

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

Bachelor thesis

Student: Maive Hanni  
Student code: 213004  
Supervisor: Alīna Reķēna,  
Department of Chemistry and  
Biotechnology, Early Stage  
Researcher  
Study program: Applied  
Chemistry and Gene  
Technology

Tallinn 2024



Ülegenoomsete metaboolsete mudelite võrdlus *Rhodotorula  
toruloides* lipogeneesi uurimiseks  
Bakalaureusetöö

Üliõpilane: Maive Hanni  
Üliõpilaskood: 213004  
Juhendaja: Alina Reķēna,  
Keemia ja biotehnoloogia  
instituut,  
doktorant-nooremteadur  
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

Signed digitally, 28.05.2024

The paper conforms to requirements in force.

Supervisor: Alīna Reķēna

Signed digitally, 28.05.2024

# Contents

<b>Abbreviations</b>	<b>5</b>
<b>Introduction</b>	<b>6</b>
<b>1 Theoretical background</b>	<b>7</b>
1.1 Need for new technologies to reduce reliance on fossil-based resources . . . . .	7
1.2 <i>Rhodotorula toruloides</i> . . . . .	8
1.3 Overview of growth laws in oleaginous microorganisms . . . . .	9
1.4 Overview of microbial cultivation methods . . . . .	10
1.5 Genome-scale metabolic modeling . . . . .	11
1.5.1 Constraint-based modeling . . . . .	12
1.5.2 Flux balance analysis . . . . .	13
1.6 Genome-scale metabolic models of <i>Rhodotorula toruloides</i> . . . . .	15
<b>2 Aims of the thesis</b>	<b>18</b>
<b>3 Methods</b>	<b>19</b>
3.1 Models . . . . .	19
3.2 Selecting experimental data . . . . .	19
3.3 Biomass equation in the models . . . . .	20
3.4 Flux balance analysis . . . . .	20
<b>4 Results</b>	<b>22</b>
4.1 Biomass maximisation as an objective function . . . . .	22
4.2 Non-growth associated maintenance minimisation as an objective function . .	32
<b>5 Conclusion</b>	<b>36</b>
<b>Acknowledgements</b>	<b>37</b>
<b>References</b>	<b>38</b>
<b>Abstract</b>	<b>43</b>
<b>Annotatsioon</b>	<b>44</b>
<b>A Appendix</b>	<b>45</b>
A.1 Non-exclusive licence for reproduction and publication of a graduation thesis .	45

# Abbreviations

**ACL** ATP-citrate lyase

**acetyl-CoA** acetyl-coenzyme A

**ATP** adenosine triphosphate

**BiGG knowledge-base** biochemical, genetic, and genomic knowledge-base

**C/N** carbon-to-nitrogen ratio

**cMDH** cytosolic malate dehydrogenase

**DNA** deoxyribonucleic acid

**DW** dry cellular weight

**FAEE** fatty acid ethyl ester

**FAME** fatty acid methyl ester

**FAS** fatty acid synthase

**FBA** flux balance analysis

**GAM reaction** growth-associated maintenance reaction

**GEM** genome-scale model

**ME** malic enzyme

**NADPH** nicotinamide adenine dinucleotide phosphate (reduced form)

**NGAM reaction** non-growth-associated maintenance reaction

**oxPPP** oxidative part of pentose phosphate pathway

**PPP** pentose phosphate pathway

**RNA** ribonucleic acid

**SBML** Systems Biology Markup Language

**SCO** single-cell oil

**TAG** triacylglycerol

**TCA cycle** tricarboxylic acid cycle (citric acid cycle)

# Introduction

To combat climate change and reduce dependence on fossil-based resources, numerous countries worldwide are transitioning towards a bio-based economy. This transition necessitates innovative processes for the sustainable production of chemicals, materials, and fuels. However, the current production of biodiesel from oilseeds and waste oils is insufficient to meet global demand [1], highlighting the need for biofuels derived from non-edible sources.

Microbial oils, also known as single-cell oils (SCOs), represent a promising alternative. These oils, produced by oleaginous microorganisms, do not compete with the food sector and can utilize low-value waste streams. The oleaginous yeast *Rhodotorula toruloides* is particularly noteworthy - it is capable of accumulating very high amounts of lipids and what is more, it has good tolerance to inhibitory compounds found in biomass hydrolysates and a broad substrate range [2]. Metabolic pathways enabling lipid production in *R. toruloides* have been generally defined from earlier systems biology studies, however their operation is not fully understood.

Genome-scale models (GEMs) offer a comprehensive approach to investigate microbial metabolism. For *R. toruloides*, several genome-scale metabolic models have been developed independently, but no comprehensive comparison of the simulated metabolic fluxes through pathways generating lipid precursors for these GEMs has been done.

This thesis presents a comparison of predictions from rhto-GEM [3], iRhtoC [4], Rt\_IFO0880 [5], and Rt\_IFO0880\_LEBp2023 [6] in lipogenesis focused central carbon metabolism. To present a comprehensive overview of current situation in metabolic model development for *R. toruloides*, flux balance analysis (FBA) using two different objective functions, biomass maximization and non-growth associated maintenance (NGAM), was performed to carry out metabolic model simulations using continuous cultivation data of *R. toruloides*.

Results are presented in two chapters. The first chapter presents the differences in flux predictions using biomass maximization as an objective function by all models. The second chapter explores the differences in flux predictions by Rt\_IFO0880-based models using the NGAM as an objective function aiming at comparison of the metabolic activity in pathways devoted to nicotinamide adenine dinucleotide phosphate (NADPH) regeneration, a cofactor essential for lipid biosynthesis.

# 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 [7]. 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 [8]. Bio-based products may enhance environmental sustainability compared to fossil based equivalents [7].

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. Advancements have highlighted the potential of chemicals derived from plant oils as alternative feedstocks to the petrochemical industry [9]. 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) [1].

The demand for vegetable oils has increased rapidly due to the expansion of the demand for edible oils in the food market (representing over 80%). The increasing demand in biodiesel sector also represents an increasing part in the growth of the demand of vegetable oils. [10] But the production of biodiesel from oilseeds and waste oils does not sustain the global demand [1] and food crisis has shown the need for the development of second-generation biofuels derived from non-edible sources, such as lignocellulosic raw materials and industrial waste streams [11].

Another promising source of fatty acids for oleochemical production are microbial oils, also called single-cell oils (SCOs), that represent the triacylglycerides produced by microorganisms [12]. Research has focused on the development of biodiesel production from SCOs that are produced via cultivation using oleaginous microorganisms (microorganisms capable of accumulating lipids at more than 20% of the total cellular dry weight (DW)). Biodiesel production from SCOs 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. [1]

## 1.2. *Rhodotorula toruloides*

*Rhodotorula toruloides* (previously *Rhodosporidium toruloides*) is a non-conventional, oleaginous yeast, which can accumulate lipids up to 76.1% of cell dry weight [13]. What is more, *R. toruloides* has good tolerance to inhibitory compounds that are naturally found in biomass hydrolysates [2, 14]. *Rhodotorula toruloides* is an exceptional microbial lipid producer and has recently emerged as one of the most promising yeasts for bioproduction [15, 16].

*R. toruloides* occurs naturally in leaves, soil, sea water, etc. It has a broad substrate range, which has made this yeast popular for producing biological oils from inedible substrates such as pentose sugars and crude glycerol. 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 [13, 17]. *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 [18].

Fatty acids mainly accumulate as TAGs, and they are produced via four cyclic enzymatic reactions that require 1 adenosine triphosphate (ATP) and 2 nicotinamide adenine dinucleotide phosphate (NADPH) molecules for adding 1 acetyl-coenzyme A (acetyl-CoA) to the fatty acid chain [19]. During fatty acid synthesis and elongation, acetyl-CoA is the donor of C2-carbon. NADPH is required for reduction steps and it is known to be mainly produced by malic enzyme (decarboxylating malate dehydrogenase, ME), and by glucose 6-phosphate dehydrogenase and phosphogluconate dehydrogenase in the pentose phosphate pathway (PPP). [20]

Metabolic pathways producing acetyl-CoA and a cofactor NADPH have been subject to metabolic studies in *R. toruloides*. Compared to *Saccharomyces cerevisiae*, *R. toruloides* has several different enzymatic pathways that facilitate the generation of lipid precursors. Important difference is that *R. toruloides* possesses the enzyme ATP-citrate lyase (ACL), which synthesises acetyl-CoA from citrate. ACL has been demonstrated to be upregulated in *R. toruloides* during lipid accumulation [21] and it has been suggested to be the main source of acetyl-CoA for lipid synthesis in oleaginous species [22]. It has been found that *R. toruloides* also possesses the enzyme phosphoketolase [23], which is able to convert fructose-6-phosphate and/or xylulose-5-phosphate into acetyl-phosphate + erythrose-4-phosphate and/or acetyl-phosphate + glyceraldehyde-3-phosphate [24], producing an acetyl residue (acetyl-phosphate) directly from a C5-substrate. Thus, phosphoketolase is a more efficient way of providing acetyl-CoA compared to the decarboxylation of pyruvate to acetyl-CoA, where one third of the carbon substrate is lost [23]. But the upregulation of phosphoketolase during nutrient limitation in *R. toruloides* has not been acknowledged [23].

It has been found that in oleaginous yeast malic enzyme is an important enzyme involved in the production of NADPH [25], but the role of ME in *R. toruloides* is not clearly understood. Proteomics analysis of *R. toruloides* has suggested that NADPH is mainly produced through the pentose phosphate pathway (when grown on xylose and glucose) [21]. Lipid biosynthetic



reactions downstream of acetyl-CoA synthesis do not significantly differ between oleaginous and non-oleaginous yeast species [26].

### 1.3. Overview of growth laws in oleaginous microorganisms

It has been well-established that an imbalance of nutrients in the culture medium triggers lipid accumulation in oleaginous microorganisms. When a crucial nutrient, typically nitrogen, is depleted, cells continue to assimilate excess carbon substrate and transform it into storage fat. [25] Thus, the carbon-to-nitrogen ratio (C/N) is a significant factor in initiating lipid accumulation [27]. The cells take in carbon faster than they can convert it into new cells, so the surplus carbon is stored by turning it into lipid. This lipid accumulation necessitates a slower cell growth rate, allowing the excess carbon to be assimilated more quickly than it can be converted into biomass, thus directing the surplus carbon into lipid. This process of lipid accumulation can also be accomplished in continuous culture with oleaginous yeast, where it is essential to maintain a sufficiently low dilution rate (growth rate) to enable the cells to assimilate the glucose. Continuous cultivation studies have clearly demonstrated that the lipid synthesis rate is slower than the maximum growth rate. [25]

The first major biochemical distinction identified between oleaginous and non-oleaginous yeast species was the presence of ATP-citrate lyase in oleaginous yeast during lipid accumulation. This enzyme has been shown to be crucial for a eukaryotic microbial cell to accumulate significant amounts of triacylglycerol lipids. Yeasts without ACL invariably had low lipid cell contents. However, some yeasts that had ACL activity did not accumulate lipids, suggesting that some other enzyme activities are also necessary for lipid accumulation. [25]

It has been found that another important enzyme in lipogenesis is malic enzyme, which generates NADPH used by fatty acid synthase (FAS). NADPH is also generated by glucose-6-phosphate dehydrogenase, hosphogluconate dehydrogenase, and NADP-dependent isocitrate dehydrogenase. When malic enzyme activity was inhibited using selective inhibitors (sesamol), lipid content in the cells decreased by almost 90% (from 24% of the cell biomass to 2%), without significantly affecting growth. This led to the conclusion that sesamol was specifically inhibiting both the cytoplasmic and membrane-bound malic enzymes, and without malic enzyme, the cell was unable to accumulate lipid or carry out its desaturations. [25]

Wynn and Ratledge (1997) further demonstrated that in a mutant of *Aspergillus Nidulans* that lacked malic enzyme activity, only half the lipid that had been previously produced by a competent strain under nitrogen-limited growth conditions, was now produced. Fatty acid biosynthesis itself was still functional, and phospholipids were produced. Meaning that the cells can function without malic enzyme, but they cannot produce storage triacylglycerols in any significant quantity - without malic enzyme activity, the flow of carbon from glucose to lipid was significantly reduced, and only essential lipids were produced, presumably using other sources of NADPH. [25]

## 1.4. Overview of microbial cultivation methods

Microorganisms play a crucial role in biotechnology, being utilized for the production of a variety of bioproducts. In an industrial setting, microbes are cultivated in large-scale bioreactors to manufacture biopharmaceuticals, dietary supplements, biofuels, or other chemical substances. The cultivation process requires careful control of various parameters to ensure optimal growth conditions for the microbes. These parameters include temperature, pH, oxygen levels, agitation (stirring), and pressure. It is vital to regulate these factors to provide a suitable physical and chemical environment for the cells, thereby enhancing their productivity. There are three main methods for microbial cultivation: batch, fed-batch and steady-state. [28]

In a batch culture, no nutrients are added or waste removed. Microorganisms growing in such a closed culture follow a pattern known as the growth curve, which, when plotted against time, reveals different phases. The first phase of the growth curve, the lag phase, represents a small number of cells (known as an inoculum) introduced into a fresh culture medium, a nutrient-rich broth that promotes growth. During this phase, the cell count remains unchanged, but the cells increase in size and are metabolically active, producing proteins necessary for growth. The log phase follows next, where the cells divide actively, and their count increases exponentially. Cells in the log phase exhibit a constant growth rate and uniform metabolic activity, making them ideal for industrial applications and research work. [29]

However, as the cell count rises during the log phase, several factors contribute to a slowdown in the growth rate. Accumulation of waste products, gradual depletion of nutrients, and limited oxygen availability due to increased consumption all contribute to this slowdown. This leads to a plateau in the total number of live cells, known as the stationary phase. In this phase, the number of new cells created by cell division equals the number of cells dying, resulting in a relatively stagnant total population of living cells. As the culture medium becomes saturated with toxic waste and nutrients get exhausted, cell death outpaces cell division, leading to an exponential decrease in the cell count. This phase is named the death or decline phase. [29]

Fed-batch cultivation is a variation of batch cultivation. Microorganisms are initially grown under batch conditions, after which nutrients are incrementally added to the fermenter throughout the remaining cultivation duration. The addition of fresh nutrients typically results in significant biomass accumulation during the exponential growth phase. Therefore, fed-batch cultivation is particularly useful for bioprocesses aiming for high biomass density or high product yield when the desired product is positively correlated with microbial growth. [28]

In industrial applications and research work, it is beneficial to keep cells in the logarithmic phase of growth. A steady-state cultivation, also called chemostat, enables the maintenance of a continuous culture thanks to the addition and removal of fluids, adjusted to keep the culture in the logarithmic phase of growth. [29] Fresh medium is continuously added to the fermenter, while used medium, (toxic) metabolites and cells are simultaneously harvested. Unlike fed-batch cultivation, the maximum working volume of the vessel does not limit the amount of fresh medium or feed solution that can be added to the culture during the process.

When the addition and removal rates are equal, the culture volume remains constant. The cellular growth rate and environmental conditions, like the concentrations of metabolites, remain constant. Steady-state cultures can last for days, weeks, or even months, significantly reducing downtime and making the process more economically competitive. [28]

Metabolic analysis requires experimental data for the determination of metabolic fluxes. These flux calculations are based on the measurements of specific rates for the uptake of substrate and the formation of product during cultivations. The specific substrate uptake rate, denoted as  $r_s$  (expressed in grams or moles of the compound per gram or mole of biomass and unit time), represents the rate at which cells take up a substrate (the flux into the cells). In steady-state cultivations, this rate can be calculated as the difference between the glucose concentration in the feed and that in the bioreactor, multiplied by the dilution rate. The specific product formation rate  $r_p$  represents the fluxes of metabolic products out of the cell, that is, the rate at which products are secreted. Additionally, there is an accumulation of biomass within the cell and that rate is represented as a flux of the specific growth rate,  $\mu$  expressed in 1/h, that equals the dilution rate in a steady-state chemostat. [30]

## 1.5. Genome-scale metabolic modeling

Cellular metabolism involves numerous reactions that are part of the conversion of resources into energy and precursors needed for biosynthesis. Rates of these reactions are called fluxes and they illustrate metabolic activity. Flux of a reaction results from a combined regulation of many biological levels (transcription, translation, post-translational modifications and protein-protein interactions). [31] Hence, metabolic fluxes represent cellular phenotype under certain conditions and therefore analyzing the flux distribution of metabolites is very useful for studying cell metabolism [32]. It is difficult to measure intracellular fluxes experimentally, but it is possible to predict these fluxes thanks to metabolic models [31].

When the first full genome sequences were published in the 1990s, in principle it became possible to identify all the gene products involved in a given organism’s biological processes. This, with well studied biochemistry of metabolism, allowed the reconstruction of metabolic networks on a genome-scale for a target organism. Such reconstructions, containing biochemical, genetic, and genomic (BiGG) knowledge, can be converted into a mathematical format allowing the formulation of genome-scale models (GEMs). [33] Thanks to the fact that GEMs account for all known genes, proteins, and biochemical reactions, it is possible to conduct systematic analysis of a given organism’s metabolism, where typically the objective is to obtain an overview of possible flux patterns. It is possible to integrate omics data and experimental metabolic fluxes to GEMs for generating a holistic view of metabolism in different physiological states, enabling a greater understanding of cellular physiology, providing valuable information for metabolic engineering to develop better microbial factories [34].

The metabolic reconstruction process usually is very labor- and time consuming. 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 for 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. [35]

### 1.5.1. Constraint-based modeling

Genome-scale metabolic models (GEMs) are commonly used to compute metabolic phenotypes. However, these models also depend on a set of constraints due to various factors that limit cellular functions. These constraints fall into four categories: basic physico-chemical constraints, spatial or topological constraints, condition-dependent environmental constraints, and self-imposed or regulatory constraints. [36]

Physico-chemical constraints are fundamental and provide inviolable constraints on cell functions, including the conservation of mass, energy, and momentum. Topobiological constraints arise from the crowding of molecules inside cells, affecting the form and function of biological systems. For instance, bacterial DNA, which is about 1,000 times longer than a cell, must be tightly packed yet easily accessible for transcription. [36] Environmental constraints, which are time and condition dependent, include factors like nutrient availability, pH, temperature, osmolarity, and the availability of electron acceptors. These constraints are crucial for the quantitative analysis of microorganisms and require defined media and well-documented environmental conditions for integrating data into accurate and predictive quantitative models. Regulatory constraints are self-imposed and subject to evolutionary change, allowing the cell to eliminate suboptimal phenotypic states. These constraints are implemented in various ways, including the amount of gene products made and their activity. [36]

A significant limitation of conventional GEMs is that they do not account for enzyme abundances and kinetics, which limit metabolic fluxes. These models often assume that the uptake rate of the carbon source limits production, which may oversimplify the situation. [37] The synthesis of enzymes is resource- and energy-intensive, and their catalytic capacities are limited by their kinetics. Furthermore, the quantity of enzymes is space-constrained. [34]

An increase in the requirement of an enzyme or a pathway would be a trade-off for other functions. Experimental evidence suggests that resource re-allocation could be an effective strategy in response to nutrient and growth shifts, demonstrating the biological significance of proteome constraints. [38] Incorporating such constraints into a metabolic model can lead to more realistic results by reducing simulated flux distributions to those that are most economic and limiting the phenotypes that the model can simulate. [34] Overview of a reconstruction of a GEMs is shown in the Figure 1.1.

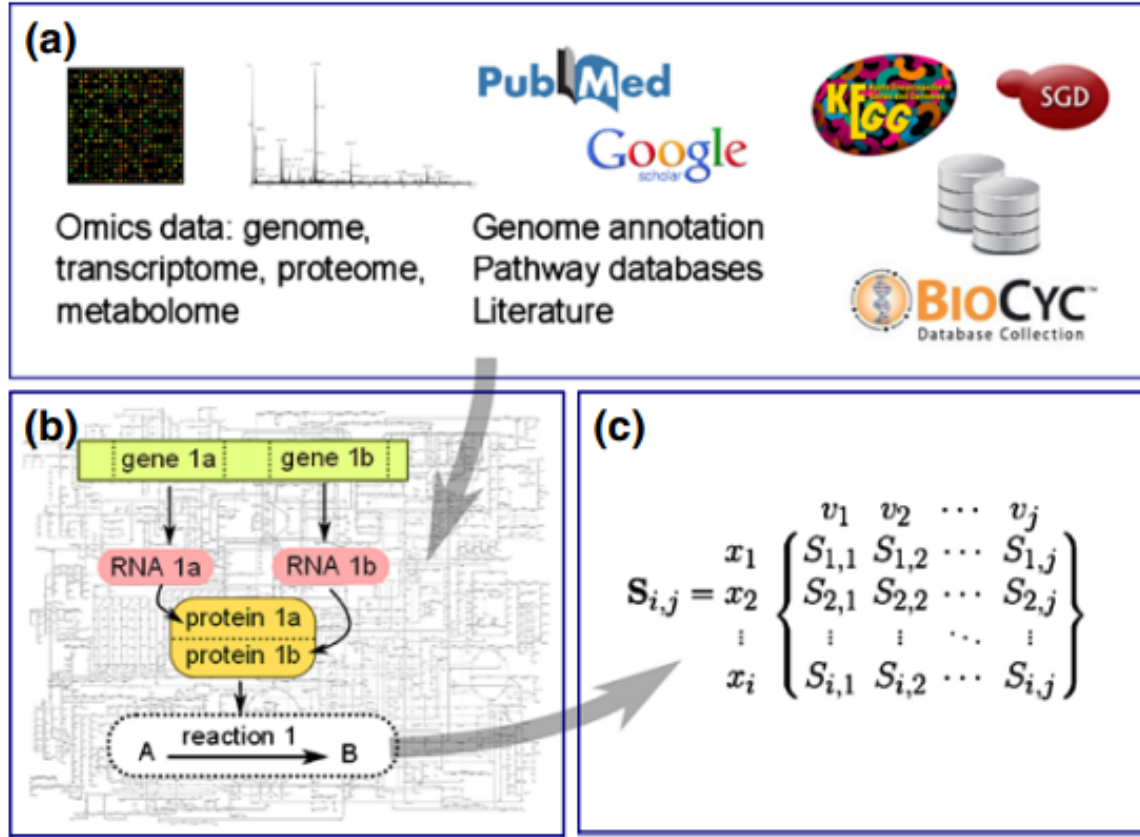


Figure 1.1: Reconstruction of a GEM. (a) The genome annotation is used to reconstruct the draft. (b) Gene-protein-reaction relationships are defined for the metabolic model. (c). A solution space is defined from the constraints applied to the model. Figure is from article [39].

### 1.5.2. Flux balance analysis

Flux balance analysis (FBA) is a mathematical approach for analyzing the flow of metabolites through a metabolic network. It is a commonly employed method for investigating biochemical networks, especially the genome-scale metabolic network reconstructions. Using FBA, it is possible to predict an organism's growth rate or the production rate of a biotechnologically important metabolite. [40]

The initial phase in FBA involves the mathematical representation of metabolic reactions. The reconstructed genome-scale networks can be transformed into mathematical stoichiometric matrices  $S \in \mathbb{R}^{(m \times n)}$ , where each row corresponds to one unique metabolite (for a system with  $m$  metabolites) and each column corresponds to an individual reaction ( $n$  reactions). Each column's entries are the stoichiometric coefficients of the metabolites involved in a reaction. A negative coefficient is assigned to each metabolite that is consumed, while a positive coefficient is assigned to each metabolite that is produced. A stoichiometric coefficient of zero is assigned to each metabolite that does not participate in a specific reaction. The stoichiometric matrix  $S$  is sparse, as most biochemical reactions involve only a few different metabolites. The vector  $v$  represents the fluxes of all the reactions in the network and it has a length of  $n$ . [40]

At steady-state, which is simulated by GEMs, there is neither accumulation nor depletion of metabolites in a metabolic network, meaning the rate of production of each metabolite in the network must equal its rate of consumption. This flux balance can be mathematically represented as in Equation (1.1) [36]:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0}, \quad (1.1)$$

where  $\mathbf{S}$  is the stoichiometric matrix and  $\mathbf{v}$  is the vector of the fluxes. Any vector  $\mathbf{v}$  that satisfies this equation is said to be in the null space of  $\mathbf{S}$ . In any realistic large-scale metabolic model, there are more reactions than there are compounds ( $n > m$ ). This means that there are more unknown variables than equations, so there is no unique solution to this system of equations. [40] Thus, the solution space is further constrained by a set of upper and lower bounds on the fluxes:

$$a_i < v_i < b_i, \quad (1.2)$$

where  $v_i$  is the flux of a certain reaction (usually experimentally measured reactions), and  $a_i$  and  $b_i$  are the upper and lower bounds of the flux, respectively.

The next phase in flux balance analysis involves defining a biological objective that aligns with the research problem being studied. Constraints define a spectrum of potential solutions, but it is still possible to pinpoint and analyze individual solutions within this space. From a mathematical standpoint, the objective is represented by an objective function, which quantifies the contribution of each reaction to the phenotype. For instance, the biomass reaction, which drains precursor metabolites from the system in accordance with their relative stoichiometries, predicts biomass production. The biomass growth rate reaction is then scaled so that its flux matches the organism’s exponential growth rate, denoted by  $\mu$ . [40]

Another examples of functions used as objective function, are the maximum ATP generation or a specific product formation [39]. Also growth-associated maintenance (GAM) and non-growth-associated maintenance (NGAM) reactions can be used. GAM and NGAM reactions account for additional energetic requirements in the form of ATP for growth, beyond metabolic costs [41]. GAM accounts for the energy needed for cell replication, including macromolecular synthesis (proteins, deoxyribonucleic acid (DNA), and ribonucleic acid (RNA)). Determining GAM is best achieved through chemostat growth experiments. NGAM represents ATP requirements for cell maintenance that is not related to growth (in GEMs it is noted as an ATP hydrolysis reaction). The rate of this reaction can be estimated from growth experiments. [35]

FBA uses linear programming to find such vector  $\mathbf{v}$  that satisfies constraints mentioned above (Equations (1.1) and (1.2)) and also maximizes or minimizes the given objective function (see Figure 1.2). [40] The COBRA [42] and RAVEN [43] toolboxes can be used for solving this kind of optimization problem for large systems of equations efficiently. Both toolboxes are freely available in Matlab, and the COBRA package is also available in Python. These toolboxes use models that are saved in the Systems Biology Markup Language (SBML) [44] format.

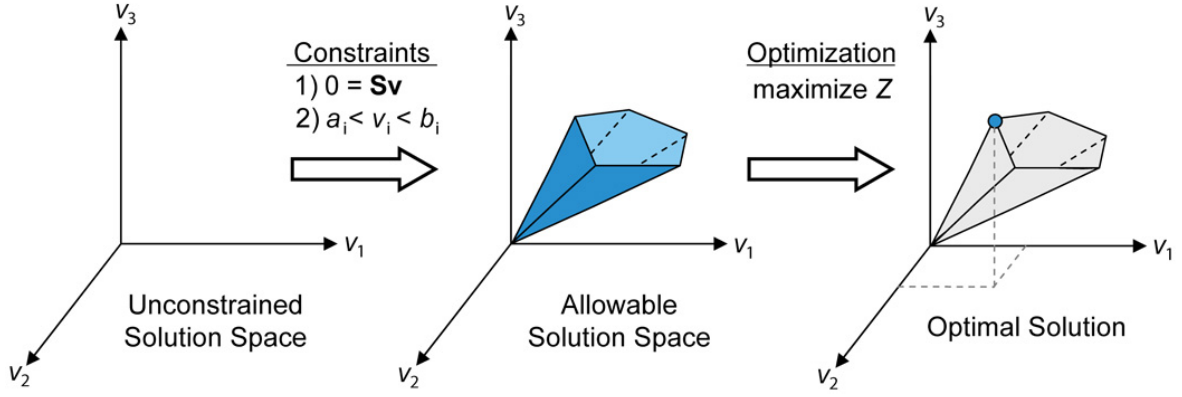


Figure 1.2: Conceptual basis of constraint-based modeling and FBA. The figure is from [40].

Flux balance analysis provides only a single optimal solution of the solution space, limiting insight into all of the solution space [42]. An alternative approach to FBA is uniform random sampling of the flux space, which fully determines the range of feasible steady-state fluxes, within the network, given certain physicochemical constraints [45]. This method does not necessitate the definition of an objective function [46]. From sampling solutions, the average and standard deviation for each metabolic flux in the GEM can be calculated. The flux distributions derived from this sampling can answer questions about the most probable flux value for any reactions and correlations between two reactions under given constraints. [42]

## 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. [3]. The model includes 4869 genes, 897 reactions, and 3334 metabolites. This model is based on the genome sequence of *R. toruloides* strain NP11 [21] (which is accessible from NCBI database [47]). For the reconstruction of the metabolism parts that are relatively conserved between fungal species, the well-curated GEM of *Saccharomyces cerevisiae* was utilized as template model (yeast-GEM version 8.2.0, 16). Orthologous genes were identified through bi-directional BLASTP against the *S. cerevisiae* S288c reference genome.

To transform the 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 or otherwise deprecated. The 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 [48]. 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. [3]

The model incorporates knowledge derived from genomics and proteomics data generated for *R. toruloides* and was validated using cultivation data. Simulations of rhto-GEM on various carbon sources showed good match with experimentally reported growth rates. The model analysis helped to identify potential genetic engineering strategies for enhanced lipid production. [3]

### **iRhto1108**

In the same year that Tiukova et al. introduced rhto-GEM, Dinh et al. presented another *R. toruloides* genome-scale metabolic model named iRhto1108 [4]. This model was built upon functional genomics data from [49] and prior knowledge. The model is based on the metabolic network of the strain IFO0880 [49] (available from JGI database [50]). It includes 2204 reactions, 1985 metabolites and 1108 genes.

The authors supplemented and integrated previous knowledge with in-house generated biomass composition and experimental measurements related to the metabolic capabilities of the organism. The iRhto1108 model incorporates yeast biochemistry information from (i) previously constructed genome-scale models (*S. cerevisiae* yeast 7.6 [51], (ii) KBase fungal models [52]), and (iii) *R. toruloides* specific information extracted from the primary literature [49][53][54].

The essential metabolic functions and growth capability of the model were thoroughly validated with experimental results, including gene essentiality [49] and growth data. The iRhto1108 model was successful in reproducing the lipid accumulation phenotypes observed in experiments. It can effectively represent the metabolism of *R. toruloides* and provide valuable predictions that have been validated with experimental data, including suggestions for genetic alterations that could lead to triacylglycerol overproducing strains. [4] Considering that *R. toruloides* is a non-conventional microorganism, the iRhto1108 model has been a promising start for GEMs of *R. toruloides*, and it holds potential to assist in future research on that yeast.

Two versions of the model were created, namely iRhtoC and iRhtoN, corresponding to conditions limited by carbon and nitrogen, respectively. These two versions are identical with the exception of the biomass reaction, acyl composition reaction, and the growth associated maintenance reaction. Experimental measurements were conducted to determine the organism-specific macromolecular composition and ATP maintenance requirements under these two separate growth conditions. [4]

### **Rt\_ IFO0880**

In 2021, a comprehensive multi-omics analysis of lignocellulosic carbon utilization in *R. toruloides* was conducted by Kim et al., leading to the reconstruction of a genome-scale metabolic network named Rt\_ IFO0880 [5]. This metabolic reconstruction consists of 1106 genes, 1934 reactions, and 2010 metabolites across nine compartments.



The initial draft of the metabolic network reconstruction was built using high-quality metabolic network models of model organisms and orthologous protein mapping. The draft was then manually curated into a metabolic model, with the aid of functional annotation and a variety of multi-omics data, including transcriptomics, proteomics, metabolomics, and RB-TDNA sequencing. The authors identified numerous incorrect reactions, particularly in fatty acid biosynthesis and beta-oxidation. The reactions and genes in the central metabolic pathways were manually verified for their co-factor usage and localization. The biomass reaction was updated using multi-omics and other experimental measurements. Updates included the DNA composition (using the genome sequence), RNA composition (using transcriptomics data), amino acid composition (using proteomics data), and lipid composition (using fatty acid methyl ester analysis). The authors carried out a genome-scale evaluation and iterative improvement of the model, utilizing high-throughput growth phenotyping and functional genomics. The metabolic model’s ability to predict growth on various carbon, nitrogen, sulfur, and phosphate sources was tested. The model was further refined to resolve inconsistencies, and several genes with erroneous ortholog mapping were removed. [5]

The metabolic network model was validated against high-throughput growth phenotypes in 213 growth conditions and conditional gene essentiality in 27 growth conditions. The model demonstrated high prediction accuracies and significantly expanded the breadth and depth of metabolic coverage compared to previously published models [3, 4]. The authors of the model believe that the developed metabolic network Rt\_IFO0880 is the most complete and accurate to date. [5]

### **Rt\_IFO0880\_LEBp2023**

In a doctoral thesis focused on the carotenoid production of *R. toruloides*, the author compared the four genome-scale metabolic models of *R. toruloides*. These included the previously mentioned models and also model rhto-GEM\_BioEng, a version of the rhto-GEM with integrated carotenoids into the biomass composition and an alternative xylose assimilation pathway [55]. The model Rt\_IFO0880 was selected for further enhancement by the author due to its superior representation of the metabolic pathways involved in the biosynthesis of carotenoids in *R. toruloides* and its higher accuracy and sensitivity in predicting gene essentiality. [6]

Several modifications to the model Rt\_IFO0880 were made, including (1) the addition of the reaction and gene-protein-reaction relation (GPR) corresponding to cytosolic malate dehydrogenase (cMDH), (2) the lower and upper flow limits of the xylokinase and phytoene dehydrogenase enzymes were equalized to zero to reflect the absence of detectable activity of the former and the deletion of the gene encoding the latter, (3) the creation of a phytoene transport reaction for the lipid body compartment, and (4) the modification of the lower limit of the cytosol-to-phytoene NADP<sup>+</sup> transport reaction. The updated model, named Rt\_IFO0880\_LEBp2023, was further validated against experimental data. [6]

## 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 genome-scale metabolic models of *R. toruloides*. For that, flux balance analysis is carried out with all of the models using different:

- (i) constraints, such as the specific glucose uptake  $r_{glu}$  and the specific growth rate  $\mu$ , and
- (ii) objective functions, such as biomass maximisation and non-growth associated maintenance reaction minimisation.

The intracellular fluxes of pentose phosphate pathway and citric acid cycle enzymes are predicted by these models to understand the main sources of lipid precursors, acetyl-CoA and NADPH, within these models.

## 3. Methods

### 3.1. Models

Genome-scale metabolic models of *R. toruloides* rhto-GEM, iRhtoC, Rt\_IFO0880 (JSON files) were obtained from supplemental files from publications [3–5] and the fourth model Rt\_IFO0880\_LEBp2023 [6], made by a group member, was obtained directly from him.

All scripts are available in Github repository ([https://github.com/maivehanni/BSc\\_thesis](https://github.com/maivehanni/BSc_thesis)) upon request sent to maive.hanni@gmail.com.

### 3.2. Selecting experimental data

In the simulations, the experimental steady-state cultivation data of *R. toruloides* strain IFO0880 in 1 L lab-scale bioreactors (Applikon Biotechnology, Delft, Netherlands) was used. The experiment was done by TalTech Food Tech and Bioengineering research group members and the results have not yet been published. The cultivation process is described in more detail in [56], with the exception that the continuous cultivation regime was used here instead of batch. Briefly, the cultivations were performed at pH 6.0, controlled by the addition of 2 mol/L KOH; dissolved oxygen was maintained at greater than 25% thanks to keeping the airflow at 1-vvm and stirring speeds 400-600 rpm. Cultivation medium contained glucose 10 g/L as the sole carbon source, 5 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3 g/L KH<sub>2</sub>PO<sub>4</sub>, 0.5 g/L MgSO<sub>4</sub> heptahydrate [57], supplemented with vitamins and minerals according to Verduyn [58]. Bioreactors were equipped with gas analyser (BlueSens gas sensor GmbH, Herten, Germany) used for measuring the composition of CO<sub>2</sub> and O<sub>2</sub> in the gas outflow. Specific rates of consumption and production 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: *R. toruloides* strain IFO0880 continuous cultivation results from lab experiments. Cultivations were carried out in 1 L bioreactors, at pH 6.0, airflow at 1-vvm, stirring 400-600 rpm. Cultivation medium contained glucose 10 g/L, 5 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3 g/L KH<sub>2</sub>PO<sub>4</sub>, 0.5 g/L MgSO<sub>4</sub> heptahydrate, supplemented with vitamins and minerals. Bioreactors were equipped with gas analyser.

Specific growth rate $\mu$ 1/h	Specific glucose uptake rate $r_{glu}$ mmol/gDW/h	Specific O <sub>2</sub> uptake rate $r_{O_2}$ mmol/gDW/h	Specific CO <sub>2</sub> secretion rate $r_{CO_2}$ mmol/gDW/h	Specific glycerol secretion rate $r_{gly}$ mmol/gDW/h
0.049	0.476	1.083	1.171	< 0.02
0.100	1.114	2.521	2.521	
0.151	1.648	3.851	3.854	
0.203	2.305	4.352	5.834	
0.25	-	-	-	
0.301	3.1	6.327	7.415	

### 3.3. Biomass equation in the models

For all simulations, the default biomass composition of each model was used, respectively. In the models, the biomass composition is represented by the biomass equation and corresponds to the biomass contents measured in respective publications. In the models rhto-GEM, iRhtoC and Rt\_IFO0880-based models, the lipid content, which is the most important parameter for present study, is around %, 12.3% and < 10%, respectively. In all models, the lipid composition is comparable. What is more, these values are similar to the lipid content of continuous cultivation of *R. toruloides* reported by Shen et al. [59] and thus correspond to the physiological data used in the simulations (section 3.2).

### 3.4. Flux balance analysis

Model simulations were performed using the COBRApy package (version 0.29.0) [60] in Python (version 3.11.4) with all four models. Throughout the process, metabolic flux patterns were predicted using flux balance analysis [40] from COBRApy package with the Gurobi mathematical optimization solver (version 11.0.0, Gurobi Optimization Inc.).

The following functions from COBRApy were used:

- (i) `read_sbml_model()` for importing the metabolic model in SBML format;
- (ii) `model.objective` for defining the objective function;
- (iii) `model.reactions.get_by_id().bounds` for assigning FBA bounds;
- (iv) `model.optimize()` for calculating the solution (default is the maximization of the objective function and the minimization is achieved by using `model.optimize('minimize')`);
- (v) `loopless_solution()` for obtaining a new flux distribution, where the sum of abso-

lute non-exchange fluxes is minimized (`loopless_solution()` is based on a previously obtained reference flux distribution with the function `model.optimize()`);

- (vi) `model.reactions.reaction.name` was used for obtaining the names of reactions instead of their IDs.

Calculated fluxes were stored in a Pandas (version 2.1.3) dataframe and the fluxes of interest were further visualized using Matplotlib (version 3.8.2) package.

Model simulations were carried out on five different growth rates. Glucose uptake rate  $r_{glu}$  values over five growth rates, were used as lower and upper bounds ( $a_i$  and  $b_i$ ) on the glucose exchange reaction (reaction ID: `r_1714` in model `rhto-GEM` and `EX_glc__D_e` in others; equation: extracellular D-glucose  $\rightleftharpoons$ ) to reach an allowable solution space in simulations for model constraining. Amino acid uptake was not allowed, as in experiments, from where the physiological data used in simulations is obtained, defined mineral medium was used.  $CO_2$  and  $O_2$  exchange rates were left unconstrained. The flux results of the five simulations, with different constrains on specific glucose uptake rate, were later concatenated into one dataframe.

Firstly, simulations were carried out on carbon limitation, with the objective function set to biomass maximisation (reaction ID: `r_4041` (in model `rhto-GEM`), `Biomass_Rt_Clim` (in `iRhtoC`), `BIOMASS_RT` (in `Rt_IFO0880`-based models)). Secondly, simulations with minimisation of non-growth associated maintenance (NGAM) reaction as an objective function were carried out (`r_4046`, `ATPM_c`, `ATPM`, respectively;  $ATP[c] + H_2O[c] \Rightarrow ADP[c] + H^+[c] + \text{phosphate}[c]$  (`[c]` indicates that the respective metabolite is in cytoplasm)). Solution space was constrained by setting upper and lower bounds to glucose exchange and biomass reaction. 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 model was optimized for biomass maximisation. On average, model overestimation of glucose uptake per growth rate was around 25% of experimental glucose uptake rate and it was considered a reasonable assumption in this case.

For comparison of cofactor balances between the models, pie plots were made using Matplotlib to visualize the production and consumption of NADPH. Cofactor balances were visualized for simulations with both, biomass maximisation and NGAM minimisation, as the objective function. Information about production and consumption of specific cofactors, was extracted from solution fluxes using the COBRApy functions:

- (i) `model.metabolites.metabolite.summary().producing_flux` and
- (ii) `model.metabolites.metabolite.summary().consuming_flux`.

These functions filter reactions containing the selected metabolite and provide an overview of all producing/consuming flux rates involving that selected metabolite, respectively. Reactions that had the same absolute value in producing and consuming fluxes were excluded. Results were plotted on a pie chart. The reactions that had a lower proportion than 2.2% of the total flux of the corresponding cofactor, were summed together and represented in the sector ‘Other consuming’ or ‘Other producing’, respectively.

## 4. Results

*Rhodotorula toruloides* can naturally accumulate high amounts of lipids, but the metabolic principles that make this possible and differentiate, if true, *R. toruloides* from other oleaginous and 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 adequate. 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. Biomass maximisation as an objective function

Firstly, simulations with biomass maximization as an objective function were carried out. The solutions were constrained over five experimental glucose uptake rates - 0.476, 1.114, 1.648, 2.305 and 3.1 mmol/gDW/h (Table 3.1). All simulated fluxes are available on a Github repository ([https://github.com/maivehanni/BSc\\_thesis/tree/main/All\\_simulated\\_fluxes](https://github.com/maivehanni/BSc_thesis/tree/main/All_simulated_fluxes)). In further analysis, selected exchange and intracellular fluxes were in the focus, and they are visualized over the biomass growth rate and shown below.

To explore the intracellular flux patterns over increasing growth rates, especially the pathways that generate acetyl coenzyme A in glucose catabolic pathways, all figures show the fluxes over growth rates from 0.05 to 0.25 1/h, that is the range predicted by the models, when constrained over experimentally measured specific glucose uptakes. For the investigation of the predicted source of NADPH regeneration in metabolism, NADPH producing and consuming fluxes were visualized on pie charts on growth rates 0.05 – 0.25 1/h, but only the ones that significantly differed on different growth rates, are shown below.

#### **rhto-GEM**

Model rhto-GEM estimates that the biomass growth rate, when glucose uptake is constrained to the experimentally measured specific glucose uptake rate, is lower compared to the experimental

specific growth rate during the same glucose uptake rate, reaching only 0.25 1/h in simulations *versus* 0.3 1/h experimentally. As the fluxes are visualized in the graphs over the growth rate, this is the reason why there are no predicted fluxes on growth rate over 0.25 1/h.

Model predicted that exchange fluxes are higher than experimentally measured per growth rate (Figure 4.1.a). Model predicted that the fluxes of specific O<sub>2</sub> uptake and CO<sub>2</sub> secretion are 1.5 – 7.6 and 1.6 – 8.4 mmol/gDW/h per growth rates from 0.03 to 0.25 1/h, respectively. Whereas physiological data from experiments showed that the exchange fluxes of O<sub>2</sub> and CO<sub>2</sub> were 1.1 – 6.3 and 1.2 – 7.4 mmol/gDW/h, per growth rates from 0.05 to 0.3 1/h respectively. This result was expected as conventional GEMs (without resource allocation constraints, such as enzyme constraints) are known to overestimate exchange fluxes.

All the investigated intracellular fluxes were predicted by the model to increase linearly over the predicted range of specific growth rates (4.1.b). Flux through glucose 6-phosphate dehydrogenase oxPPP was predicted to increase from 0.18 to 1.45 mmol/gDW/h, representing around 48% of the carbon from D-glucose 6-phosphate branching point. From the D-xylulose 5-phosphate branching point, 32% of the flux was predicted to go to transketolase 1 (fluxes from 0.04 to 0.34 mmol/gDW/h), 28% to transketolase 2 (0.04 – 0.3 mmol/gDW/h) and 40% to phosphoketolase (fluxes from 0.05–0.4 mmol/gDW/h). The flux of transaldolase was predicted to be zero on all rates. Fructose biphosphate aldolase was predicted to increase from 0.28 to 1.15 mmol/gDW/h, representing 82% of the carbon from D-fructose 6-phosphate. Pyruvate decarboxylase represents 15% and pyruvate dehydrogenase 59% of the carbon from cytosolic pyruvate branching point having fluxes from 0.06 to 0.5 and 0.46 – 2.25 mmol/gDW/h, respectively (Figure 4.1.b). The model did not predict any flux for ACL, instead it predicts that acetyl-CoA is produced through phosphoketolase.

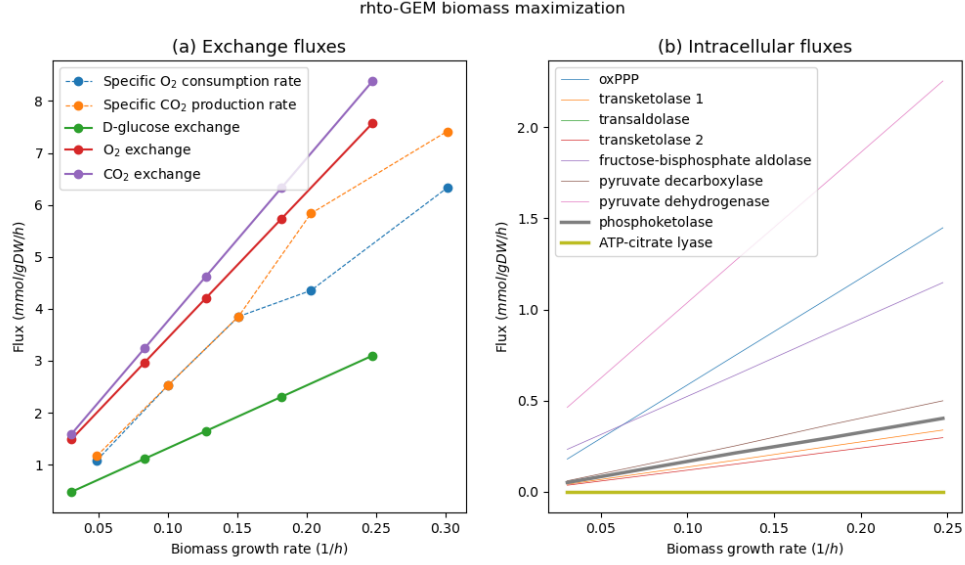


Figure 4.1: Simulated exchange (a) and intracellular (b) fluxes in *R. toruloides* with model rhto-GEM optimized for biomass maximization and constrained over five specific glucose uptake rates. Exchange fluxes plot (a) shows the predicted specific glucose uptake (D-glucose exchange), specific oxygen consumption (O<sub>2</sub> exchange) and specific carbon dioxide production rate (CO<sub>2</sub> exchange). The experimentally measured exchange rates of oxygen (Specific O<sub>2</sub> consumption rate) and carbon dioxide (Specific CO<sub>2</sub> production rate) are visualized with dashed lines. Intracellular fluxes graph (b) shows the fluxes of glucose 6-phosphate dehydrogenase (oxPPP), transketolase 1, transaldolase, transketolase 2, fructose-bisphosphate aldolase, pyruvate decarboxylase, pyruvate dehydrogenase, and in bold the fluxes of phosphoketolase and ATP-citrate lyase.

For NADPH production and consumption there are no differences between different growth rates. In all cases the model predicts that around 90% of the NADPH is produced by glucose 6-phosphate dehydrogenase and phosphogluconate dehydrogenase (oxPPP). Ca 6% is produced by methylenetetrahydrofolate dehydrogenase (Figure 4.2). NADPH is primarily consumed by fatty acid synthases 50% (fatty-acyl-CoA synthase (n-C16:0CoA) and fatty-acyl-CoA synthase (n-C18:0CoA)) and glutamate dehydrogenase by 30%.



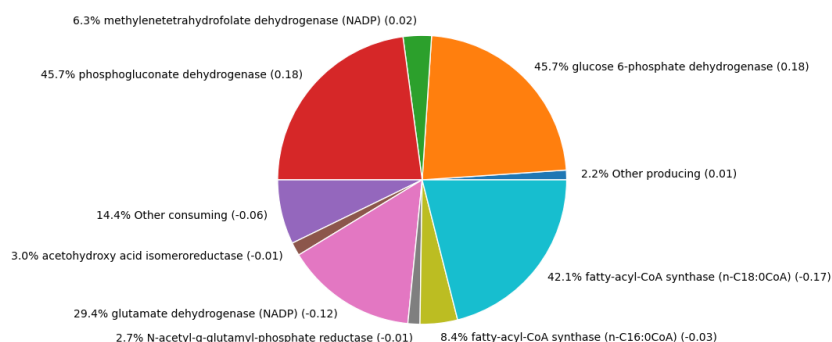


Figure 4.2: Simulated NADPH producing and consuming fluxes in *R. toruloides* with model rhto-GEM optimized for biomass maximization. Glucose uptake was constrained to the lowest rate. The upper part of the pie shows producing fluxes and the lower part shows consuming fluxes. The flux of the enzyme is shown between the brackets after the name of the metabolite, and in front of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux, respectively.

## iRhtoC

iRhtoC model predicts very similar exchange fluxes as the previous model, predicting higher  $O_2$  and  $CO_2$  rates than experimentally measured (Figure 4.3.a). Model estimated that over the predicted range of specific growth rates, specific  $O_2$  uptake and  $CO_2$  production are 1.3 – 7.9 and 1.4 – 8.8 mmol/gDW/h, respectively. Intracellular fluxes are predicted to increase linearly over increasing growth rates. However, the predictions of intracellular fluxes by iRhtoC differ from model rhto-GEM. iRhtoC predicts that over the five growth rates the flux through oxPPP increase from 0.24 to 1.69 mmol/gDW/h (55% from D-glucose 6-phosphate branching point), which is slightly more than rhto-GEM predicted. Fluxes of transketolase 1 and 2 are also a bit higher than predicted by rhto-GEM (0.08 – 0.56 and 0.07 – 0.5 mmol/gDW/h representing 53% and 47% from D-xylulose 5-phosphate branching point, respectively). Fluxes of transaldolase and pyruvate decarboxylase are predicted to be zero. Fructose-bisphosphate aldolase (representing 93% from D-fructose 6-phosphate) and pyruvate dehydrogenase (80% from cytosolic pyruvate branching point) fluxes are from 0.24 to 1.49 and from 0.53 to 3.29 mmol/gDW/h, respectively. ATP-citrate lyase has a flux from 0.18 to 1.29 mmol/gDW/h representing 40% of carbon from mitochondrial citrate. The flux of phosphoketolase is predicted to be zero (Figure 4.3.b).

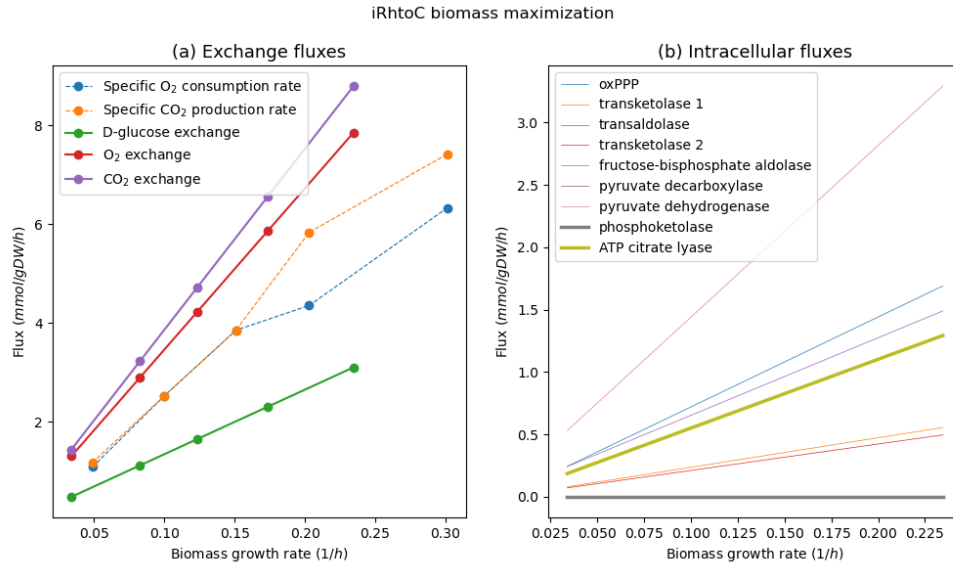


Figure 4.3: Simulated exchange (a) and intracellular (b) fluxes in *R. toruloides* with model iRhtoC optimized for biomass maximization and constrained over five specific glucose uptake rates. Exchange fluxes plot (a) shows the predicted specific glucose uptake (D-glucose exchange), specific oxygen consumption (O<sub>2</sub> exchange) and specific carbon dioxide production (CO<sub>2</sub> exchange) rate. Intracellular fluxes graph (b) shows the fluxes of glucose 6-phosphate dehydrogenase (oxPPP), transketolase 1, transaldolase, transketolase 2, fructose-bisphosphate aldolase, pyruvate decarboxylase, pyruvate dehydrogenase, and in bold the fluxes of phosphoketolase and ATP-citrate lyase.

This model predicts that almost 90% of NADPH is produced by glucose 6-phosphate dehydrogenase and phosphogluconate dehydrogenase (oxPPP) on all rates, but around 6% is produced by malic enzyme on growth rates 0.03 and 0.23 mmol/gDW/h and on other rates the model predicts isocitrate dehydrogenase instead (Figures 4.4 and 4.5). Predicted consuming fluxes of NADPH do not differ between growth rates. Similarly to model rhto-GEM, it is predicted that NADPH is consumed by glutamate dehydrogenase by 35%. This model predicts lower use of fatty acid synthesis system (FAS), which is 25%. (In this model, the enzymes of fatty acid synthesis system are distinguished in more detail than in rhto-GEM, resulting in 16 FAS enzymes. Each respective enzyme's flux is around 1.5% of total consumption flux of NADPH and they were included in the sector 'Other consuming' in the pie chart, as 2.2% was used as a cut off for including fluxes independently in the pie).

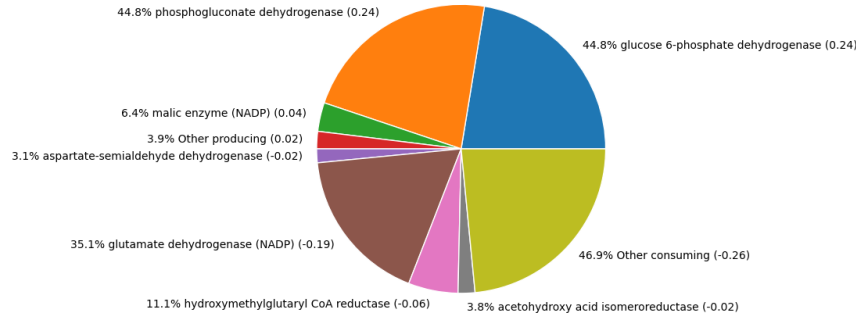


Figure 4.4: Simulated NADPH producing and consuming fluxes in *R. toruloides* with model iRhtC. The model was optimized for biomass maximization and glucose uptake was constrained to the lowest rate. The fluxes are same when glucose uptake is constrained to highest rate. The upper part of the pie shows producing fluxes and the lower part shows consuming fluxes. The flux of the enzyme is shown between the brackets after the name of the metabolite, and in front of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux, respectively.

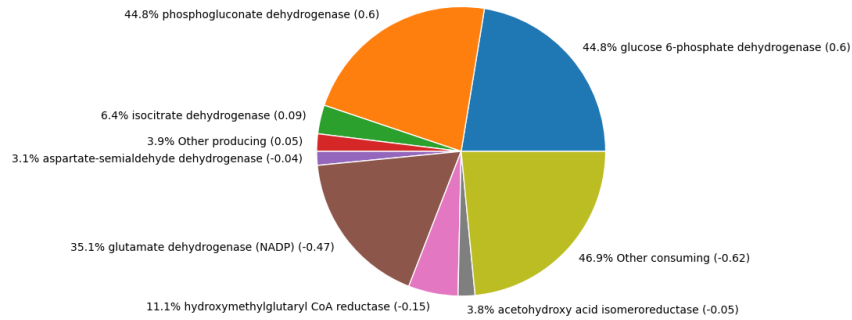


Figure 4.5: Simulated NADPH producing and consuming fluxes in *R. toruloides* with model iRhtC. The model was optimized for biomass maximization. Results are the same when glucose uptake is constrained to 1.114, 1.648 or 2.305 mmol/gDW/h. The upper part of the pie shows producing fluxes and the lower part shows consuming fluxes. The flux of the enzyme is shown between the brackets after the name of the metabolite, and in front of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux, respectively.

## Rt\_IFO0880

Rt\_IFO0880 also predicts higher exchange fluxes than experimentally measured (Figure 4.6.a). Model predicted that over the predicted range of specific growth rates, specific O<sub>2</sub> uptake and CO<sub>2</sub> production fluxes are 0.5 – 3.1, 1.1 – 6.8 and 1.3 – 7.7 mmol/gDW/h, respectively. 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. Model predictions differ from models rhto-GEM and iRhtoC by not having any flux in oxPPP and transketolase 2. Transketolase 1 represents 13% of the carbon from glyceraldehyde 3-phosphate. Transaldolase represents 19%, fructose-bisphosphate aldolase 52% and FPK 21% carbon from D-glucose 6-phosphate. XPK represents 100% of carbon from D-xylulose 5-phosphate. Summed flux of XPK and FPK is from 0.18 to 1.25 mmol/gDW/h over the rates. Pyruvate decarboxylase represents 7% and pyruvate dehydrogenase 58% of carbon from cytosolic pyruvate branching point (Figure 4.6.b). Similarly to rhto-GEM, this model also predicts the use of phosphoketolase and no use of ACL.

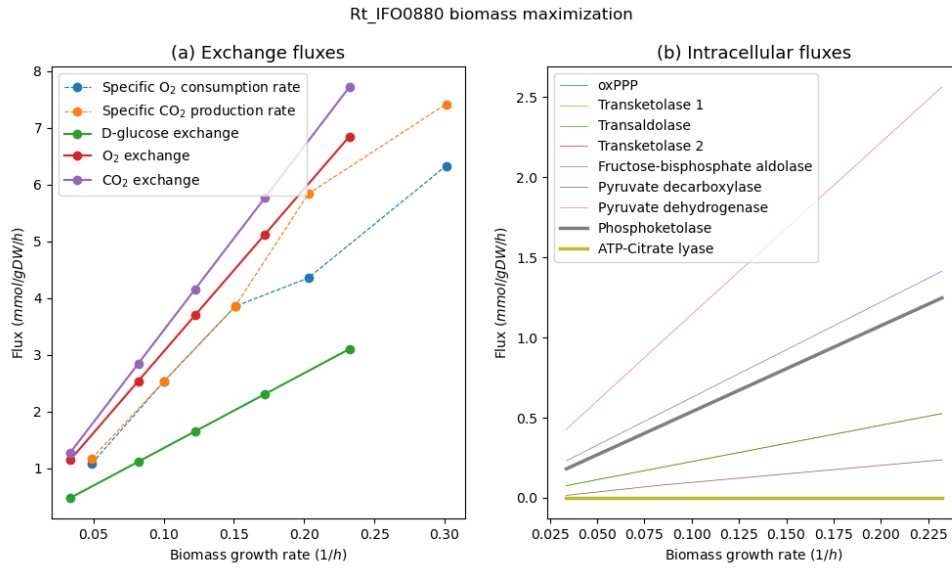


Figure 4.6: Simulated exchange (a) and intracellular (b) fluxes in *R. toruloides* with model Rt\_IFO0880 optimized for biomass maximization and constrained over five specific glucose uptake rates. Exchange fluxes plot (a) shows the predicted specific glucose uptake (D-glucose exchange), specific oxygen consumption (O<sub>2</sub> exchange) and specific carbon dioxide production (CO<sub>2</sub> exchange) rate. Intracellular fluxes graph (b) shows the fluxes of glucose 6-phosphate dehydrogenase (oxPPP), transketolase 1, transaldolase, transketolase 2, fructose-bisphosphate aldolase, pyruvate decarboxylase, pyruvate dehydrogenase, and in bold the fluxes of phosphoketolase and ATP-citrate lyase.

In NADPH production and consumption on different rates there are no significant differences. Around 90% of NADPH is produced by alcohol dehydrogenase (ID: ALCD2y, Nicotinamide adenine dinucleotide phosphate[c] + Ethanol[c] => H<sup>+</sup>[c] + Nicotinamide adenine dinucleotide phosphate - reduced[c] + Acetaldehyde[c]) on all rates (Figure 4.7). This is very different from other models. On all rates, NADPH is consumed by glutamate dehydrogenase by around 35% similarly to previous models. Around 7% is consumed by fatty acyl CoA synthase n C80CoA lumped reaction and 10% is the summed flux of five other fatty acyl CoA synthases (n C100CoA, n C120CoA, n C140CoA, n C160CoA and n C180CoA). The total flux of FAS is around 17% and is comparable to iRhtoC.

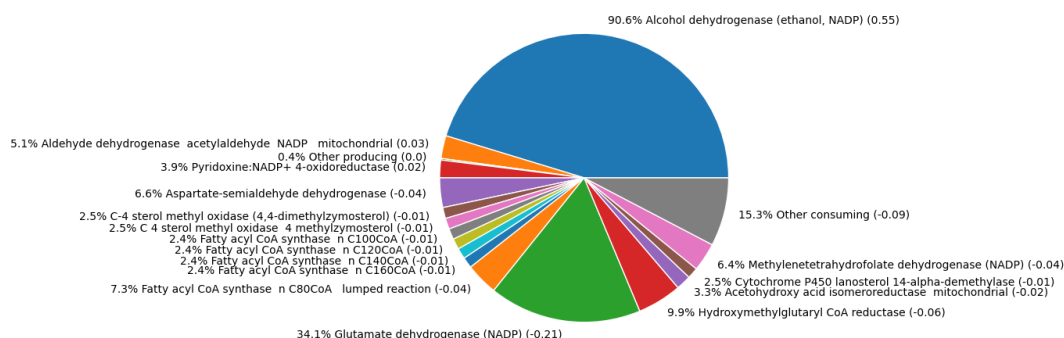


Figure 4.7: Simulated NADPH producing and consuming fluxes in *R. toruloides* with model Rt\_IFO0880. The model was optimized for biomass maximization and glucose uptake was constrained. The upper part of the pie shows producing fluxes and the lower part shows consuming fluxes. The flux of the enzyme is shown between the brackets after the name of the metabolite, and in front of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux, respectively.

## Rt\_IFO0880\_LEBp2023

Exchange flux predictions are almost the same as with model Rt\_IFO0880. Model predicted that over the predicted range of specific growth rates, specific  $O_2$  uptake and  $CO_2$  production fluxes are 1.1 – 6.8 and 1.3 – 7.7 mmol/gDW/h, respectively (Figure 4.8.a). From D-glucose 6-phosphate branching point 9% of the carbon is represented by oxPPP. Transketolase 1 and 2 represent very low carbon percentage, only 0.3% and 1.3%, from glyceraldehyde 3-phosphate branching point, respectively. Transaldolase represents 0.6%, transketolase 2 2% and fructose-bisphosphate aldolase 89% from D-glucose 6-phosphate. Pyruvate decarboxylase represents 5% and pyruvate dehydrogenase 62% carbon from cytosolic pyruvate. 33% of carbon from mitochondrial citrate is represented by ACL. (4.8.B) On all growth rates the use of ACL is predicted instead of phosphoketolase, which is similar with the model iRhtoC. (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.)

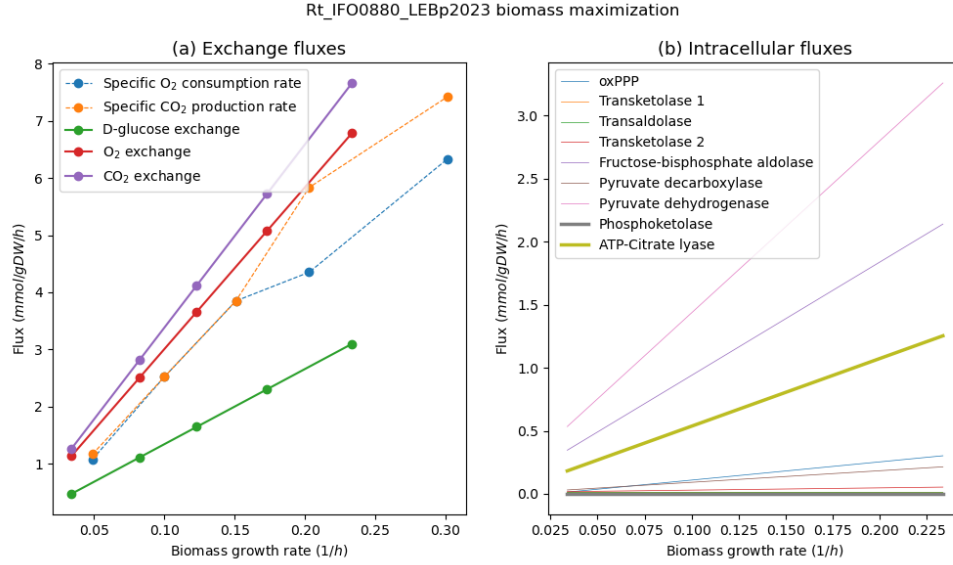


Figure 4.8: Simulated exchange (a) and intracellular (b) fluxes in *R. toruloides* with model Rt\_IFO0880\_LEBp2023 optimized for biomass maximization and constrained over five specific glucose uptake rates. Exchange fluxes plot (a) shows the predicted specific glucose uptake (D-glucose exchange), specific oxygen consumption (O<sub>2</sub> exchange) and specific carbon dioxide production (CO<sub>2</sub> exchange) rate. Intracellular fluxes graph (b) shows the fluxes of glucose 6-phosphate dehydrogenase (oxPPP), transketolase 1, transaldolase, transketolase 2, fructose-bisphosphate aldolase, pyruvate decarboxylase, pyruvate dehydrogenase, and in bold the fluxes of phosphoketolase and ATP-citrate lyase.

On lowest growth rate, NADPH is predicted to be produced mainly ( $\sim 85\%$ ) by aldehyde dehydrogenase (ALDD19xr, Phenylacetaldehyde[c] + H<sub>2</sub>O H<sub>2</sub>O[c] + Nicotinamide adenine dinucleotide[c]  $\rightleftharpoons$  2 H<sup>+</sup>[c] + Nicotinamide adenine dinucleotide - reduced[c] + Phenylacetic acid[c]) and on all other rates mainly ( $\sim 75\%$ ) by alcohol dehydrogenase and  $\sim 15\%$  by glucose-6-phosphate dehydrogenase and phosphogluconate dehydrogenase (Figures 4.9 and 4.10). Interestingly, as the growth rate increases the use of aldehyde dehydrogenase slightly decreases and use of oxidative pentose phosphate pathway for NADPH production increases. On all growth rates, NADPH is primarily consumed by glutamate dehydrogenase (35%) and FAS (20%) (fatty acyl CoA synthase n C80CoA lumped reaction, fatty acyl CoA synthase n C100CoA, C120CoA, C140CoA and C160CoA).

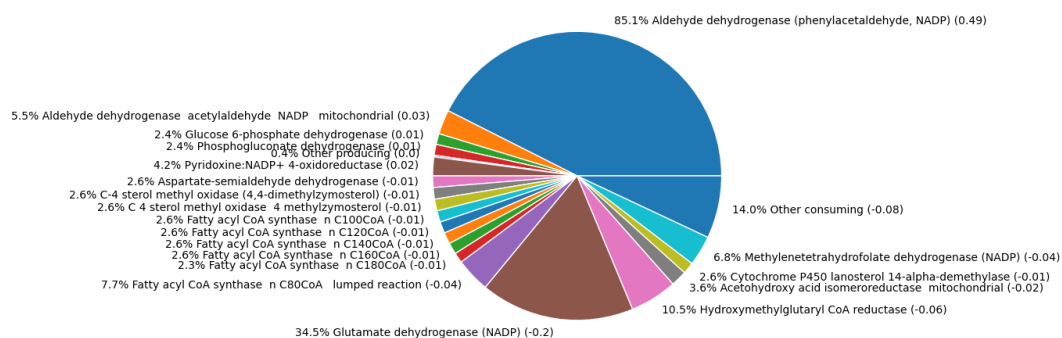


Figure 4.9: Simulated NADPH producing and consuming fluxes in *R. toruloides* with model Rt\_IFO0880\_LEBp2023. The model was optimized for biomass maximization. Glucose uptake was constrained on the lowest rate. The upper part of the pie shows producing fluxes and the lower part shows consuming fluxes. The flux of the enzyme is shown between the brackets after the name of the metabolite, and in front of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux, respectively.

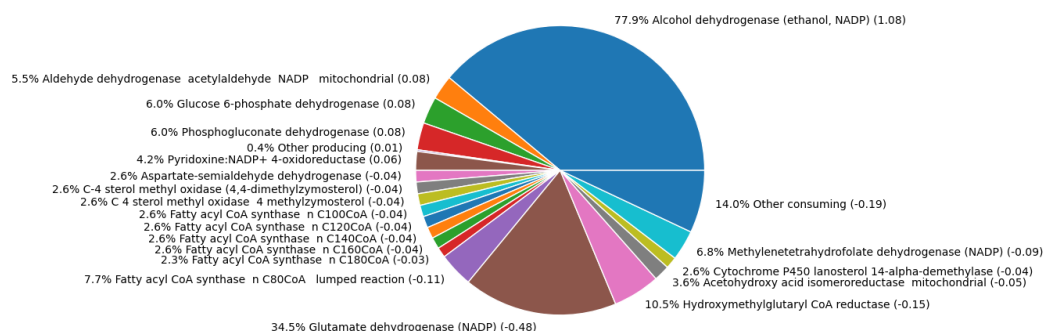


Figure 4.10: Simulated NADPH producing and consuming fluxes in *R. toruloides* with model Rt\_IFO0880\_LEBp2023. The model was optimized for biomass maximization and glucose uptake constrained to 1.114. The results are very similar when glucose uptake is constrained to 1.114, 1.648, 2.305 or 3.1. The upper part of the pie shows producing fluxes and the lower part shows consuming fluxes. The flux of the enzyme is shown between the brackets after the name of the metabolite, and in front of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux, respectively.

Simulations with biomass maximisation as an objective function under carbon limitation, with glucose as the sole source of carbon, 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) predicted the production of acetyl-CoA using ACL. In many studies, ACL has been found to be present in oleaginous yeast, whereas phosphoketolase has been

predicted to be active in *R. toruloides* during growth on xylose by [55, 56] and on glucose by [55], but in both studies protein levels of ACL were also reported and in [56] ACL was found in even higher levels under nitrogen limitation, when lipid accumulation was increased. In experiments where phosphoketolase pathways were transferred to the oleaginous model yeast *Y. lipolytica*, the lipid production in the engineered yeast increased by 53% when grown on glucose [61].

NADPH sources also differ among model predictions. On all growth rates, models rhto-GEM and iRhtoC predict that most of the NADPH is produced by the oxidative pentose phosphate pathway and  $\sim 6\%$  by ethylenetetrahydrofolate dehydrogenase in rhto-GEM and by malic enzyme in iRhtoC. Model Rt\_IFO0880 predicts that most of the NADPH is produced by alcohol dehydrogenase and Rt\_IFO0880\_LEBp2023 predicts that on the lowest growth rate NADPH is mainly ( $\sim 85\%$ ) produced through aldehyde dehydrogenase and on all other rates,  $\sim 75\%$  is produced by alcohol dehydrogenase and  $\sim 15\%$  by oxPPP. Malic enzyme, which is considered a key enzyme in the regeneration of NADPH for lipid biosynthesis [25], was not predicted to be the primary NADPH producer in any of the models. However, this is in accordance with the proteomics data of *R. toruloides* grown on xylose by [56], where it was found that most of the NADPH is produced in oxPPP. It was also suggested by Wasylenko et al. 2015 that in *Yarrowia lipolytica* the oxidative pentose phosphate pathway is the primary source of NADPH during lipid accumulation on growth on glucose [62]. Still, the results by Li et al. 2012 indicated that the lipid content in *Rhodotorula glutinis* grown on glucose increased from 18.7% to 39.4% thanks to overexpression of malic enzyme, indicating importance of ME [63]. These results in this study as well as in others, demonstrate the carbon source-dependent differences in yeast metabolism.

## 4.2. Non-growth associated maintenance minimisation as an objective function

To see whether different objective function changes the flux patterns, metabolic flux patterns were further investigated using the minimisation of non-growth associated maintenance (NGAM) reaction as an objective function. Specifically, whether Rt\_IFO0880-based models could re-arrange NADPH regeneration through different pathways than alcohol and aldehyde dehydrogenase. When optimizing for NGAM minimisation, it is assumed that cells strive for satisfying physiological parameters with least energy expenditure as NGAM minimisation decreases the ATP demand.

Same experimentally measured specific glucose uptake rates were used throughout the simulations (Table 3.1). This objective function also needed constraints on biomass growth rate because otherwise the simulation chooses zero as its flux. Experimentally measured specific growth rates together with experimental specific glucose uptakes were infeasible for the models. Because of that, growth rate was constrained to the values that each model predicted in the simulations optimized for biomass maximisation. These values



slightly varied between the models, but when rounded were 0.03, 0.08, 0.12-0.13, 0.17-0.18 and 0.23-0.25 1/h. All simulated fluxes are available on a Github repository ([https://github.com/maivehanni/BSc\\_thesis/tree/main/All\\_simulated\\_fluxes](https://github.com/maivehanni/BSc_thesis/tree/main/All_simulated_fluxes)).

As a result, model Rt\_IFO0880 optimized for NGAM minimization predicted the production of NADPH differently - around 90% of NADPH is produced by homoserine dehydrogenase on rates 0.08 and 0.17 1/h and on other rates by alcohol dehydrogenase (Figures 4.11 and 4.12). NADH consuming fluxes did not change, like previously, majority is being consumed by glutamate dehydrogenase and 17% by FAS.

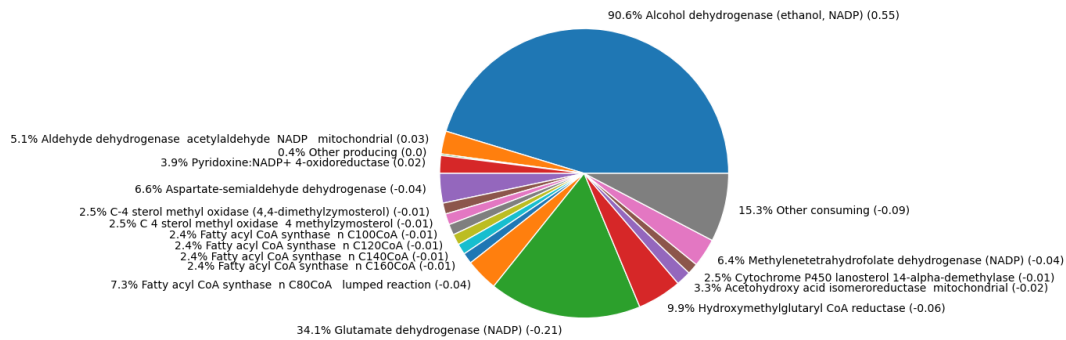


Figure 4.11: Simulated NADPH producing and consuming fluxes in *R. toruloides* with model Rt\_IFO0880. The model was optimized for NGAM minimization. Glucose uptake and growth rate were constrained on the lowest rate. The results are the same when glucose uptake is constrained to 1.648 or 3.1 mmol/gDW/h. The upper part of the pie shows producing fluxes and the lower part shows consuming fluxes. The flux of the enzyme is shown between the brackets after the name of the metabolite, and in front of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux, respectively.

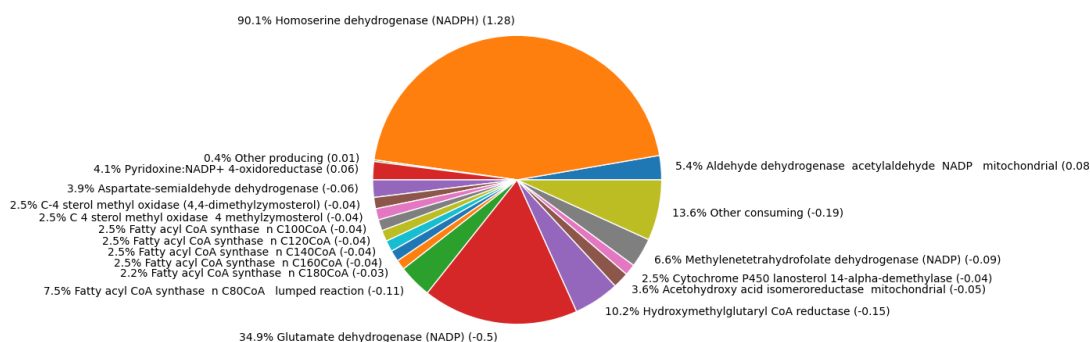


Figure 4.12: Simulated NADPH producing and consuming fluxes in *R. toruloides* with model Rt\_IFO0880. The model was optimized for NGAM minimization, growth rate and glucose uptake were constrained to 0.08 1/h and 1.114 mmol/gDW/h, respectively. The results are similar when glucose uptake is constrained to 2.305 mmol/gDW/h. The upper part of the pie shows producing fluxes and the lower part shows consuming fluxes. The flux of the enzyme is shown between the brackets after the name of the metabolite, and in front of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux, respectively.

Model Rt\_IFO0880\_LEBp2023 optimized for NGAM minimization also predicted the production of NADPH differently compared to the simulation with biomass maximization as the objective function. On rates 0.08, 0.12 and 0.17 1/h use of homoserine dehydrogenase is predicted instead of alcohol dehydrogenase (Figure 4.13). On lowest rate aldehyde dehydrogenase and on highest rate alcohol dehydrogenase is predicted, which is the same as previously, when biomass maximisation was used as an objective function. NADPH consuming fluxes do not differ from previous simulation results: 35% being consumed by glutamate dehydrogenase and 20% by FAS.

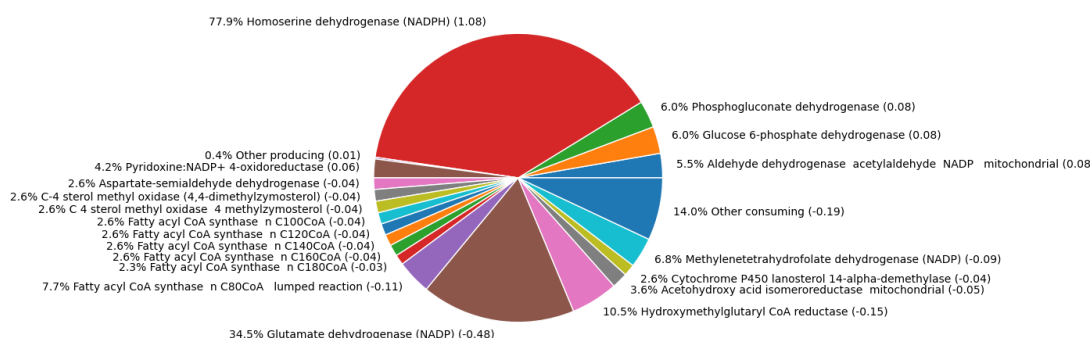


Figure 4.13: Simulated NADPH producing and consuming fluxes in *R. toruloides* with model Rt\_IFO0880\_LEBp2023. The model was optimized for NGAM minimization, growth rate and glucose uptake were constrained to 0.08 1/h and 1.114 mmol/gDW/h, respectively. The results are very similar when glucose uptake is constrained to 1.648 or 2.305 mmol/gDW/h. The upper part of the pie shows producing fluxes and the lower part shows consuming fluxes. The flux of the enzyme is shown between the brackets after the name of the metabolite, and in front of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux, respectively.

All in all, these simulations showed that Rt\_IFO0880-based models could re-arrange NADPH regeneration through different pathways than alcohol and aldehyde dehydrogenase when NGAM minimization is used as an objective function. However, the predictions did not lead towards the oxidative part of pentose phosphate pathway, but instead, to homoserine dehydrogenase for majority of simulated biomass production rates (on 0.08 and 0.17 1/h for Rt\_IFO0880 and on 0.08, 0.12 and 0.17 1/h for Rt\_IFO0880\_LEBp2023). This result was unexpected and it is not clear what made the models predict alcohol and aldehyde dehydrogenase only on some growth rates.

## 5. Conclusion

In this work, four genome-scale metabolic models of *Rhodotorula toruloides* were compared during simulations with glucose as a sole carbon substrate uptake, using two different objective functions - maximisation of biomass and minimisation of NGAM. The fluxes through phosphoketolase and ACL pathways, that are found to be important for lipid production, were explored. Results demonstrated clear difference in model predictions in regard to the use of one (models that predicted XPK: rhto-GEM, Rt\_IFO0880) or the other pathway (models that predicted ACL: iRhtoC and Rt\_IFO0880\_LEBp2023), respectively, and this did not change with different objective functions.

The consuming fluxes of NADPH also differed between the models. Two models (rhto-GEM and iRhtoC) predicted that most of the NADPH is produced by the oxidative part of pentose phosphate pathway and Rt\_IFO0880-based models predicted NADPH production mainly by alcohol or aldehyde dehydrogenase when model was optimized for biomass maximisation, whereas when optimized for NGAM minimisation, homoserine dehydrogenase was predicted on most growth rates instead. Only model iRhtoC predicted that malic enzyme was involved in the production of NADPH, but that corresponded to only 6% of total produced NADPH.

This study gave more insight about how the models rhto-GEM, iRhtoC, Rt\_IFO0880 and Rt\_IFO0880\_LEBp2023 predict central carbon metabolism of *R. toruloides* than previously found in the literature. Still, as the models lead to different understanding of the production of lipid precursors acetyl-CoA and NADPH in *R. toruloides*, there is a need for a consensus genome-scale metabolic model of *R. toruloides* for obtaining more accurate predictions of *R. toruloides* phenotype.

## Acknowledgements

I would like to express my deepest gratitude to my supervisor Alīna Reķēna for proposing the thesis topic and her invaluable guidance and feedback throughout the practical work and writing process. I am also very thankful to the Food Tech and Bioengineering research group leader Petri-Jaan Lahtvee, who generously provided his knowledge and expertise for further defining the aim and structure of the work. Lastly, I am grateful to my family and friends, and especially my partner. Their belief and support in me has maintained my motivation high throughout this process.

## References

- [1] 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.
- [2] Bonturi, N. et al. “Microbial oil production in sugarcane bagasse hemicellulosic hydrolysate without nutrient supplementation by a *Rhodospiridium toruloides* adapted strain”. In: *Process Biochemistry* 57 (2017), pp. 16–25. DOI: 10.1016/j.procbio.2017.03.007.
- [3] 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.
- [4] Dinh, H. V. et al. “A comprehensive genome-scale model for *Rhodospiridium toruloides* IFO0880 accounting for functional genomics and phenotypic data”. In: *Metabolic Engineering Communications* 9 (2019), e00101. DOI: 10.1016/j.mec.2019.e00101.
- [5] Kim, J. et al. “Multi-Omics Driven Metabolic Network Reconstruction and Analysis of Lignocellulosic Carbon Utilization in *Rhodospiridium toruloides*”. In: *Frontiers in Bioengineering and Biotechnology* 8 (2021). DOI: 10.3389/fbioe.2020.612832.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] Durrett, T. P., Benning, C., and Ohlrogge, J. “Plant triacylglycerols as feedstocks for the production of biofuels”. In: *The Plant Journal* 54.4 (2008), pp. 593–607. DOI: 10.1111/j.1365-3113x.2008.03442.x.
- [10] Rosillo-Calle, F., Pelkmans, L., and Walter, A. “A global overview of vegetable oils, with reference to biodiesel”. In: *A report for the IEA Bioenergy Task 40* (2009).
- [11] 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.
- [12] Adrio, J. L. “Oleaginous yeasts: Promising platforms for the production of oleochemicals and biofuels”. In: *Biotechnology and Bioengineering* 114.9 (2017), pp. 1915–1920. DOI: 10.1002/bit.26337.

- [13] Li, Y., Zhao, Z. (, and Bai, F. “High-density cultivation of oleaginous yeast *Rhodospiridium 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.
- [14] Hu, C. et al. “Effects of biomass hydrolysis by-products on oleaginous yeast *Rhodospiridium toruloides*”. In: *Bioresource Technology* 100.20 (2009), pp. 4843–4847. DOI: 10.1016/j.biortech.2009.04.041.
- [15] Wu, C.-C., Honda, K., and Kazuhito, F. “Current advances in alteration of fatty acid profile in *Rhodotorula toruloides*: a mini-review”. In: *World Journal of Microbiology and Biotechnology* 39.9 (2023). DOI: 10.1007/s11274-023-03595-3.
- [16] Park, Y.-K., Nicaud, J.-M., and Ledesma-Amaro, R. “The Engineering Potential of *Rhodospiridium toruloides* as a Workhorse for Biotechnological Applications”. In: *Trends in Biotechnology* 36.3 (2018), pp. 304–317. DOI: 10.1016/j.tibtech.2017.10.013.
- [17] 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.
- [18] Buzzini, P. et al. “Carotenoid profiles of yeasts belonging to the genera *Rhodotorula*, *Rhodospiridium*, *Sporobolomyces*, and *Sporidiobolus*”. In: *Canadian Journal of Microbiology* 53.8 (2007), pp. 1024–1031. DOI: 10.1139/w07-068.
- [19] 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.
- [20] Tehlivets, O., Scheuringer, K., and Kohlwein, S. D. “Fatty acid synthesis and elongation in yeast”. In: *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1771.3 (2007), pp. 255–270. DOI: 10.1016/j.bbalip.2006.07.004.
- [21] Zhu, Z. et al. “A multi-omic map of the lipid-producing yeast *Rhodospiridium toruloides*”. In: *Nature Communications* 3.1 (2012). DOI: 10.1038/ncomms2112.
- [22] 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.
- [23] Tiukova, I. A. et al. “Proteome analysis of xylose metabolism in *Rhodotorula toruloides* during lipid production”. In: *Biotechnology for Biofuels* 12.1 (2019). DOI: 10.1186/s13068-019-1478-8.
- [24] Evans, C. T. and Ratledge, C. “Induction of xylulose-5-phosphate phosphoketolase in a variety of yeasts grown on d-xylose: the key to efficient xylose metabolism”. In: *Archives of Microbiology* 139.1 (1984), pp. 48–52. DOI: 10.1007/bf00692711.
- [25] 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.

- [26] Fakas, S. “Lipid biosynthesis in yeasts: A comparison of the lipid biosynthetic pathway between the model nonoleaginous yeast *Saccharomyces cerevisiae* and the model oleaginous yeast *Yarrowia lipolytica*”. In: *Engineering in Life Sciences* 17.3 (2016), pp. 292–302. DOI: 10.1002/elsc.201600040.
- [27] 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.
- [28] Yang, Y. and Sha, M. *A Beginner’s Guide to Bioprocess Modes – Batch, Fed-Batch, and Continuous Fermentation*. Eppendorf Inc., Application Note No. 408. Contact: bioprocess-experts@eppendorf.com. 2019.
- [29] University, O. S. “Microbial Growth”. In: [Online; accessed 2024-05-14]. 2021.
- [30] Stephanopoulos, G. N., Aristidou, A. A., and Nielsen, J. “Material Balances and Data Consistency”. In: *Metabolic Engineering*. Elsevier, 1998, pp. 115–146. DOI: 10.1016/b978-012666260-3/50005-4.
- [31] Nidelet, T. et al. “Diversity of flux distribution in central carbon metabolism of *S. cerevisiae* strains from diverse environments”. In: *Microbial Cell Factories* 15.1 (2016). DOI: 10.1186/s12934-016-0456-0.
- [32] Nielsen, J. “It Is All about MetabolicFluxes”. In: *Journal of Bacteriology* 185.24 (2003), pp. 7031–7035. DOI: 10.1128/jb.185.24.7031-7035.2003.
- [33] Palsson, B. “Metabolic systems biology”. In: *FEBS Letters* 583.24 (2009), pp. 3900–3904. DOI: 10.1016/j.febslet.2009.09.031.
- [34] 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.
- [35] 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.
- [36] Price, N. D., Reed, J. L., and Palsson, B. Ø. “Genome-scale models of microbial cells: evaluating the consequences of constraints”. In: *Nature Reviews Microbiology* 2.11 (2004), pp. 886–897. DOI: 10.1038/nrmicro1023.
- [37] 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.
- [38] 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.
- [39] 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.



- [40] 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.
- [41] Feist, A. M. et al. “A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information”. In: *Molecular Systems Biology* 3.1 (2007). DOI: 10.1038/msb4100155.
- [42] 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.
- [43] Kerkhoven, E. et al. *SysBioChalmers/RAVEN: v2.9.0*. 2024. DOI: 10.5281/ZENODO.11107015.
- [44] 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.
- [45] Price, N. D., Schellenberger, J., and Palsson, B. O. “Uniform Sampling of Steady-State Flux Spaces: Means to Design Experiments and to Interpret Enzymopathies”. In: *Biophysical Journal* 87.4 (2004), pp. 2172–2186. DOI: 10.1529/biophysj.104.043000.
- [46] Bordel, S., Agren, R., and Nielsen, J. “Sampling the Solution Space in Genome-Scale Metabolic Networks Reveals Transcriptional Regulation in Key Enzymes”. In: *PLoS Computational Biology* 6.7 (2010). Ed. by J. L. Reed, e1000859. DOI: 10.1371/journal.pcbi.1000859.
- [47] Biotechnology Information, N. C. for. *Genome assembly RHOziaDV1.0 for Rhodotorula toruloides NP11*. Accessed: 2024-05-15. 2013. URL: [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_000320785.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000320785.1/) (visited on 05/15/2024).
- [48] Sánchez, B. J. et al. “SLIMER: probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework”. In: *BMC Systems Biology* 13.1 (2019). DOI: 10.1186/s12918-018-0673-8.
- [49] Coradetti, S. T. et al. “Functional genomics of lipid metabolism in the oleaginous yeast *Rhodospiridium toruloides*”. In: *eLife* 7 (2018). DOI: 10.7554/elife.32110.
- [50] Joint Genome Institute, M. F. P. *Rhodospiridium toruloides IFO0880 v4.0*. Accessed: 2024-05-15. 2018. URL: [https://mycocosm.jgi.doe.gov/Rhoto\\_IF00880\\_4/Rhoto\\_IF00880\\_4.home.html](https://mycocosm.jgi.doe.gov/Rhoto_IF00880_4/Rhoto_IF00880_4.home.html) (visited on 05/15/2024).
- [51] 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.
- [52] 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.

- [53] Jagtap, S. S. and Rao, C. V. “Production of d-arabitol from d-xylose by the oleaginous yeast *Rhodospiridium toruloides* IFO0880”. In: *Applied Microbiology and Biotechnology* 102.1 (2017), pp. 143–151. DOI: 10.1007/s00253-017-8581-1.
- [54] 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.
- [55] Reḱēna, 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.
- [56] 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.
- [57] Lahtvee, P.-J. et al. “Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast”. In: *Cell Systems* 4.5 (2017), 495–504.e5. DOI: 10.1016/j.cels.2017.03.003.
- [58] Verduyn, C. et al. “Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation”. In: *Yeast* 8.7 (1992), pp. 501–517. DOI: 10.1002/yea.320080703.
- [59] Shen, H. et al. “Kinetics of continuous cultivation of the oleaginous yeast *Rhodospiridium toruloides*”. In: *Journal of Biotechnology* 168.1 (2013), pp. 85–89. DOI: 10.1016/j.jbiotec.2013.08.010.
- [60] 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.
- [61] Xu, P. et al. “Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals”. In: *Proceedings of the National Academy of Sciences* 113.39 (2016), pp. 10848–10853. DOI: 10.1073/pnas.1607295113.
- [62] Wasylenko, T. M., Ahn, W. S., and Stephanopoulos, G. “The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in *Yarrowia lipolytica*”. In: *Metabolic Engineering* 30 (2015), pp. 27–39. DOI: 10.1016/j.ymben.2015.02.007.
- [63] Li, Z. et al. “Overexpression of malic enzyme (ME) of *Mucor circinelloides* improved lipid accumulation in engineered *Rhodotorula glutinis*”. In: *Applied Microbiology and Biotechnology* 97.11 (2012), pp. 4927–4936. DOI: 10.1007/s00253-012-4571-5.

# Abstract

The transition towards bioeconomy to reduce the dependence on fossil-based resources, necessitates innovative methods for producing chemicals and fuels from sustainable materials. The current production of biodiesel from oilseeds and waste oils is insufficient to meet the global demand of the biodiesel industry, highlighting the need for second-generation oleochemicals derived from non-edible sources. Microbial oils (SCOs), which utilize low-value waste streams and do not compete with the food sector, are a promising source of fatty acids for oleochemical production. The non-conventional, oleaginous yeast *Rhodotorula toruloides* is one of the most promising yeasts for bioproduction of oleochemicals. Although its metabolic pathways that enable lipid production are defined, the way of enabling such a high lipid production is not fully understood.

Genome-scale metabolic models (GEMs) can be used to predict metabolic fluxes, enabling a greater understanding of cellular physiology, providing valuable information for metabolic engineering to develop better microbial factories. Several genome-scale metabolic models have been developed for *Rhodotorula toruloides*, but a comprehensive overview of simulations focused on central carbon metabolism with these models has not yet been presented. The objective of this thesis was to compare the predictions of central carbon metabolism focused on lipogenesis of four classical GEMs of *R. toruloides* (rhto-GEM, iRhtoC, Rt\_IFO0880, and Rt\_IFO0880\_LEBp2023) using FBA optimized for biomass maximization and NGAM minimization over five increasing specific growth rates. The simulations showed that over the growth rates, investigated fluxes increased linearly.

In regard to cofactor NADPH metabolism, the rhto-GEM and iRhtoC models predicted most of the NADPH to be produced through oxPPP, while the Rt\_IFO0880-based models predicted that most of the NADPH is produced by alcohol dehydrogenase, aldehyde dehydrogenase, or homoserine dehydrogenase, depending on the objective function and specific growth rate, indicating the need for a consensus genome-scale metabolic model of *R. toruloides*.

## Annotatsioon

Kliimamuutuste vastu võitlemiseks ja vähendamaks sõltuvust fossiilsetest ressurssidest on paljud riigid üleminekul bio-põhisele majandusele. See üleminek nõuab innovaatilisi protsesse keemiliste ainete, materjalide ja kütuste jätkusuutlikuks tootmiseks. Praegune biodiisli tootmine õliseemnetest ja jäätmeõlidest ei ole piisav, et rahuldada globaalset nõudlust [1], rõhutades vajadust biokütuste järele, mis pärinevad mittesöödavatest allikatest. Mikroobsed lipiidid (*microbial oils*), tuntud ka kui ühe-raku õlid (*single-cell oils*, SCOs), on paljutootav allikas kemikaalide tootmiseks, kuna selleks saab kasutatada väheväärtuslikke jäätmeid, tänu millele ei konkureeri antud tootmine toiduainetööstusega.

Mitte-traditsiooniline õlirikas pärm *Rhodotorula toruloides* on märkimisväärne mikroorganism bioproduktide tootmiseks, suutes akumulierida väga suurtes kogustes lipiide. Metaboolsed rajad, mis võimaldavad antud pärmis lipiidide tootmist, on üldiselt määratletud, kuid ainulaadsed metaboolsed omadused, mis võimaldavad kõrget lipiidide sünteesi, pole veel täielikult teada. Täpsemad teadmised vastavatest mehhanismidest võimaldaks tõhusamate rakuvabrikute bioinseneerimist. Ülegenoomsed metabolismi mudelid (*genome-scale metabolic models*, GEMs) võimaldavad raku ainevahetuse põhjalikku *in silico* uurimist. Pärmis *Rhodotorula toruloides* jaoks on sõltumatult välja töötatud mitu ülegenoomset mudelit, kuid puudub põhjalik ülevaade antud mudelitega ennustatud kesksest süsiniku metabolismist.

Käesoleva töö eesmärk oli võrrelda nelja mudeli (rhto-GEM, iRhtoC, Rt\_IFO0880 ja Rt\_IFO0880\_LEBp2023) ennustusi, keskendudes lipogeneesi prekursoreid, nimelt atsetüül-CoA-d ja kofaktor NADPH-d, tootvatele radadele. Metaboolsete voogude simuleerimiseks kasutati metaboolsete voogude analüüsi (*flux balance analysis*, FBA). Raku fenotüüpi prognoositi viie spetsiifilise kasvukiiruse korral, kasutades eksperimentaalseid *R. toruloides* pidevkultuuri (*continuous cultivation*) andmeid.

Simulatsioonidest selgus, et uuritud reaktsioonide vood (*fluxes*) kasvasid lineaarselt koos kasvukiiruse tõusuga. Tulemused näitasid selget erinevust mudelite ennustustes atsetüül-CoA tootmiseks: mudelid rhto-GEM ja Rt\_IFO0880 ennustasid fosfoketolaasi rada ning mudelid iRhtoC ja Rt\_IFO0880\_LEBp2023 ennustasid ATP tsitraadi lüaasi (*ATP-citrate lyase*) kasutamist. Ka kofaktor NADPH tootmise ennustused erinesid. Mudelid rhto-GEM ja iRhtoC ennustasid, et suurem osa NADPH-st toodetakse pentoosfosfaadi raja oksüdatiivses osas, samas kui Rt\_IFO0880-põhised mudelid ennustasid enamiku NADPH tootmist alkoholi dehüdrogenaasi, aldehüüd dehüdrogenaasi või homoseriin dehüdrogenaasi kaudu. Leitud tulemused näitavad vajadust konsensus *R. toruloides* ülegenoomse metabolismi mudeli järele.

## A. Appendix

### A.1. Non-exclusive licence for reproduction and publication of a graduation thesis

I, Maive Hanni,

1. grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis Comparison of genome-scale metabolic models for investigating lipogenesis metabolism in *Rhodotorula toruloides*, supervised by Alīna Reķēna,
  - 1.1 to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
  - 1.2 to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

29.05.2024