal points within a constrained space (see figure below).

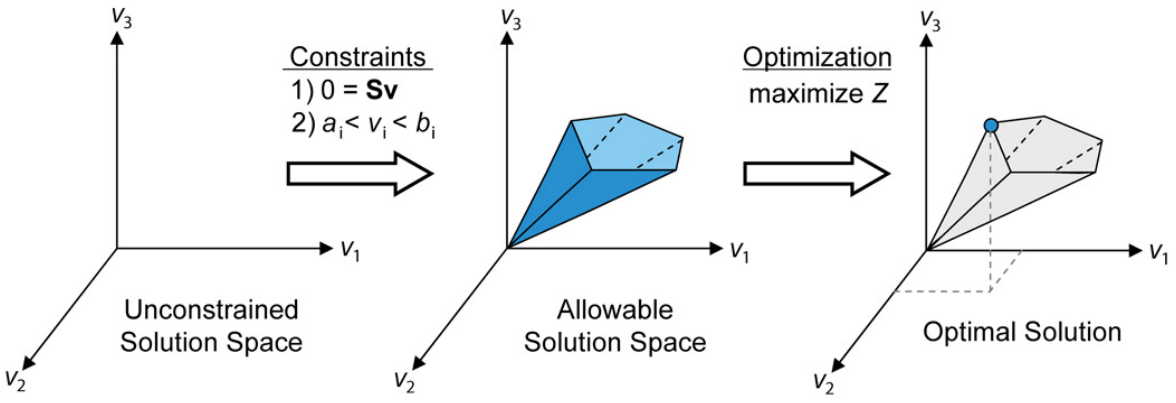

Figure 1.2: Conceptual basis of constraint-based modeling and FBA. The figure is from [25].

With no constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints imposed by the stoichiometric matrix S (1) and

capacity constraints imposed by the lower and upper bounds (a_i and b_i) (2) are applied to a network, it defines an allowable solution space. The network may acquire any flux distribution within this space, but points outside this space are denied by the constraints.

Through optimization of an objective function, FBA can identify a single optimal flux distribution that lies on the edge of the allowable solution space. Optimization of such a system is accomplished by linear programming. FBA can thus be defined as the use of linear programming to solve the equation $S \cdot v = 0$ given a set of upper and lower bounds on v and a linear combination of fluxes as an objective function. The output of FBA is a particular flux distribution, v, which maximizes or minimizes the objective function. [25]

The next step in FBA is to define a biological objective that is relevant to the problem being studied. In the case of predicting growth, the objective is biomass production, the rate at which metabolic compounds are converted into biomass constituents such as nucleic acids, proteins, and lipids. Mathematically, the objective is represented by an objective function that indicates how much each reaction contributes to the phenotype. A biomass reaction that drains precursor metabolites from the system at their relative stoichiometries to simulate biomass production is selected by the objective function in order to predict growth rates. This reaction is scaled so that the flux through it is equal to the exponential growth rate (μ) of the organism. [25]

The most commonly used objective function in FBA is maximization of the specific growth rate, ATP generation or a certain product formation. FBA is often used for estimating the biotechnological potential of microorganisms and pinpoint genetic manipulations that could improve the performance of a cell. The main applications of FBA are:
(1) Instructions for metabolic engineering purposes;
(2) Biological interpretation and discovery through contextualizing high-throughput data;
(3) Development of a computational framework;
(4) Evolutionary elucidation;
(5) Description of multispecies communities. [17]

Many computational linear programming algorithms exist, and they can very quickly identify optimal solutions to large systems of equations. The COBRA Toolbox [26] is a freely available Matlab toolbox for performing these calculations. Models for the COBRA Toolbox are saved in the Systems Biology Markup Language (SBML) [27] format.

However, FBA has limitations. Because it does not use kinetic parameters, it cannot predict metabolite concentrations. It is also only suitable for determining fluxes at steady state. Except in some modified forms, FBA does not account for regulatory effects such as activation of enzymes by protein kinases or regulation of gene expression, so its predictions may not always be accurate. [25]

Limitations

GEMs and kinetic models both have their advantages and drawbacks. GEMs are comprehensive

but are dependent on a pseudosteady state and in their current form do not take regulation, such as gene-protein and protein-protein level interactions, allosteric regulation or regulation at post-translational level, into account. At the same time, kinetic models require parameter values that are difficult to estimate at the global scale. As neither of them can fully replace the other, there has been a considerable effort on combining these two approaches in a singular model or applying them in succession. [17]

GEMs in combination with global datasets are invaluable for detecting the metabolic bottlenecks, but due to lack of kinetic data about enzyme activities or regulation mechanisms they are limited in their predictive power. Applying kinetic models on metabolic bottlenecks previously detected with GEMs can help to understand the regulation or kinetics of these specific enzymatic steps, as one would not need model parameters for the whole system. [17]

1.6. Genome-scale metabolic models of *Rhodotorula toruloides*

rhto-GEM

The first genome-scale model of *R. toruloides* metabolism named rhto-GEM was presented in 2019 by Tiukova et al. The model includes 4869 genes, 897 reactions, and 3334 metabolites. The model rhto-GEM, is based on the genome sequence of *R. toruloides* strain NP11. To reconstruct those parts of metabolism that are relatively conserved between fungal species, the well-curated GEM of Saccharomyces cerevisiae was taken as template model (yeast-GEM version 8.2.0, 16), while orthologous genes were identified via bi-directional BLASTP against the S. cerevisiae S288c reference genome. [12]

To transform this functional draft model to the first version of the *R. toruloides* GEM, additional manual curation was performed where remaining template-derived genes were replaced by their *R. toruloides* ortholog where possible and otherwise deprecated. Lipid metabolism of *R. toruloides* was described applying the SLIMEr formalism as previously described for *S. cerevisiae*, which allows direct integration of lipid class and acyl chain experimental distribution data. As the acyl chain distribution of *R. toruloides* is different from *S. cerevisiae*, e.g. the presence of C18:2 and C18:3, this required extensive manual curation of the SLIMEr reactions. *R. toruloides* specific reactions and pathways, such as carotene and torulene biosynthesis, synthesis and degradation of C18:2 and C18:3 fatty acids, and mitochondrial beta-oxidation were subsequently manually curated. [12]

The model incorporates knowledge obtained from genomics and proteomics data generated for *R. toruloides* and was validated using cultivation data. Simulations of rhto-GEM on various carbon sources demonstrated good agreement with experimentally reported growth rates. Analysis of model allowed to identify potential genetic engineering strategies for enhanced lipid production. Some of these genetic targets were found to agree with published experimental studies. As such, rhto-GEM emerges as a valuable tool for future analysis of oleaginous and lipid metabolism. [12]

iRhto1108

During the same year as rhto-GEM by Tiukova et al. [12] was presented, Dinh et al. [28] presented another *R. torluoides* genome-scale metabolic model named iRhto1108, collecting and organizing functional genomics data [29] and prior knowledge. The model is based on the strain IFO0880's metabolic network and it accounts for 2204 reactions, 1985 metabolites and 1108 genes.

R. toruloides has been target of significant research efforts including genome (re)sequencing, functional genomics analyses, differential omics characterization, determination of macromolecular composition, and growth kinetics in a continuous culture. These experiments have ushered an improved understanding of *R. toruloides* metabolism and provided the basis for the reconstruction of a metabolic model with genome-wide coverage. [28]

The authors integrated and supplemented the current knowledge with in-house generated biomass composition and experimental measurements pertaining to the organism's metabolic capabilities. iRhto1108 model integrates yeast biochemistry information from (i) previously built genome-scale models (*S. cerevisiae* yeast 7.6 [30], (ii) KBase fungal models [31]), and (iii) *R. toruloides* specific information extracted from the primary literature [29][32][33] or generated herein. Organism-specific macromolecular composition and ATP maintenance requirements were experimentally measured for two separate growth conditions: (i) carbon and (ii) nitrogen limitations. Overall, iRhto1108 reproduced *R. toruloides* utilization capabilities for 18 alternate substrates, matched measured wild-type growth yield, and recapitulated the viability of 772 out of 819 deletion mutants. [28]

Essential cellular metabolism and growth capability of the model were validated extensively with experimental results, including gene essentiality [29] and growth data. iRhto1108 was also able to recapitulate experimentally-observed lipid accumulation phenotypes. iRhto1108 can comprehensively capture *R. toruloides* metabolism and provide meaningful predictions that were validated with experimental data including suggestion of genetic perturbations leading to triacylglycerol overproducing strains. Overall, iRhto1108 has undergone a detailed range of testing and validation studies promising to aid in future investigations of *R. toruloides*. [28]

Rt_IFO0880

In 2021 Kim et al. performed multi-omics analysis of lignocellulosic carbon utilization in *R. toruloides* and reconstructed the genome-scale metabolic network named Rt_IFO0880. The curated metabolic reconstruction consisted of 1106 genes, 1934 reactions, and 2010 metabolites (1246 unique metabolites) in nine compartments. [34]

High-quality metabolic network models for model organisms and orthologous protein mapping were used to build a draft metabolic network reconstruction. The reconstruction was manually curated to build a metabolic model using functional annotation and multi-omics data including transcriptomics, proteomics, metabolomics, and RB-TDNA sequencing. [34]

There were many incorrect reactions involved in fatty acid biosynthesis and beta-oxidation, especially for unsaturated fatty acids. The reactions and genes in the central metabolic pathways were manually checked for their co-factor usage and localization. Authors used multi-omics and other experimental measurement to update the biomass reaction. The DNA composition was updated using the genome sequence, RNA composition was updated using transcriptomics data, amino acid composition was updated using proteomics data, and lipid composition was updated using fatty acid methyl ester analysis. [34]

Authors performed a genome-scale evaluation and iteratively improved the model using high-throughput growth phenotyping and functional genomics. They tested the developed metabolic model's capability to predict growth on different carbon, nitrogen, sulfur, and phosphate sources. The model was further refined to resolve the inconsistencies and several genes with erroneous ortholog mapping were removed from the model. [34]

The metabolic network model was validated against high-throughput growth phenotypes in 213 growth conditions and conditional gene essentiality in 27 growth conditions with high prediction accuracies, significantly expanding the breadth and depth of metabolic coverage from previously published models ([28];[12]). [34]

Authors believe that the developed metabolic network for *R. toruloides* is most complete and accurate to date, and the multi-omics data and metabolic model presented in this study will be useful for studying and engineering *R. toruloides* for lignocellulosic biomass conversion. [34]

Rt_IFO0880_LEBp2023

In a doctoral thesis focused on the carotenoid production of *R. torluoides* [20] the author compared the four existing GEMs of *R. torluoides* (the ones mentioned before and also rhto-GEM_BioEng, which is a version of the rtho-GEM with carotenoids integrated into the biomass composition and an alternative xylose assimilation pathway [35]) and selected the model Rt_IFO0880 as the best one from the four for further enhancement.

The author chose the model Rt_IFO0880, because it was the one that needed least improvement and its description of metabolism was the one that best represented the pathways involved in the biosynthesis of carotenoids in *R. toruloides*, according to the study carried out by [20]. Furthermore, its accuracy and sensitivity are higher than those of the other models, meaning that the Rt_IFO0880 is more accurate in predicting the essentiality of genes compared to the other models. [20]

The following changes were made by [20]: (i) The reaction and gene-protein-reaction relation (GPR) corresponding to cytosolic malate dehydrogenase (cMDH) were added. The lower and upper flow limits of the xylokinase (produces xylulose-5-phosphate) and phytoene dehydrogenase (enzyme in carotenoid byosynthesis) enzymes were equalized to zero to reflect, respectively, the absence of detectable activity of the former and the deletion of the gene encoding the latter; (iii) A phytoene transport reaction was created for the lipid body compartment to release the flow through the synthesis of this compound; (iv) The lower limit of the cytosol-to-phytoene

NADP+ transport reaction was changed to a non-zero value (-1000 mmol/gMCS.h) to confer reversibility to this reaction. The model updated in this way, called Rt_IFO0880_LEBp2023, was validated against the experimental data. [20]

Unlike its predecessor, GEM Rt_IFO0880_LEBp2023 provides for the use of ACL in all maximum theoretical yield of phytoenes scenarios. This enzyme is extremely abundant in *R. toruloides* [36] so its prediction by GEMs would be expected. However, from other models only the iRhtoC predicted the use of this enzyme. The modification that seems to have allowed the Rt_IFO0880_LEBp2023 to predict the use of ACL was the addition of cytosolic malate dehydrogenase. This enzyme allows the conversion of oxaloacetate produced by ACL into malate, which in turn is transported into the mitochondria into an antiporte with mitochondrial citrate, which is taken to the cytosol to be substrate of ACL. This pathway has recently been described as an alternative to the classic tricarboxylic acid pathway and has been called the "non-canonical TCA cycle" [37]. In addition to the Rt_IFO0880_LEBp2023 model, the iRhtoC model was also able to predict it. The inability of the other models to predict the non-canonical TCA cycle can be explained by the absence of cMDH (Rt_IFO0880) and citrate-malate antiporte between the cytosol and the mitochondria (rhtoGEM and rhtoGEM_BioEng). [20]

The update of the chosen model resulted in the model Rt_IFO0880_LEBp2023, presented a better fit to the experimental data than the others. In addition, this GEM was able to predict the use of the ACL enzyme, an enzyme present in *R. toruloides* according to omics studies [36] but which use was not predicted in three of the four models. The prediction of the use of ACL also allowed the observation that *R. toruloides* possibly uses a non-canonical TCA cycle to avoid the production of CO2 in the generation of mitochondrial NADH, something previously observed for this yeast. [20]

The updated model (Rt_IFO0880_LEBp2023) showed a high capacity to adjust to experimental data and to predict the use of biologically relevant enzymes for this yeast, such as ACL. Despite the few updates, the best predictions of the Rt_IFO0880_LEBp2023 when compared to the other GEMs of *R. toruloides* accredit it as the most faithful to the metabolism of this yeast to date. GEM Rt_IFO0880_LEBp2023, generated in this work, is the best representative of the metabolism of *R. toruloides*, because it describes more precisely its xylose assimilation pathways, the conversion of acetyl-P into acetyl-CoA and the absence of HMG-CoA transport from the mitochondria to the cytosol, (characteristics inherited from its predecessor GEM) and predicts the use of ACL and its involvement in a non-canonical TCA cycle. [20]

Comparison of the models

The first model is based on the genome of the NP11 strain, while the other three are based on that of the IFO0880 strain. Both are haploid strains [38]; [36]. The similarity between the genomes of these strains is approximately 95% [39].

The differences between the models cannot be explained by differences in the genomes of the NP11 and IFO0880 strains, since they exist between the iRhtoC and Rt_IFO0880 models,

which are based on the same genome annotation as the latter. In addition, the annotations in the genome of *R. toruloides* are still flawed because they are based on genes and metabolic pathways of conventional yeasts. [20]

In term of genome coverage, iRhto1108 contains slightly more genes than rhto-GEM (i.e., 1108 vs. 926 genes) or after the removal of blocked reactions (i.e., 806 vs. 624 genes). [28]

ACL, an enzyme found to be present in *R. toruloides*, was predicted only by the models iRhto1108 and Rt_IFO0880_LEBp2023.

2. Aims of the thesis

The objective of the study is investigating *R. toruloides* lipogenesis focused central carbon metabolism based on a comparison between the four GEMs of *R. toruloides*. For that FBA is carried out with all of the models using different objective functions and constraints. The intracellular fluxes of PPP and TCA enzymes are compared between the models to understand the main sources of lipid precursors acetyl-CoA and NADPH within these models.

3. Methods

Genome-scale metabolic models of *R. toruloides* (JSON files) were obtained from supplemental files from publications [12, 28, 34] and the fourth model [20], made by a group member, was obtained directly from him.

With all these models, simulations were performed using the COBRApy package (version 0.29.0) [40] in Python (version 3.11.4). Throughout the process, metabolic flux patterns were predicted using flux balance analysis [25] from COBRApy package with the Gurobi mathematical optimization solver (version 11.0.0, Gurobi Optimization Inc.).

All scripts are available in Github repository upon request sent to maive.hanni@gmail.com.

3.1. Selecting experimental data

In the simulations the experimental steady-state cultivation data of *R. toruloides* strain IFO0880 in 1 L bioreactor was used. The experiment was done by our group and the results have not yet been published. The cultivation process is described in [10]. All rates are expressed in $mmol/gDW/h$, and the biomass specific growth rate is expressed as $1/h$. The data is shown in the table 3.1.

Table 3.1: Continuous cultivation results from lab experiments

Biomass growth rate	**Glucose uptake**	**CO_2**	**O_2**	**Glycerol**
0.049	0.476	1.171	1.083	almost nothing
0.100	1.114	2.521	2.521	
0.151	1.648	3.854	3.851	
0.203	2.305	5.834	4.352	
0.25	-	-	-	
0.301	3.1	7.415	6.327	

3.2. Biomass integration to the model

3.3. Flux balance analysis and sampling of the solution space

Growth phenotypes were obtained using FBA with several objective functions. First, before having lab data for glucose uptake, the objective function was set to glucose uptake maximisation and the biomass pseudoreaction, whose flux is equivalent to the growth rate ($h^{1/h}$), was constrained at five different rates (0.05, 0.10, 0.15, 0.2, 0.25, 0.30) (as glucose

uptake flux is notated in the model with negative sign, setting the objective to glucose uptake maximisation corresponds to finding the solution space with minimal absolute value of glucose uptake flux). Model solutions were calculated on five different growth rates and Python's Pandas package was used for constructing dataframes of extracellular and the fluxes of central carbon metabolism enzymes. Matplotlib was used for data visualization.

Once lab data for steady-state cultivation of *R. toruloides* strain IFO0880 was available, glucose uptake values on five growth rates were used in simulations for model constraining. On carbon limitation the objective function was set to biomass maximisation. Again dataframes of the fluxes of interest were made and visualized.

Metabolic flux patterns were further investigated using the minimisation of non-growth associated maintenance (NGAM) as objective function with constraining both glucose uptake and growth rate to see if that changes the flux patterns. As the model overestimates glucose uptake need for a specific growth rate, it was not possible to constrain both glucose uptake and growth rate to the values obtained in lab - the solution was infeasible. Because of that, in this simulation glucose uptake was constrained to the values obtained in lab, but growth rate was constrained to the growth rate values obtained in previous simulation when solution was optimized for maximum growth rate on specified glucose uptake.

For comparison of cofactor balances between the models, pie plots were made to visualize the production and consumption of NADPH, NADH and ATP. Cofactor balances were simulated with biomass maximisation and NGAM minimisation as the objective function, as the lab data was available.

Finally, for some models the central carbon metabolism fluxes were visualized with Escher [41]. On Escher platform it is possible to create metabolic maps based on the reactions of a given GEM, which can then be fed with the simulated flux distributions to have a better overview of used metabolic pathways.

4. Results

Rhodotorula toruloides can naturally accumulate high amounts of lipids, but the metabolic mechanism that make this possible and differentiate *R. toruloides* from non-oleaginous yeast are not fully understood. Biosynthesis of the main lipid precursors acetyl-CoA and NADPH takes place in the central carbon metabolism. A better understanding of which metabolic pathways are used in production of these precursors and thus contribute to lipid accumulation, would aid in designing better metabolic engineering strategies for increasing lipid production.

Genome-scale metabolic models contain all known biochemical reactions of the specific organism and allow the calculation of metabolic fluxes, which represent the activity of metabolic pathways under specified conditions. This makes GEMs important tool for studying metabolism, but it is important that the predictive power of the GEM is good. For *R. toruloides* several genome-scale metabolic models are available, but so far comprehensive overview of simulations focused on central carbon metabolism with these models has not been presented. This work will help in the future to design *R. toruloides* as the leading microbial cell factory for production of microbial oils.

4.1. Metabolic map

4.2. Intracellular fluxes and supply of NADPH

Four genome-scale metabolic models of *Rhodotorula toruloides* - rhto-GEM, iRhto1108, Rt_IFO0880 and Rt_IFO0880_LEBp2023 - were used for simulating this yeast's metabolic flux distribution using flux balance analysis. Different objective functions, maximisation of biomass and minimisation of NGAM, were tested under carbon limitation. To see the flux patterns over different biomass growth rates, the solutions were constrained over five glucose uptakes - $0.476, 1.114, 1.648, 2.305, 3.1 (mmol/gDW/h)$ - that were measured in lab experiments.

Fluxes through notable phosphoketolase and ATP-citrate lyase pathways, that are known to be important for production of lipid precursors in olegeanous yeast, were in the center of interest and are distinguished in intracellular flux figures from other enzymes using dashed line. What is more, lipid synthesis demands high amounts of NADPH, but only a few enzymes can generate it. Pentose phosphate pathway and malic enzyme have been proposed as the main candidate enzymes for recycling NADPH in *R. toruloides* [42]. Main intracellular fluxes of central carbon metabolism and production and consumption of NADPH were compared between the simulations of four models.

For all simulations exchange fluxes were also observed for evaluating model prediction compared to experimental data. For easier comparison of model predicted exchange fluxes with experimentally measured exchange fluxes, the experimental data is illustrated on a graph 4.1.

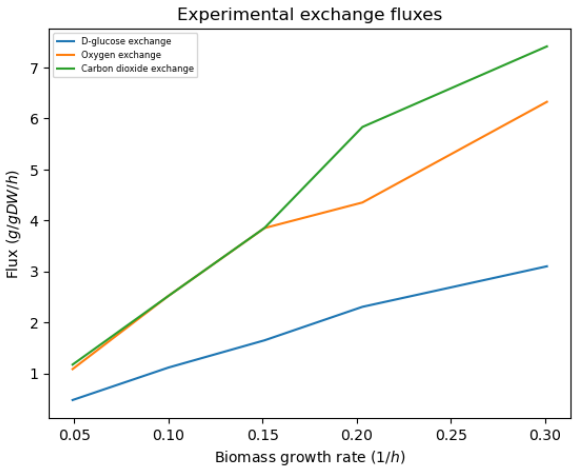

Figure 4.1: Experimental exchange fluxes in *R. toruloides* strain IFO0880 over five growth rates.

4.2.1. Biomass maximisation as objective function

In first simulations biomass maximisation was used as objective function and the calculations were made under carbon limitation constraining glucose uptake. Simulation results with all four models are shown in the figures below. All figures show the fluxes over growth rates from 0.05 - 0.25 $(1/h)$. Exchange fluxes graphs show the exchange (uptake or secretion) rate of glucose, oxygen, ammonium, sulphate, phosphate and carbon dioxide. Intracellular fluxes graphs show the fluxes of NGAM, glucose uptake, glucose 6-phosphate dehydrogenase (oxPPP), transketolase 1, transaldolase, transketolase 2, fructose-biphosphate aldolase, pyruvate decarboxylase, pyruvate dehydrogenase, phosphoketolase and ATP-citrate lyase.

These simulations, with biomass maximisation as objective function under carbon limitation, showed that the models prefer different pathways for production of acetyl-CoA. The models rhto-GEM and Rt_IFO0880 produced acetyl-CoA through phosphoketolase, whereas models iRhtoC and the modified model of Rt_IFO0880 (Rt_IFO0880_LEBp2023) produced acetyl-CoA using ACL. NADPH sources also differ. Each simulatio figures and results are described in more detail below.

rhto-GEM

The simulation results with rhto-GEM model for exchange and intracellular fluxes are shown in figure 4.2 and fluxes of NADPH producing and consuming enzymes are shown in 4.3. Compared to the values obtained in lab experiments the model predicts higher exchange fluxes on a given growth rate. This is also a reason why it was chosen to constrain the simulations with lab glucose uptake data and model simulated growth rates, because otherwise the model

predicted exchange fluxes would be even higher. This result was expected as GEMs are known to overestimate fluxes per growth rate.

Model rhto-GEM predicts the use of phosphoketolase on all growth rates and no use of ATP-citrate lyase. For NADPH production and consumption there are no differences between lowest and highest rate. In both cases the model predicts that over 90% of the NADPH is produced by glucose 6-phosphate dehydrogenase (oxPPP) and phosphogluconate dehydrogenase. 6.3% is produced by methylenetetrahydrofolate dehydrogenase.

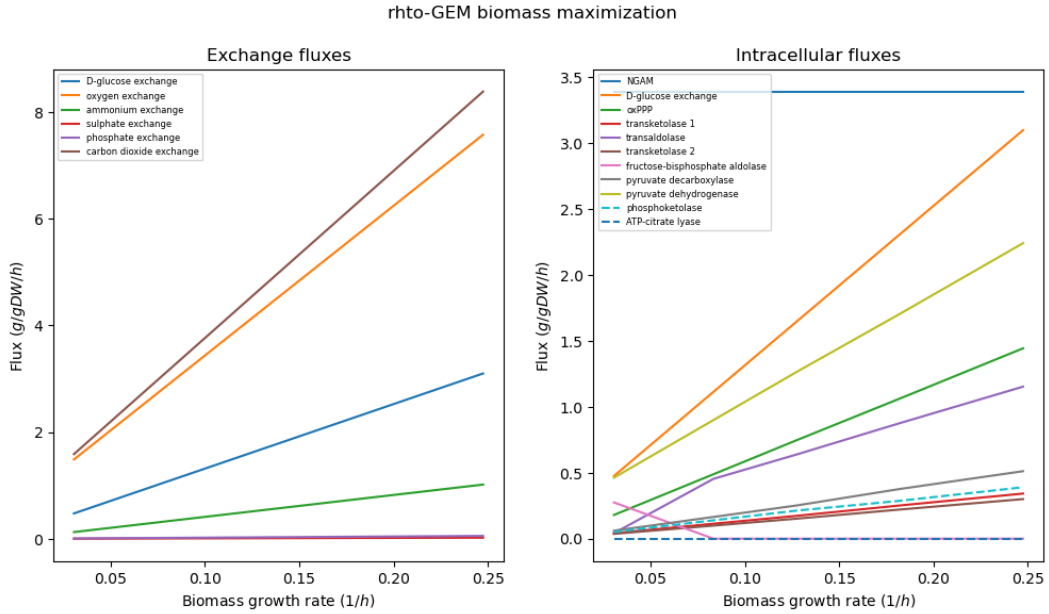

Figure 4.2: Exchange and intracellular fluxes in *R. toruloides* with model rhto-GEM optimized for biomass maximization and constrained over five glucose uptake rates.

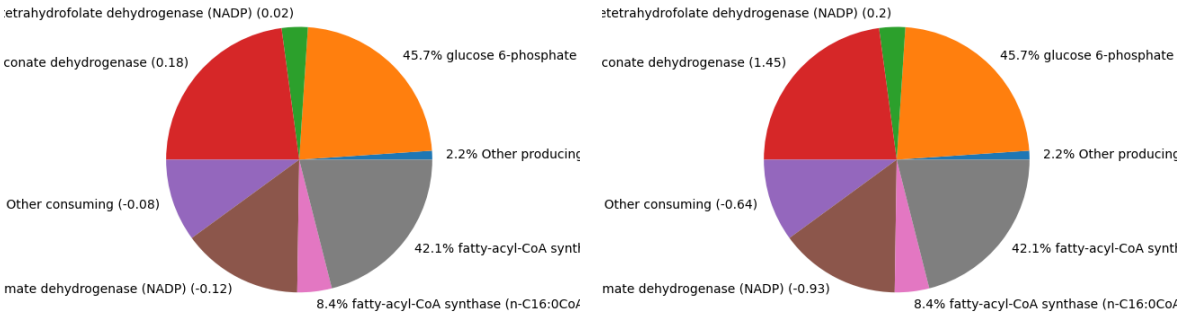

Figure 4.3: NADPH producing and consuming fluxes in *R. toruloides* with model rhto-GEM optimized for biomass maximization. Glucose uptake was constrained on the lowest (left) and highest (right) rate.

iRhtoC

The results with iRhtoC are shown in figures 4.4 and 4.5. This model predicts very similar exchange fluxes as the previous model. But predicted intracellular fluxes differ - iRhtoC

predicts the use of ATP-citrate lyase instead of phosphoketolase on all growth rates. With this model there is also no difference in NADPH production and consumption on highest and lowest rates. Almost 90% of NADPH is produced by glucose 6-phosphate dehydrogenase (oxPPP) and phosphogluconate dehydrogenase and 6.4% by malic enzyme.

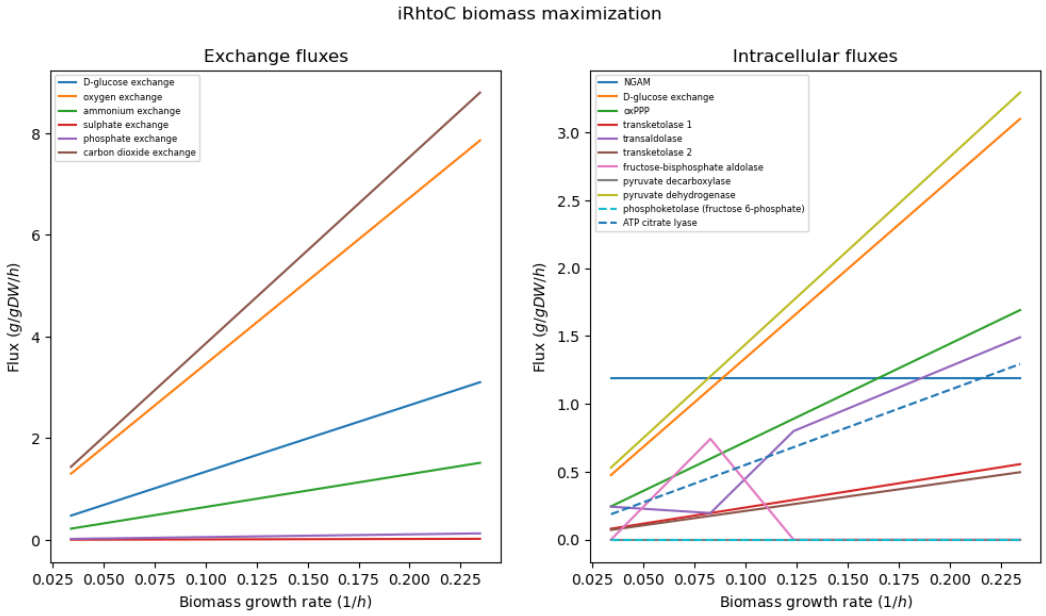

Figure 4.4: Exchange and intracellular fluxes in *R. toruloides* with model iRhtoC optimized for biomass maximization and constrained over five glucose uptake rates.

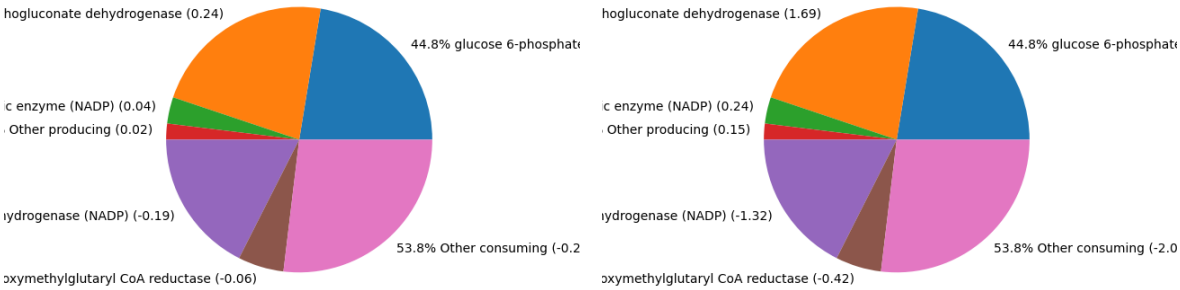

Figure 4.5: NADPH producing and consuming fluxes in *R. toruloides* with model iRhtoC. The model was optimized for biomass maximization. Glucose uptake was constrained on the lowest (left) and highest (right) rate.

Rt_IFO0880

Model Rt_IFO0880 fluxes are shown in 4.6 and 4.7. Exchange fluxes are similar with previous models. On all growth rates the use of phosphoketolase is predicted and no use of ACL. (This model has two phosphoketolases, fructose-6-phosphate phosphoketolase (FPK) and xylulose-5-phosphate phosphoketolase (XPK), but as models rhto-GEM and iRhtoC have one phosphoketolase, FPK and XPK fluxes have been summed together for easier comparison with other models.) In NADPH production and consumption on highest and lowest rates there are

not significant differences. Around 90% of NADPH is produced by alcohol dehydrogenase.

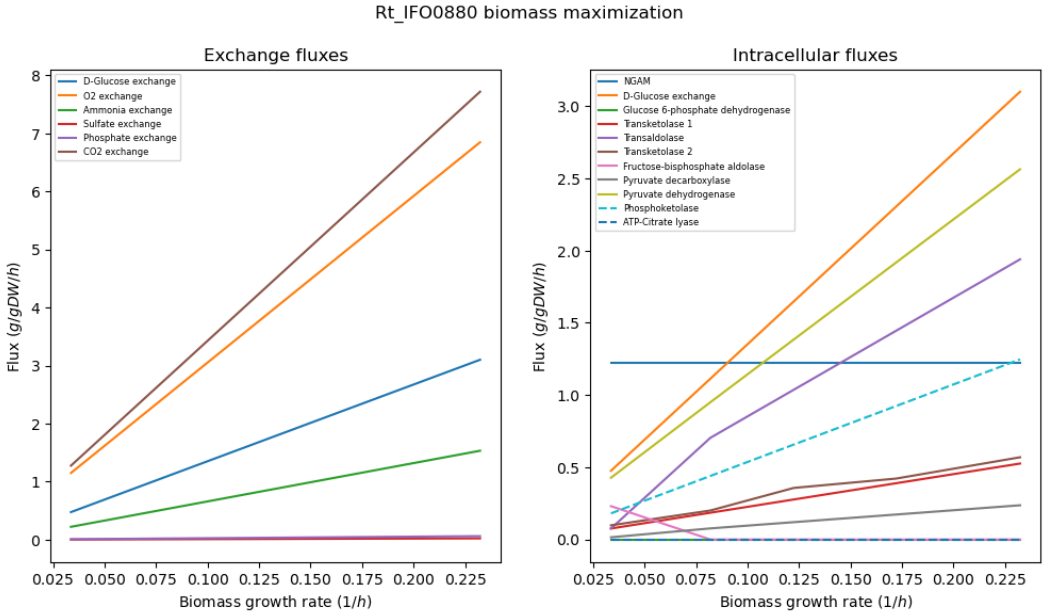

Figure 4.6: Exchange and intracellular fluxes in *R. toruloides* with model Rt_IFO0880 optimized for biomass maximization and constrained over five glucose uptake rates.

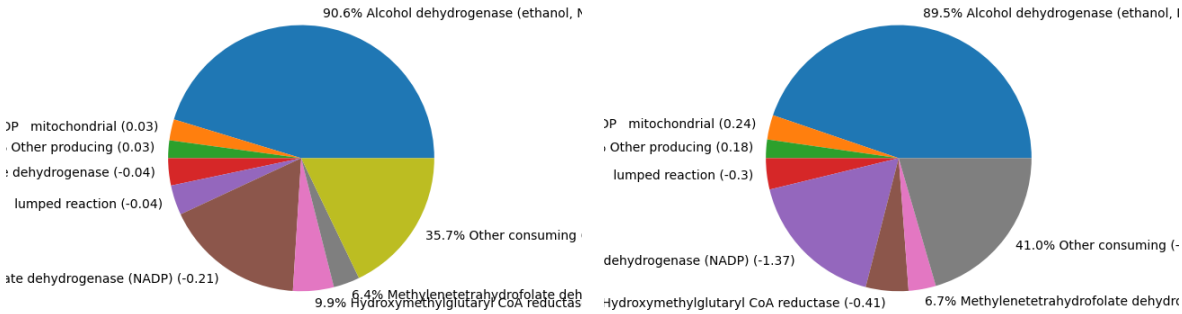

Figure 4.7: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880. The model was optimized for biomass maximization. Glucose uptake was constrained on the lowest (left) and highest (right) rate.

Rt_IFO0880_LEBp2023

In figures 4.8 and 4.9 simulation predictions with model Rt_IFO0880_LEBp2023 are shown. Exchange fluxes align with other models. On all growth rates the use of ACL is predicted instead of phosphoketolase. (As this model is an updated version of Rt_IFO0880, it also has two phosphoketolases - FPK and XPK, which fluxes have been summed together and represented as phosphoketolase.) On lowest growth rate NADPH is predicted to be produced mainly by aldehyde dehydrogenase and on highest rate by alcohol dehydrogenase.

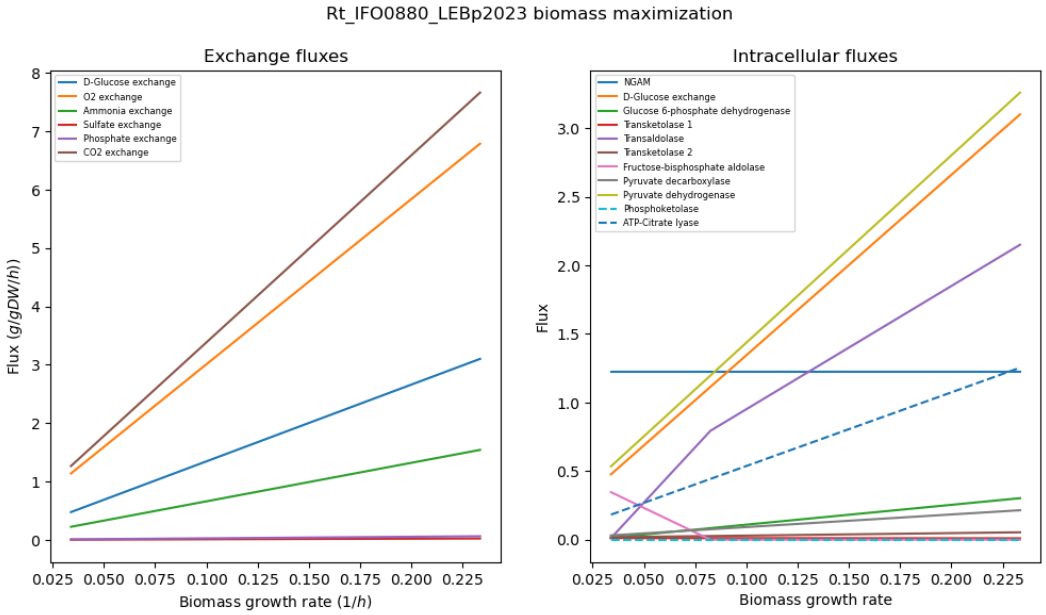

Figure 4.8: Exchange and intracellular fluxes in *R. toruloides* with model Rt_IFO0880_LEBp2023 optimized for biomass maximization and constrained over five glucose uptake rates.

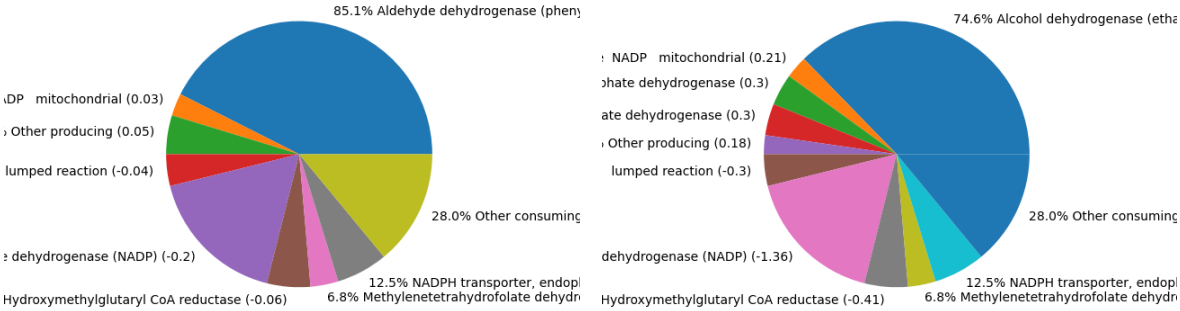

Figure 4.9: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880_LEBp2023. The model was optimized for biomass maximization. Glucose uptake was constrained on the lowest (left) and highest (right) rate.

4.2.2. NGAM minimisation as objective function

NGAM minimisation as objective function was investigated to see whether the use of different objective function gives distinct results. Again the simulations were made under carbon limitation over five glucose uptakes. This objective function also needed constraints on biomass growth rate because otherwise the simulation chooses zero as its flux. Growth rate was constrained to the values that each model predicted in the simulations optimized for biomass maximisation, because experimental growth rates together with experimental glucose uptakes were infeasible for the models.

For all simulations, the exchange fluxes are the same as with biomass maximisation as objective

function. The models that predicted phosphoketolase in previous simulations, predicted it again. Same is about prediction of ACL. For some models the production of NADPH differs with this objective function.

rhto-GEM

Model rhto-GEM optimized for NGAM minimisation (see figures 4.10 and 4.11) predicts similar fluxes as with previous objective function.

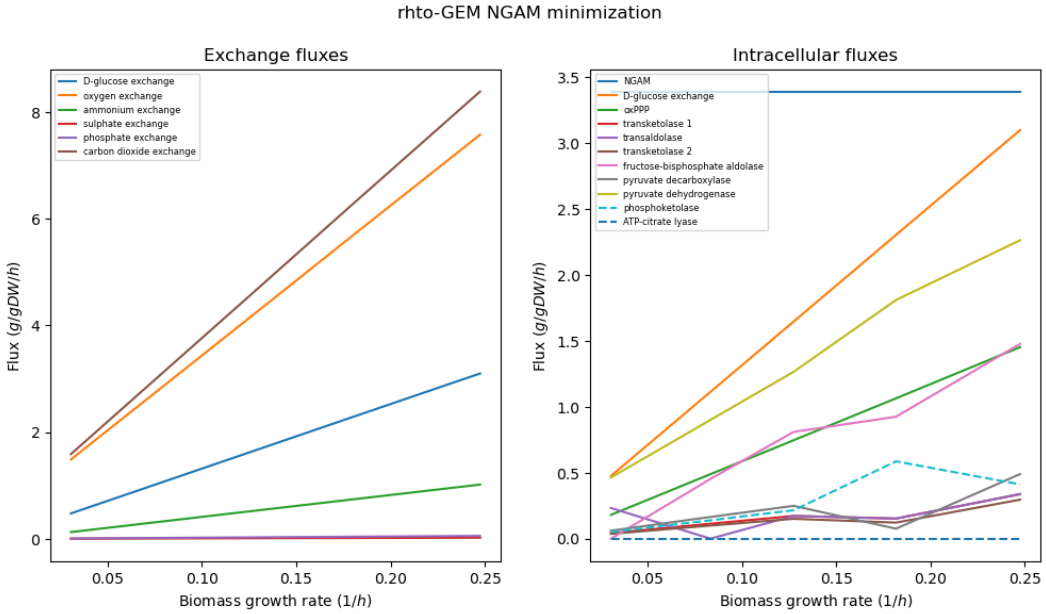

Figure 4.10: Exchange and intracellular fluxes in *R. toruloides* with model rhto-GEM optimized for NGAM minimization and constrained over five growth and glucose uptake rates.

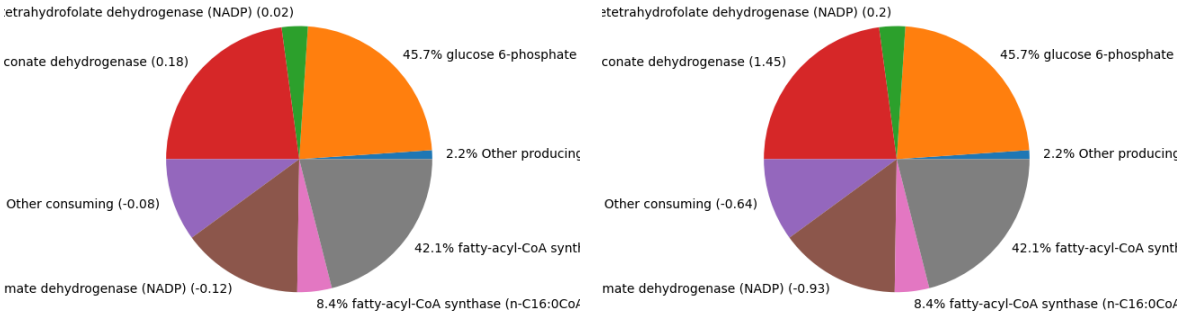

Figure 4.11: NADPH producing and consuming fluxes in *R. toruloides* with model rhto-GEM. The model was optimized for NGAM minimization. Glucose uptake with growth rate were constrained on the lowest (left) and highest (right) rate.

iRhtoC

iRhtoC optimized for NGAM minimisation (figures 4.12 and 4.13) also predicts similar fluxes as with previous objective function.

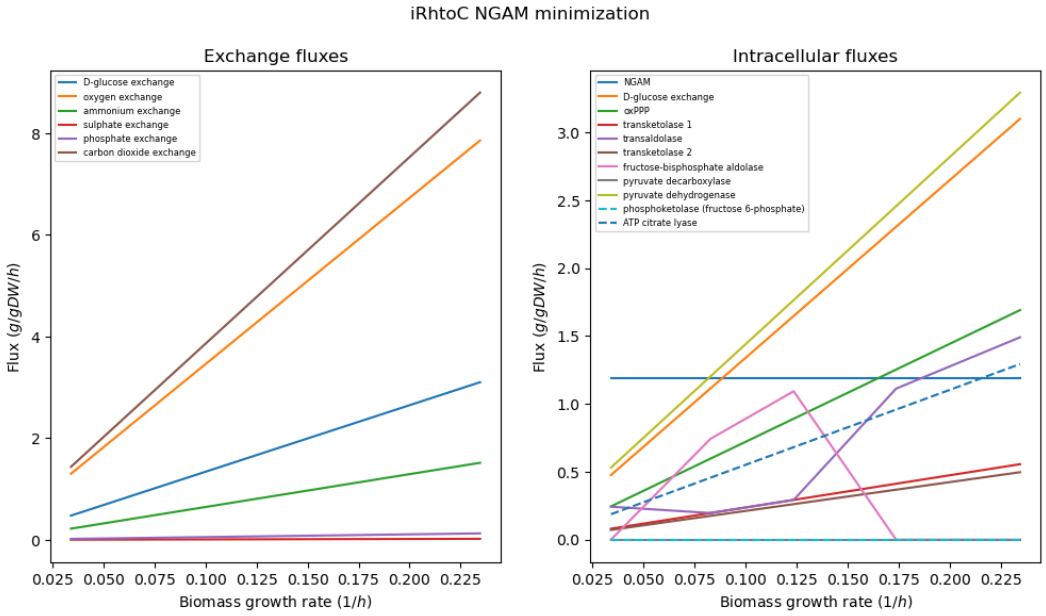

Figure 4.12: Exchange and intracellular fluxes in *R. toruloides* with model iRhtoC optimized for NGAM minimization and constrained over five growth and glucose uptake rates.

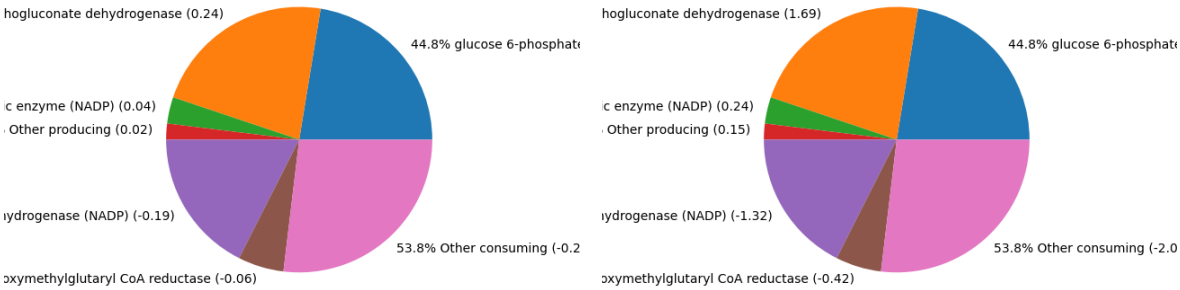

Figure 4.13: NADPH producing and consuming fluxes in *R. toruloides* with model iRhtoC. The model was optimized for NGAM minimization. Glucose uptake with growth rate were constrained on the lowest (left) and highest (right) rate.

Rt_IFO0880

Model Rt_IFO0880 optimized for NGAM minimisation (figures 4.14 and 4.15) predicts similar intracellular fluxes as before, but NADPH production differs. On the lowest growth rate it is predicted that most of NADPH is produced by homoserine dehydrogenase and on highest rate by alcohol dehydrogenase.

Rt_IFO0880_LEBp2023

Rt_IFO0880_LEBp2023 optimized for NGAM minimisation (figures 4.16 and 4.17) also predicts similar intracellular fluxes as before, but the production of NADPH differs. On the lowest growth rate it is predicted that most of NADPH is produced by aldehyde dehydrogenase

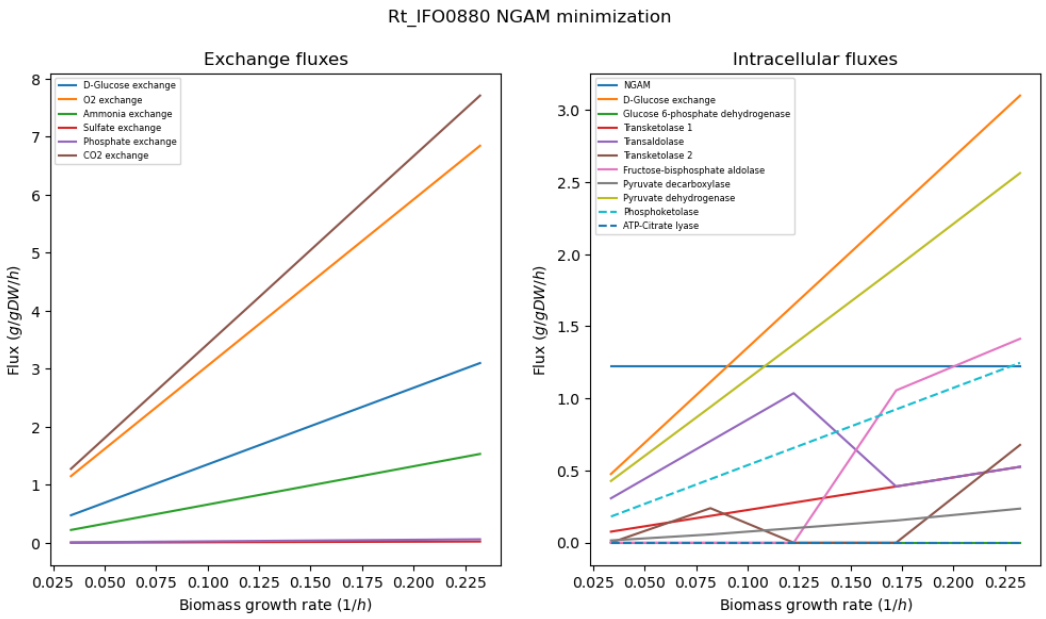

Figure 4.14: Exchange and intracellular fluxes in *R. toruloides* with model Rt_IFO0880 optimized for NGAM minimization and constrained over five growth and glucose uptake rates.

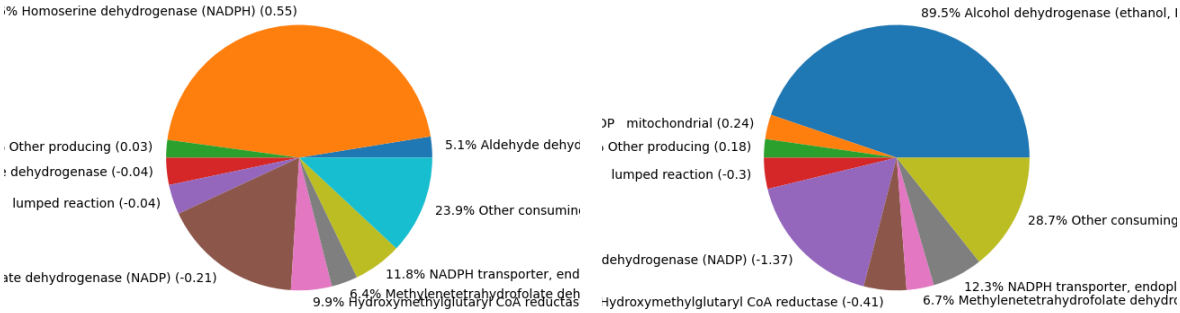

Figure 4.15: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880. The model was optimized for NGAM minimization. Glucose uptake with growth rate were constrained on the lowest (left) and highest (right) rate.

and on highest rate by homoserine dehydrogenase.

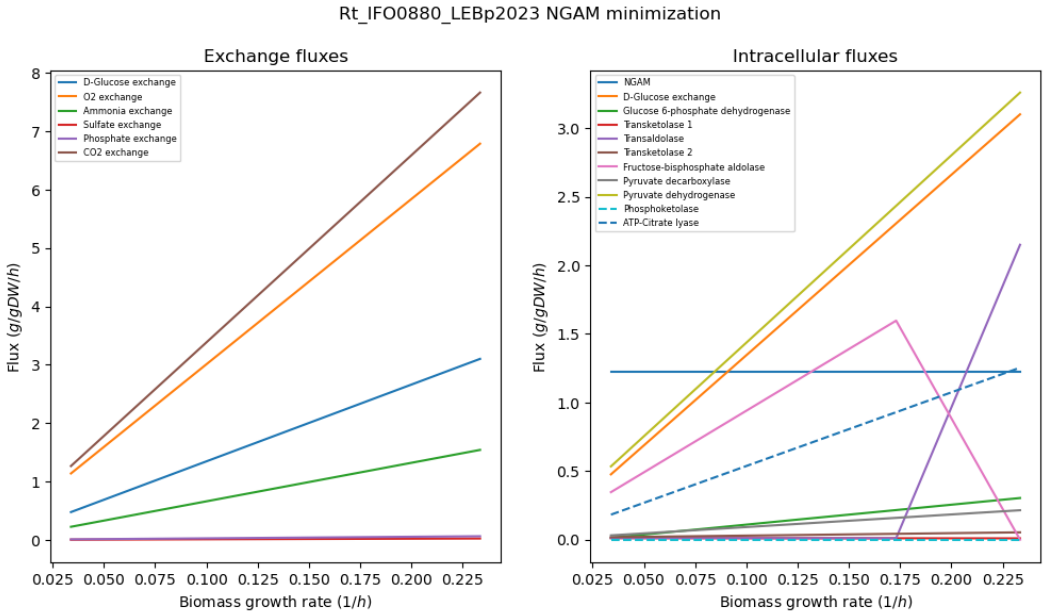

Figure 4.16: Exchange and intracellular fluxes in *R. toruloides* with model Rt_IFO0880_LEBp2023 optimized for NGAM minimization and constrained over five growth and glucose uptake rates.

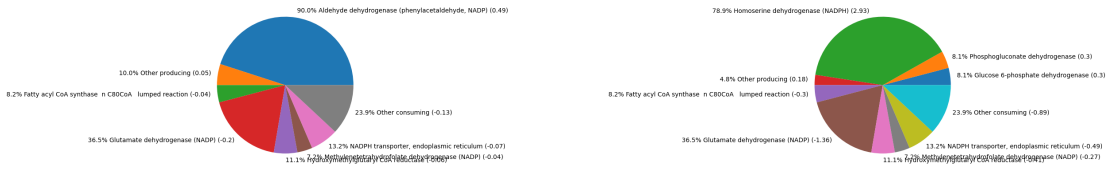

Figure 4.17: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880_LEBp2023. The model was optimized for NGAM minimization. Glucose uptake with growth rate were constrained on the lowest (left) and highest (right) rate.

5. Conclusion

In this work four genome-scale metabolic models of *Rhodotorula torluoides* were compared and different objective functions, including maximization of biomass and NGAM minimization, were tested. The fluxes through phosphoketolase and ACL pathways, that are found to be important for lipid production in all four models, were explored. This study gave more insight about how these model predict central carbon metabolism of *R. torluoides* than [20] that was based on phytoenes simulations.

The model Rt_IFO0880_LEBp2023 was found to predict *Rhodotorula torluoides* intracellular fluxes of lipogenesis precursors most accurately. This model has cytosolic malate dehydrogenase and thus is able to predict the flux of ACL, which from omics studies has been found to be present in *R. torluoides*.

Still, the autohor of the enhanced model of Rt_IFO0880 named Rt_IFO0880_LEBp2023, has pointed out that the improvements that resulted in the Rt_IFO0880_LEBp2023 model cover only a small portion of the metabolism of *R. torluoides*. "The differences are possibly the "tip of the iceberg" of many others and the differences between the existing GEMs of *R. toruloides* are significant and may lead to different understandings of this yeast's physiology if used alone" [20].

For better prediction of *R. toruloides* phenotype there is a need for a consensus GEM.

Summary

Acknowledgements

References

[1] Zuiderveen, E. A. R. et al. "The potential of emerging bio-based products to reduce environmental impacts". In: *Nature Communications* 14.1 (2023). DOI: 10.1038/s41467-023-43797-9.

[2] Research, E. C. D. G. for and Innovation. *A sustainable bioeconomy for Europe: strengthening the connection between economy, society and the environment: updated bioeconomy strategy.* Publications Office, 2018. DOI: 10.2777/478385.

[3] Lopes, H. J. S. et al. "C/N ratio and carbon source-dependent lipid production profiling in Rhodotorula toruloides". In: *Applied Microbiology and Biotechnology* 104.6 (2020), pp. 2639–2649. DOI: 10.1007/s00253-020-10386-5.

[4] Koutinas, A. A. et al. "Design and techno-economic evaluation of microbial oil production as a renewable resource for biodiesel and oleochemical production". In: *Fuel* 116 (2014), pp. 566–577. DOI: 10.1016/j.fuel.2013.08.045.

[5] Yang, X. et al. "Expression of phosphotransacetylase in Rhodosporidium toruloides leading to improved cell growth and lipid production". In: *RSC Advances* 8.43 (2018), pp. 24673–24678. DOI: 10.1039/c8ra03028f.

[6] Koutinas, A. and Papanikolaou, S. "Biodiesel production from microbial oil". In: *Handbook of Biofuels Production.* Elsevier, 2011, pp. 177–198. DOI: 10.1533/9780857090492.2.177.

[7] Unrean, P., Khajeeram, S., and Champreda, V. "Combining metabolic evolution and systematic fed-batch optimization for efficient single-cell oil production from sugarcane bagasse". In: *Renewable Energy* 111 (2017), pp. 295–306. DOI: 10.1016/j.renene.2017.04.018.

[8] Li, Y., Zhao, Z. (, and Bai, F. "High-density cultivation of oleaginous yeast Rhodosporidium toruloides Y4 in fed-batch culture". In: *Enzyme and Microbial Technology* 41.3 (2007), pp. 312–317. DOI: 10.1016/j.enzmictec.2007.02.008.

[9] Hu, C. et al. "Effects of biomass hydrolysis by-products on oleaginous yeast Rhodosporidium toruloides". In: *Bioresource Technology* 100.20 (2009), pp. 4843–4847. DOI: 10.1016/j.biortech.2009.04.041.

[10] Reķēna, A. et al. "Genome-scale metabolic modeling reveals metabolic trade-offs associated with lipid production in Rhodotorula toruloides". In: *PLOS Computational Biology* 19.4 (2023). Ed. by R. Mahadevan, e1011009. DOI: 10.1371/journal.pcbi.1011009.

[11] Vasconcelos, B. et al. "Oleaginous yeasts for sustainable lipid production—from biodiesel to surf boards, a wide range of "green" applications". In: *Applied Microbiology and Biotechnology* 103.9 (2019), pp. 3651–3667. DOI: 10.1007/s00253-019-09742-x.

[12] Tiukova, I. A. et al. "Genome-scale model of Rhodotorula toruloides metabolism". In: *Biotechnology and Bioengineering* 116.12 (2019), pp. 3396–3408. DOI: 10.1002/bit.27162.

[13] Buzzini, P. et al. "Carotenoid profiles of yeasts belonging to the genera Rhodotorula, Rhodosporidium, Sporobolomyces, and Sporidiobolus". In: *Canadian Journal of Microbiology* 53.8 (2007), pp. 1024–1031. DOI: 10.1139/w07-068.

[14] Vorapreeda, T. et al. "Alternative routes of acetyl-CoA synthesis identified by comparative genomic analysis: involvement in the lipid production of oleaginous yeast and fungi". In: *Microbiology* 158.1 (2012), pp. 217–228. DOI: 10.1099/mic.0.051946-0.

[15] Lian, J. and Zhao, H. "Recent advances in biosynthesis of fatty acids derived products in Saccharomyces cerevisiae via enhanced supply of precursor metabolites". In: *Journal of Industrial Microbiology and Biotechnology* 42.3 (2015), pp. 437–451. DOI: 10.1007/s10295-014-1518-0.

[16] Ratledge, C. and Wynn, J. P. "The Biochemistry and Molecular Biology of Lipid Accumulation in Oleaginous Microorganisms". In: *Advances in Applied Microbiology*. Elsevier, 2002, pp. 1–52. DOI: 10.1016/s0065-2164(02)51000-5.

[17] Kerkhoven, E. J., Lahtvee, P.-J., and Nielsen, J. "Applications of computational modeling in metabolic engineering of yeast". In: *FEMS Yeast Research* (2014), n/a–n/a. DOI: 10.1111/1567-1364.12199.

[18] Chen, M. et al. "Yeast increases glycolytic flux to support higher growth rates accompanied by decreased metabolite regulation and lower protein phosphorylation". In: *Proceedings of the National Academy of Sciences* 120.25 (2023). DOI: 10.1073/pnas.2302779120.

[19] Voet, D. and Voet, J. G. *Biochemistry (2nd ed.) Donald*. 2. rev. ed. Vol. [Hauptbd.] New York [u.a.]: Wiley, 1995. 1361 pp.

[20] DE BIAGGI, J. S. "Phytoene as the exclusive carotenoid in Rhodotorula toruloides: genetic modification and metabolic modelling". PhD thesis. Universidade Estadual de Campinas, Faculdade de Engenharia Química, Campinas, SP, 2023.

[21] Thiele, I. and Palsson, B. Ø. "A protocol for generating a high-quality genome-scale metabolic reconstruction". In: *Nature Protocols* 5.1 (2010), pp. 93–121. DOI: 10.1038/nprot.2009.203.

[22] Kerkhoven, E. J. "Advances in constraint-based models: methods for improved predictive power based on resource allocation constraints". In: *Current Opinion in Microbiology* 68 (2022), p. 102168. DOI: 10.1016/j.mib.2022.102168.

[23] Sánchez, B. J. et al. "Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints". In: *Molecular Systems Biology* 13.8 (2017). DOI: 10.15252/msb.20167411.

[24] Domenzain, I. et al. "Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0". In: *Nature Communications* 13.1 (2022). DOI: 10.1038/s41467-022-31421-1.

[25] Orth, J. D., Thiele, I., and Palsson, B. Ø. "What is flux balance analysis?" In: *Nature Biotechnology* 28.3 (2010), pp. 245–248. DOI: 10.1038/nbt.1614.

[26] Becker, S. A. et al. "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox". In: *Nature Protocols* 2.3 (2007), pp. 727–738. DOI: 10.1038/nprot.2007.99.

[27] Hucka, M. et al. "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models". In: *Bioinformatics* 19.4 (2003), pp. 524–531. DOI: 10.1093/bioinformatics/btg015.

[28] Dinh, H. V. et al. "A comprehensive genome-scale model for Rhodosporidium toruloides IFO0880 accounting for functional genomics and phenotypic data". In: *Metabolic Engineering Communications* 9 (2019), e00101. DOI: 10.1016/j.mec.2019.e00101.

[29] Coradetti, S. T. et al. "Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides". In: *eLife* 7 (2018). DOI: 10.7554/elife.32110.

[30] Aung, H. W., Henry, S. A., and Walker, L. P. "Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism". In: *Industrial Biotechnology* 9.4 (2013), pp. 215–228. DOI: 10.1089/ind.2013.0013.

[31] Arkin, A. P. et al. "KBase: The United States Department of Energy Systems Biology Knowledgebase". In: *Nature Biotechnology* 36.7 (2018), pp. 566–569. DOI: 10.1038/nbt.4163.

[32] Jagtap, S. S. and Rao, C. V. "Production of d-arabitol from d-xylose by the oleaginous yeast Rhodosporidium toruloides IFO0880". In: *Applied Microbiology and Biotechnology* 102.1 (2017), pp. 143–151. DOI: 10.1007/s00253-017-8581-1.

[33] Kot, A. M. et al. "Torulene and torularhodin: "new" fungal carotenoids for industry?" In: *Microbial Cell Factories* 17.1 (2018). DOI: 10.1186/s12934-018-0893-z.

[34] Kim, J. et al. "Multi-Omics Driven Metabolic Network Reconstruction and Analysis of Lignocellulosic Carbon Utilization in Rhodosporidium toruloides". In: *Frontiers in Bioengineering and Biotechnology* 8 (2021). DOI: 10.3389/fbioe.2020.612832.

[35] Pinheiro, M. J. et al. "Xylose Metabolism and the Effect of Oxidative Stress on Lipid and Carotenoid Production in Rhodotorula toruloides: Insights for Future Biorefinery". In: *Frontiers in Bioengineering and Biotechnology* 8 (2020). DOI: 10.3389/fbioe.2020.01008.

[36] Zhu, Z. et al. "A multi-omic map of the lipid-producing yeast Rhodosporidium toruloides". In: *Nature Communications* 3.1 (2012). DOI: 10.1038/ncomms2112.

[37] Arnold, P. K. et al. "A non-canonical tricarboxylic acid cycle underlies cellular identity". In: *Nature* 603.7901 (2022), pp. 477–481. DOI: 10.1038/s41586-022-04475-w.

[38] BANNO, I. "STUDIES ON THE SEXUALITY OF RHODOTORULA". In: *The Journal of General and Applied Microbiology* 13.2 (1967), pp. 167–196. DOI: 10.2323/jgam.13.167.

[39] Schultz, J. C. et al. “Metabolic engineering of Rhodotorula toruloides IFO0880 improves C16 and C18 fatty alcohol production from synthetic media”. In: *Microbial Cell Factories* 21.1 (2022). DOI: 10.1186/s12934-022-01750-3.

[40] Ebrahim, A. et al. “COBRApy: COnstraints-Based Reconstruction and Analysis for Python”. In: *BMC Systems Biology* 7.1 (2013). DOI: 10.1186/1752-0509-7-74.

[41] King, Z. A. et al. “BiGG Models: A platform for integrating, standardizing and sharing genome-scale models”. In: *Nucleic Acids Research* 44.D1 (2015), pp. D515–D522. DOI: 10.1093/nar/gkv1049.

[42] Ratledge, C. “The role of malic enzyme as the provider of NADPH in oleaginous microorganisms: a reappraisal and unsolved problems”. In: *Biotechnology Letters* 36.8 (2014), pp. 1557–1568. DOI: 10.1007/s10529-014-1532-3.

A. Genome-scale metabolic model reconstruction

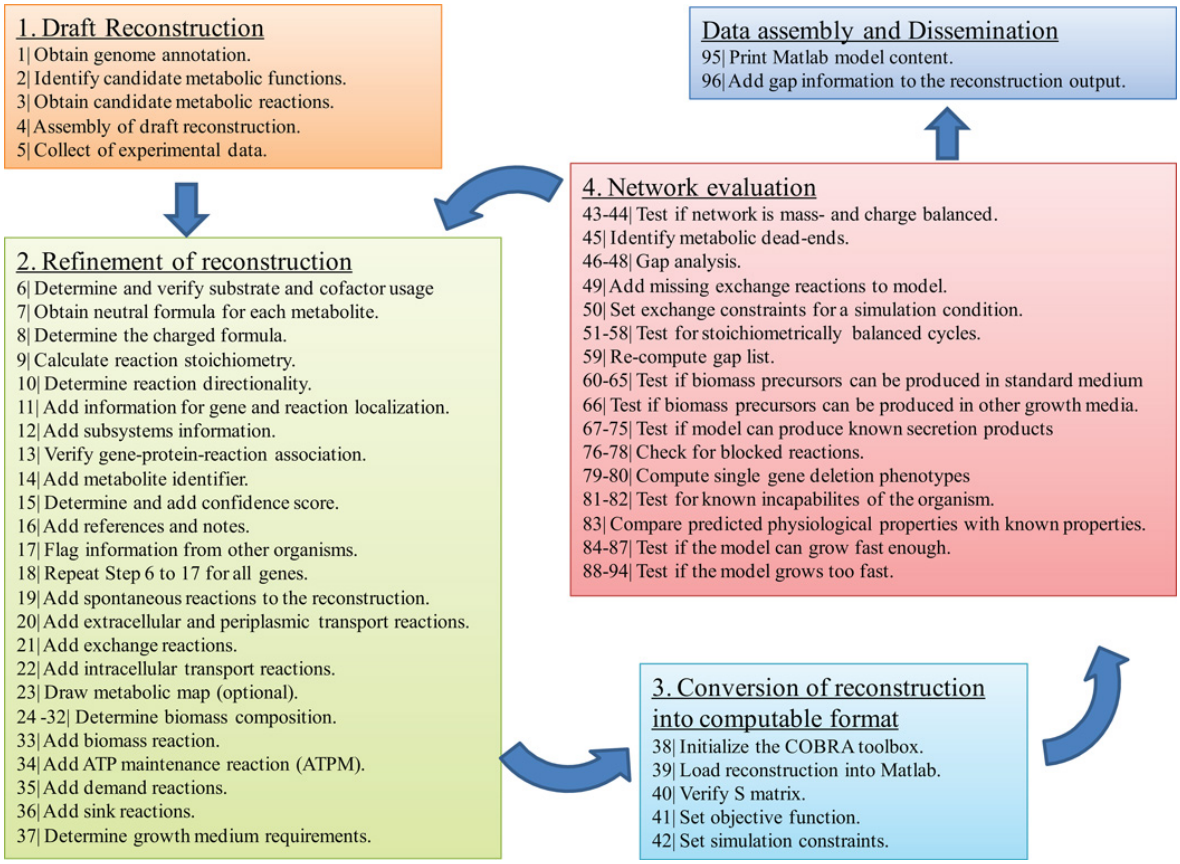

Figure A.1: The metabolic network reconstruction process. Figure is from [21].