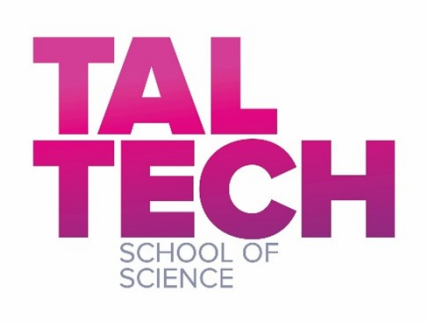

Comparison of genome-scale metabolic models for investigating lipogenesis metabolism in *Rhodotorula toruloides*

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: Alīna 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, 29.05.2024

The paper conforms to requirements in force.
Supervisor: Alīna Reķēna
Signed digitally, 29.05.2024

Contents

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

FAME fatty acid methyl ester

FAEE fatty acid ethyl ester

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

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

1. Theoretical background

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

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

The shift towards a bioeconomy needs novel processes for production of chemicals, materials, and liquid fuels from sustainable substrates, that offer improved life cycle assessments, and use less energy to produce. Advancements have highlighted the potential of chemicals derived from plant oils and animal fats as alternative feedstocks to the petrochemical industry. [3] Biodiesel is synthesized through the transesterification of triacylglycerols with short-chain alcohols (primarily methanol or ethanol) to yield monoalkyl esters, specifically fatty acid methyl esters (FAMEs) and fatty acid ethyl esters (FAEEs) [4].

The 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. [5] But the production of biodiesel from oilseeds and waste oils does not sustain the global demand [4] and the recent 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 [6].

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 [7]. Research has focused on the development of biodiesel production from SCO that are produced via fermentation using oleaginous microorganisms (microorganisms capable of accumulating lipids at more than 20% of the total cellular dry weight (DW)). Biodiesel production from SCO relies on the utilization of low-value waste streams or residues, thus presenting a sustainable alternative for biofuel production. Moreover, the production of SCOs does not require land or other resources that are typically used for food production and it is not influenced by season or climate. [4]

1.2. *Rhodotorula toruloides*

Rhodotorula toruloides (previously *Rhodosporidium toruloides*) is an oleaginous yeast which can accumulate lipids up to 76.1% of cell dry weight [8]. What is more, *R. toruloides* has good tolerance to inhibitory compounds that are naturally found in biomass hydrolysates [9]. *Rhodotorula toruloides* is an exceptional microbial lipid producer and has recently emerged as one of the most promising yeasts for bioproduction [10, 11].

R. toruloides occurs naturally in leaves, soil, sea water, etc. It has a broad substrate range, which has made this yeast a popular for producing biological oils from inedible substrates such as pentose sugars and crude glycerol. The majority of the lipids produced by *Rhodotorula toruloides* are triacylglycerol (TAG) contained long-chain fatty acids (C16:0 (palmitic acid), C16:1 (palmitoleic acid), C18:0 (stearic acid), C18:1(oleic acid), and C18:2 (linoleic acid)) and they are comparable to vegetable oils [8, 12]. *R. toruloides* lipid fraction contains also carotenoid pigments, omega-3 linolenic acid and heptadecenoic acid, which makes it a promising organism for production of pharma- and nutraceuticals [13].

Fatty acids mainly accumulate as TAGs, and they are produced via four 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 [14]. In fatty acid synthesis and elongation, Acetyl-CoA is the donor of C2-carbon. NADPH is required for reduction steps and it is mainly produced by malic enzyme (decarboxylating malate dehydrogenase, ME), and by glucose 6-phosphate dehydrogenase and phosphogluconate dehydrogenase in the pentose phosphate pathway (PPP). [15]

Metabolic pathways producing acetyl-CoA and a cofactor NADPH have been the main focus of 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). ACL has been demonstrated to be upregulated in *R. toruloides* during lipid accumulation [16] and it has been suggested to be the main source of acetyl-CoA for lipid synthesis in oleaginous species [17].

Proteomics analysis of *R. toruloides* has suggested that NADPH is mainly produced through the pentose phosphate pathway (when grown on xylose and glucose) but the role of malic enzyme is not clearly understood. The role of phosphoketolase in the generation of acetyl-CoA has not been acknowledged previously [16]. Lipid biosynthetic reactions downstream of acetyl-CoA synthesis do not differ between oleaginous and non-oleaginous yeast species [18].

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. [19] Thus, the carbon-to-nitrogen ratio (C/N) is a significant factor in initiating lipid accumulation [3]. 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. [19]

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 but did not accumulate lipids, suggesting that some other enzyme activities are also necessary for lipid accumulation. [19]

It has been found that another important enzyme in lipogenesis is malic enzyme, which generates NADPH, that is used by fatty acid synthetase. NADPH is also generated by glucose-6-phosphate dehydrogenase, phosphogluconate 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. [19]

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 be 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. [19]

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, these 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 conducive 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. [20]

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.

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. [21]

Fed-batch fermentation is a variation of batch fermentation. Microorganisms are initially grown under batch conditions, after which nutrients are incrementally added to the fermenter throughout the remaining fermentation duration. The addition of fresh nutrients typically results in significant biomass accumulation during the exponential growth phase. Therefore, fed-batch fermentation is particularly useful for bioprocesses aiming for high biomass density or high product yield when the desired product is positively correlated with microbial growth. [20]

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 maintainance of a continuous culture thanks to the addition and removal of fluids, adjusted to keep the culture in the logarithmic phase of growth. [21] Fresh medium is continuously added to the fermenter, while used medium, toxic metabolites and cells are simultaneously harvested. Unlike fed-batch

fermentation, 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. [20]

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 metabolite results from a combined regulation of many biological levels (transcription, translation, post-translational modifications and protein-protein interactions). [22] Hence, metabolic fluxes represent cellular phenotype under certain conditions and therefore analyzing the flux distribution of metabolites is very useful for studying cell metabolism [23]. It is difficult to measure intracellular fluxes experimentally, but it is possible to predict these fluxes thanks to metabolic models [22].

When the first full genome sequences were published in the 1990s, in principle it became possible to identify all the gene products involved in 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). [24] 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 [25, 26]. It is possible to integrate omics data and experimental metabolic fluxes to GEMs for generating holistic view of metabolism in different physiological states. This enables a greater understanding of cellular physiology, providing valuable information for metabolic engineering to develop better microbial factories.

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. [27]

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. [28]

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. [28]

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. [28]

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. [29] 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. [30]

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. [26] 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. [30]

Genome-scale metabolic models are constrained by three factors: (1) the stoichiometry of the network; (2) preset upper and lower limits for specific reactions; and (3) the assumption of a steady state. [25] Overview of a reconstruction of a GEMs is shown in the figure 1.1.

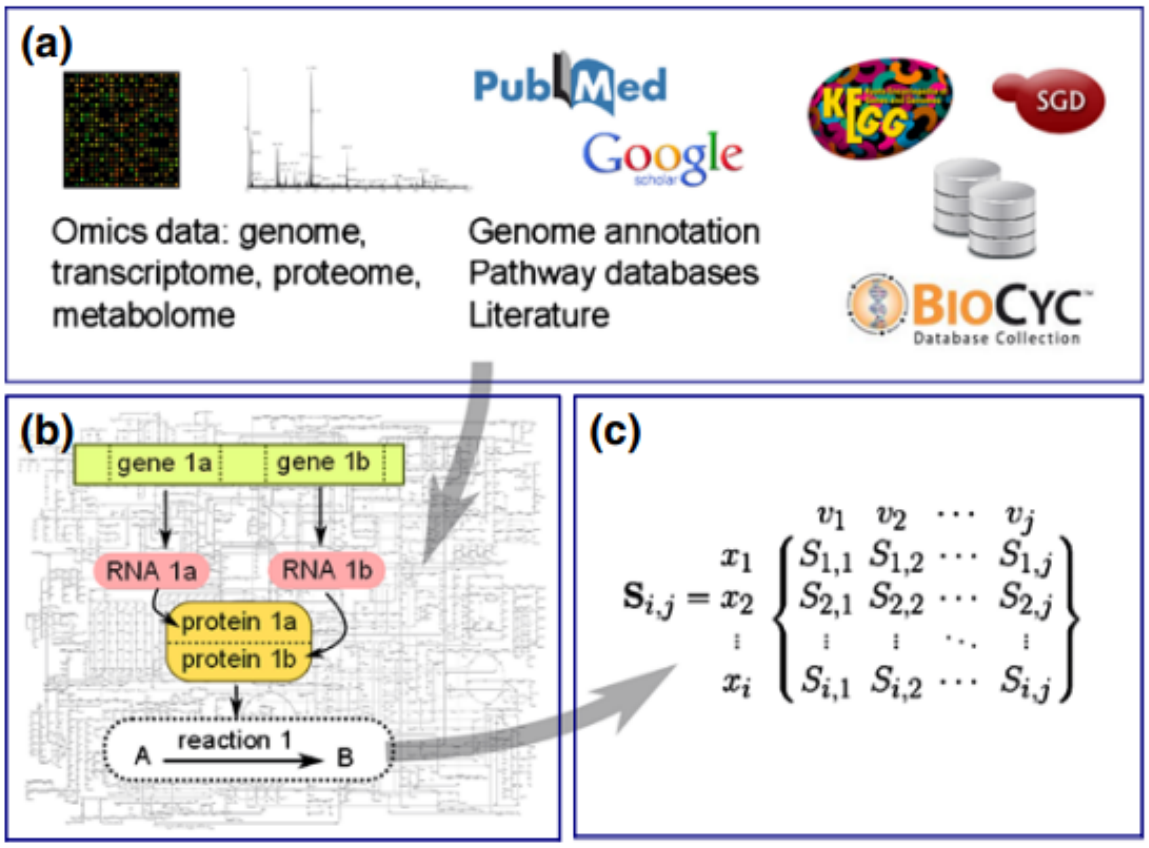

Figure 1.1: Reconstruction of a GEM. (a) The genome annotation is used to reconstruct the draft. (b) Gene-protein-reaction relationships are defined for the metabolic model. (c). A solution space is defined from the constraints applied to the model. Figure is from article [25].

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. FBA enables the computation of flow of metabolites through this metabolic network, thereby making it possible to predict an organism's growth rate or the production rate of a biotechnologically important metabolite. [31]

The initial phase in FBA involves the mathematical representation of metabolic reactions. The reconstructed genome-scale networks can be transformed into mathematical stoichiometric matrices $\mathbf{S} \in \mathbb{R}^{(m \times n)}$, where each row corresponds to one unique metabolite (for a system with m metabolites) and each column corresponds to an individual reaction (n reactions). [25] 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 $\mathbf{S}$ is sparse, as most biochemical reactions involve only a few different metabolites. The vector

$\mathbf{v}$ represents the fluxes of all the reactions in the network and it has a length of n. [31]

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 $\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$. [28] 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. [31] Thus, the solution space is further constrained by a set of upper and lower bounds on the fluxes ($a_i < v_i < b_i$).

The subsequent step in FBA is to define a biological objective that is relevant to the problem being studied. Even though constraints define a range of solutions, it is still possible to identify and analyze single points within the solution space. For instance, we may be interested in identifying which point corresponds to the maximum growth rate, the rate at which metabolic compounds are converted into biomass constituents (nucleic acids, proteins, and lipids), or to the maximum ATP production of an organism, given its particular set of constraints. FBA is one method for identifying such optimal points within a constrained space. [31]

Mathematically, the objective is represented by an objective function that indicates how much each reaction contributes to the phenotype. A biomass reaction that drains precursor metabolites from the system at their relative stoichiometries to simulate biomass production is selected by the objective function in order to predict growth rates. This reaction is scaled so that the flux through it is equal to the exponential growth rate μ of the organism. [31]

The problem of FBA is to find a vector $\mathbf{v}$ such that it satisfies constraints

$$\mathbf{S} \cdot \mathbf{v} = 0, \qquad a_i < v_i < b_i$$

and maximizes/minimizes the objective function. In order to solve this problem, linear programming methods are used. For example, the COBRA Toolbox [32] can be used for solving this kind of problems efficiently for large systems of equations. The COBRA Toolbox, freely available in Matlab and Python, uses models saved in the Systems Biology Markup Language (SBML) [33] format. The resulting flux distribution $\mathbf{v}$ maximizes or minimizes the given objective function (see figure 1.2). [31]

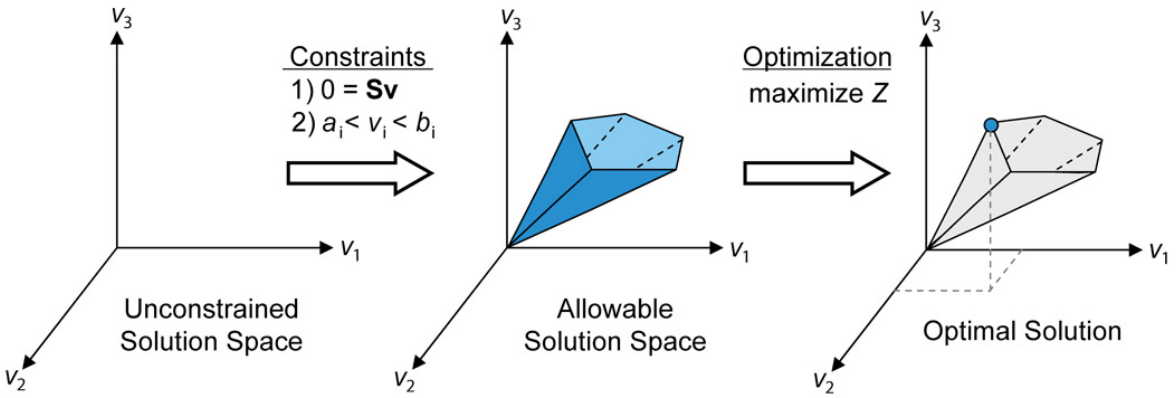

Figure 1.2: Conceptual basis of constraint-based modeling and FBA. The figure is from [31].

An alternative approach to FBA involves sampling of the solution space by considering all permissible flux distributions based on mass balance (stoichiometric) and flux capacity constraints. Uniform random sampling of the solution space is a method to understand the permissible metabolic flux space under any environmental condition. The flux distributions derived from this sampling can answer questions about the most probable flux value for any reactions and dependencies between two reactions under given constraints. [32]

The most commonly used objective function in FBA include maximization of the specific growth rate, ATP generation or a specific product formation [25]. Beyond metabolic costs, additional energetic requirements (in the form of ATP) exist for growth. These requirements account for growth-associated maintenance (GAM) and non-growth-associated maintenance (NGAM). [34] 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. [27] These reactions can also be used as objective functions.

FBA is commonly used to assess the biotechnological potential of microorganisms and identify genetic modifications that could enhance the cell performance. Its key applications include: (1) instructions for metabolic engineering; (2) biological interpretation and discovery through contextualizing high-throughput data; (3) creating a computational framework; (4) explaining evolutionary aspects; (5) describing multispecies communities. [25]

However, FBA has limitations. It lacks kinetic parameters, preventing the prediction of metabolite concentrations. Additionally, it only works at steady state and does not fully account for regulatory effects, potentially affecting accuracy. [31]

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

rhto-GEM

The first genome-scale model of *R. toruloides* metabolism named rhto-GEM was presented in 2019 by Tiukova et al. The model includes 4869 genes, 897 reactions, and 3334 metabolites. This model is based on the genome sequence of *R. toruloides* strain NP11 [16] (which is accessible from NCBI database [35]). For the reconstruction of the parts of metabolism 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. [18]

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 [36]. 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. [18]

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. [18]

iRhto1108

In the same year that Tiukova et al. introduced rhto-GEM [18], Dinh et al. presented another *R. torluoides* genome-scale metabolic model named iRhto1108. This model was built upon functional genomics data from [37] and prior knowledge [38]. The model is based on the metabolic network of the strain IFO0880 [37] (available from JGI database [39]). 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 [40], (ii) KBase fungal models [41]), and (iii) *R. toruloides* specific information extracted from the primary literature [37][42][43].

The essential metabolic functions and growth capability of the model were thoroughly validated

with experimental results, including gene essentiality [37] 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. [38] Considering that *R. toruloides* is a non-traditional microorganism, the iRhto1108 model has been a promising start for Rhto models, and it holds potential to assist in future research on *R. toruloides*.

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. [38]

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. This refined metabolic reconstruction consisted of 1106 genes, 1934 reactions, and 2010 metabolites (1246 of which were unique) across nine compartments. [44]

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. [44]

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 [18, 38]. The authors believe that the developed metabolic network Rt_IFO0880 is the most complete and accurate to date, and the presented model will be a valuable resource for studying and engineering *R. toruloides*

for lignocellulosic biomass conversion. [44]

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 rhto-GEM_BioEng, a version of the rtho-GEM with integrated carotenoids into the biomass composition and an alternative xylose assimilation pathway [45]. The model Rt_IFO0880 was selected for further enhancement 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. [46]

The author made several modifications to the model, 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+ transport reaction. The updated model, named Rt_IFO0880_LEBp2023, was then validated against experimental data. [46]

Unlike its predecessor, the GEM Rt_IFO0880_LEBp2023 accounted for the use of ACL in all maximum theoretical yield of phytoenes scenarios. This enzyme is highly abundant in *R. toruloides* [16], so its prediction by GEMs would be expected. However, from other models only the iRhtoC predicted the use of ACL. It seems that the addition of cytosolic malate dehydrogenase allowed the Rt_IFO0880_LEBp2023 to predict the use of ACL. This enzyme facilitates the conversion of oxaloacetate produced by ACL into malate, which is then transported into the mitochondria in an antiporter with mitochondrial citrate, which is taken to the cytosol to be a substrate of ACL. [46] This pathway, recently described as an alternative to the classic tricarboxylic acid (TCA) pathway, has been called the "non-canonical TCA cycle" [47]. The inability of other models to predict the non-canonical TCA cycle can be attributed to the absence of cMDH in Rt_IFO0880 and the absence of citrate-malate antiporter between the cytosol and the mitochondria in rhto-GEM and rhto-GEM_BioEng. [46]

The updated model, Rt_IFO0880_LEBp2023, demonstrated a better fit to the experimental data than the other models and was able to predict the use of the ACL enzyme, which is present in *R. toruloides* according to omics studies [16]. The model also suggested that *R. toruloides* possibly uses a non-canonical TCA cycle to avoid the production of CO_2 in the generation of mitochondrial NADH, a characteristic previously observed for this yeast. Despite only few updates, Rt_IFO0880_LEBp2023 seems to have better predictions compared to the other GEMs of *R. toruloides*. [46]

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 objective functions and constraints. The intracellular fluxes of pentose phosphate pathway and citric acid cycle enzymes are compared between the 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 [18, 38, 44] and the fourth model Rt_IFO0880_LEBp2023 [46], made by a group member, was obtained directly from him.

All scripts are available in Github repository (`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 our group and the results have not yet been published. The cultivation process is described in more detail in [48], with the exception that the continuous cultivation regime was used 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 $(NH_4)_2SO_4$, 3 g/L KH_2PO_4, 0.5 g/L $MgSO_4$ heptahydrate [49], supplemented with vitamins and minerals according to Verduyn [50]. Bioreactors were equipped with gas analyser (BlueSens gas sensor GmbH, Herten, Germany) used for measuring the composition of CO_2 and O_2 in the gas outflow. Optical density was monitored using a Bug Lab BE3000 Biomass Monitor (Bug Lab, Concord, CA, United States). 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: Continuous cultivation results from lab experiments

Biomass growth rate μ 1/h	**Glucose uptake rate** r_{glu} mmol/gDW/h	**Glycerol secretion rate** r_{gly} mmol/gDW/h
0.049	0.476	
0.100	1.114	
0.151	1.648	< 0.02
0.203	2.305	
0.25	-	
0.301	3.1	

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, Rt_IFO0880&Rt_IFO0880_LEBp202, the lipid content, which is the most important parameter for present study, is .., 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. [51] 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) [52] in Python (version 3.11.4) with all four models. Throughout the process, metabolic flux patterns were predicted using flux balance analysis [31] 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 absolute 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. Firstly, 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 <=>) 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 is simulations is obtained, defined mineral medium was used. CO_2 and O_2 exchange rates were left unconstrained.

On carbon limitation the objective function was set to biomass maximisation (reaction ID: r_4041 (in model rhto-GEM), Biomass_Rt_Clim (in iRhtoC), BIOMASS_RT (in other two)).

Minimisation of non-growth associated maintenance (NGAM) reaction was set as an objective function (r_4046, ATPM_c, ATPM, respectively; ATP[c] + H_2O[c] => ADP[c] + H^+[c] + 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, NADH and ATP. 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 5% 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 *R. toruloides* from non-oleaginous yeast are not fully understood. Biosynthesis of the main lipid precursors acetyl-CoA and NADPH takes place in the central carbon metabolism. A better understanding of which metabolic pathways are used in production of these precursors and thus contribute to lipid accumulation, would aid in designing better metabolic engineering strategies for increasing lipid production.

Genome-scale metabolic models contain all known biochemical reactions of the specific organism and allow the calculation of metabolic fluxes, which represent the activity of metabolic pathways under specified conditions. This makes GEMs important tool for studying metabolism, but it is important that the predictive power of the GEM is 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

To obtain the flux distribution through the metabolic network, simulations using firstly biomass maximization as an objective function were carried out. Calculations were made under carbon limitation constraining glucose uptake. To assess the flux patterns over different biomass growth rates, the solutions were constrained over five experimental glucose uptake rate values - $0.476, 1.114, 1.648, 2.305$ and 3.1 $(mmol/gDW/h)$.

All simulated fluxes are available on a Github repository (`https://github.com/maivehanni/BSc_thesis`). In further analysis we focused on selected exchange and intracellular fluxes, which are visualized in graphs and shown below.

All figures show the fluxes of metabolites over growth rates from $0.05 - 0.25$ $1/h$. Exchange fluxes plots show the exchange (uptake or secretion) rate of glucose, oxygen, ammonium, sulphate, phosphate and carbon dioxide. The experimental exchange rates of glucose, oxygen and carbon dioxide are also visualized on the same plot and visualized with dashed lines. Intracellular fluxes graphs show the fluxes of NGAM, glucose uptake, glucose 6-phosphate dehydrogenase (oxPPP), transketolase 1, transaldolase, transketolase 2, fructose-biphosphate aldolase, pyruvate decarboxylase, pyruvate dehydrogenase, phosphoketolase and ATP-citrate lyase.

NADPH producing and consuming fluxes were also investigated on growth rates from $0.05-0.25$

$1/h$, but only the ones that significantly differed on different growth rates, are shown below. In these pie charts, 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 infront of the name is the percent that the given enzyme makes up of the total producing or consuming NADPH flux.

rhto-GEM

Compared to the experimental physiological rate data, rhto-GEM predicts higher exchange fluxes for a measured growth rate (figure 4.1). This result was expected as GEMs are known to overestimate fluxes per growth rate.

All the investigated intracellular fluxes were predicted by the model to increase linearly over the growth rates from 0.03 to 0.25 $1/h$. Flux through oxPPP was predicted to increase from 0.18 to 1.45 $mmol/gDCW/h$, 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), 28% to transketolase 2 ($0.04 - 0.3$) and 40% to phosphoketolase (fluxes from $0.05 - 0.4$). 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 $mmol/gDCW/h$, representing 82% of the carbon from D-fructose 6-phosphate. Pyruvate decarboxylase represents 15% and pyruvate dehydrogenase 59% of the carbon from cytososlic pyruvate branching point having fluxes from 0.06 to 0.5 and $0.46 - 2.25$ $mmol/gDCW/h$, respectively. The model did not predict any flux for ACL, instead it predicts that acetyl-CoA is produced through phosphoketolase.

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).

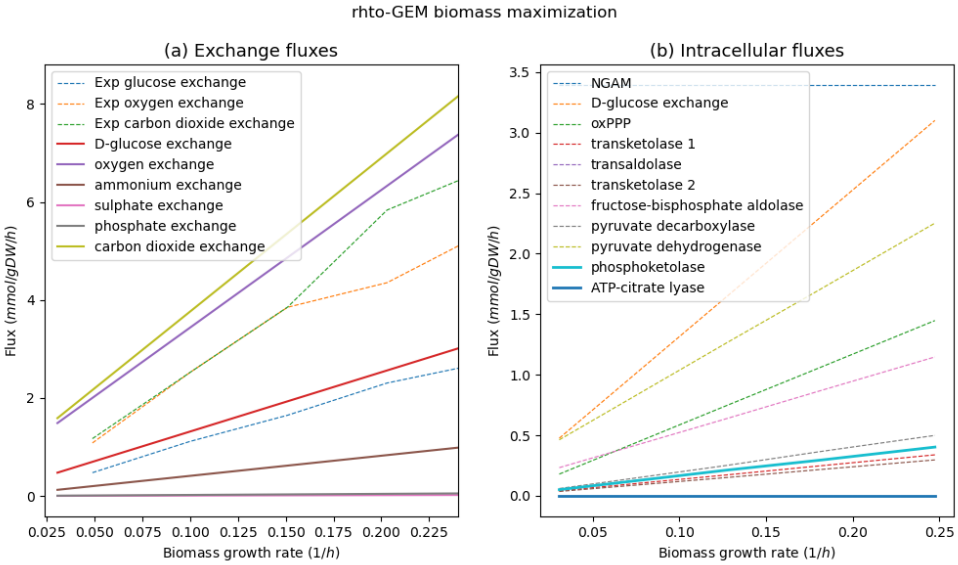

Figure 4.1: Exchange and intracellular fluxes in *R. toruloides* with model rhto-GEM optimized for biomass maximization and constrained over five glucose uptake rates.

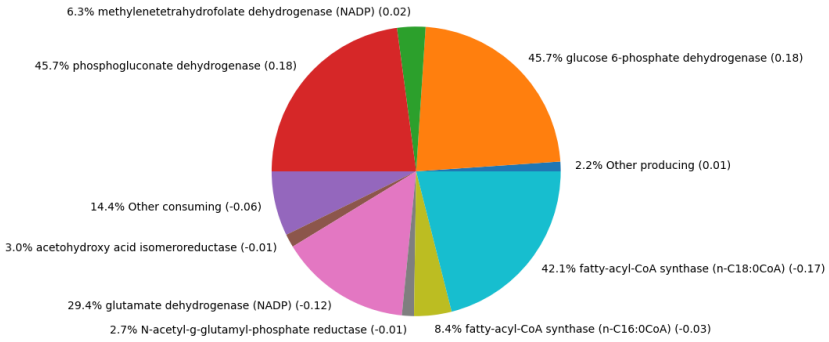

Figure 4.2: NADPH producing and consuming fluxes in *R. toruloides* with model rhto-GEM optimized for biomass maximization. Glucose uptake was constrained to the lowest rate.

iRhtoC

This model predicts very similar exchange fluxes as the previous model, predicting higher rates than experimentally measured. But intracellular fluxes differ from model rhto-GEM. iRhtoC predicts that over the five growth rates flux through oxPPP increase from 0.24 to 1.69 $mmol/gDCW/h$ (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 $0.07-0.5$ representing 53% and 47% from D-xylulose 5-phosphate branching point, respectively). Fluxes of transaldolase and pyruvate dehydrogenase are predicted to be zero. Fructose-bisphosphate aldolase (representing 93% from D-fructose 6-phosphate) and pyruvate dehydrogenase (80% from cytososlic pyruvate branching point)

fluxes are from 0.24 to 1.49 and from 0.53 to 3.29 $mmol/gDCW/h$, respectively. ATP-citrate lyase has a flux from 0.18 to 1.29 representing 40% of carbon from mithocondrial citrate (figure 4.3).

This model predicts that almost 90% of NADPH is produced by glucose 6-phosphate dehydrogenase (oxPPP) and phosphogluconate dehydrogenase on all rates, but around 6% is produced by by malic enzyme on growth rates 0.03 and 0.23 and on other rates the model predicts isocitrate dehydrogenase instead (figures 4.4 and 4.5).

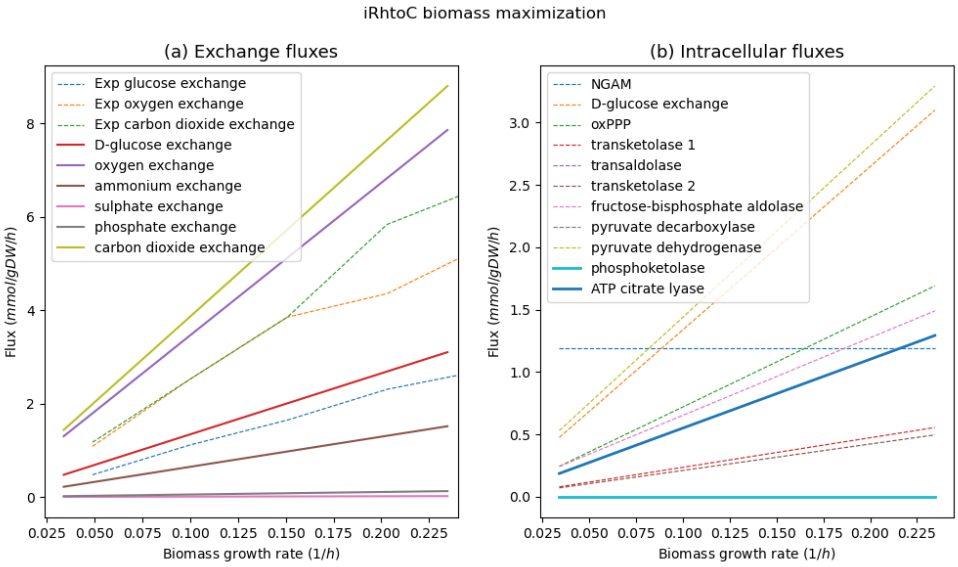

Figure 4.3: Exchange and intracellular fluxes in *R. toruloides* with model iRhtoC optimized for biomass maximization and constrained over five glucose uptake rates.

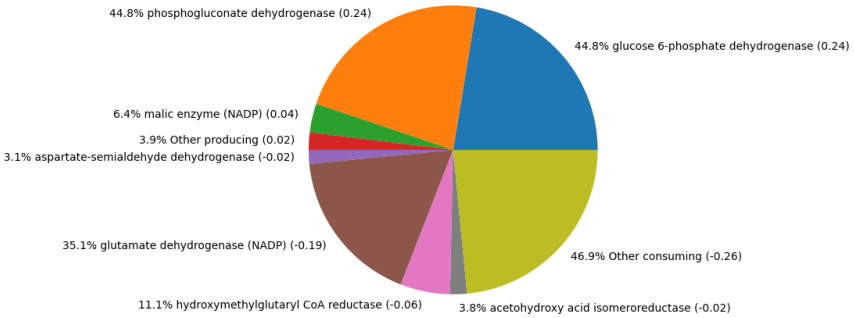

Figure 4.4: NADPH producing and consuming fluxes in *R. toruloides* with model iRhtoC. The model was optimized for biomass maximization and glucose uptake was constrained to the lowest rate. The fluxes are same when glucose uptake is constrained to highest rate.

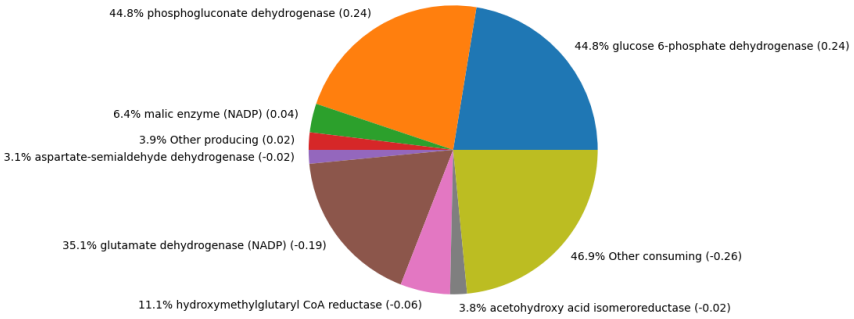

Figure 4.5: NADPH producing and consuming fluxes in *R. toruloides* with model iRhtoC. The model was optimized for biomass maximization. Results are the same when glucose uptake is constrained to $1.114, 1.648$ or 2.305 $(mmol/gDW/h)$.

Rt_IFO0880

Rt_IFO0880 also predicts higher exchange fluxes than experimentally measured. 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 over the rates. Pyruvate decarboxylase represents 7% and pyruvate dehydrogenase 58% of carbon from cytososlic pyruvate branching point (figure 4.6). Similarly to rhto-GEM, this model also predicts the use of phosphoketolase and no use of ACL.

In NADPH production and consumption on different rates there are not significant differences, as around 90% of NADPH is produced by alcohol dehydrogenase on all rates (figure 4.7). This is very different from other models.

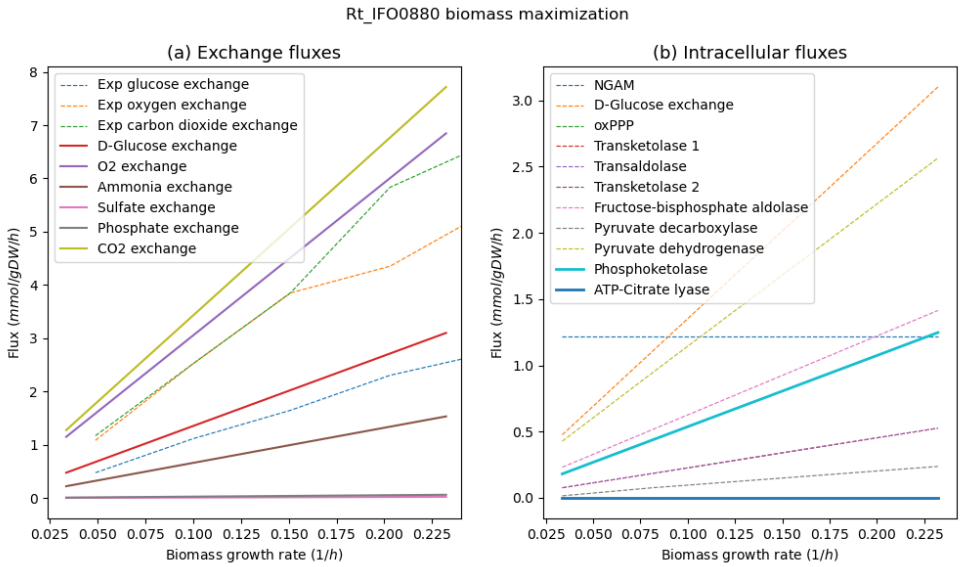

Figure 4.6: Exchange and intracellular fluxes in *R. toruloides* with model Rt_IFO0880 optimized for biomass maximization and constrained over five glucose uptake rates.

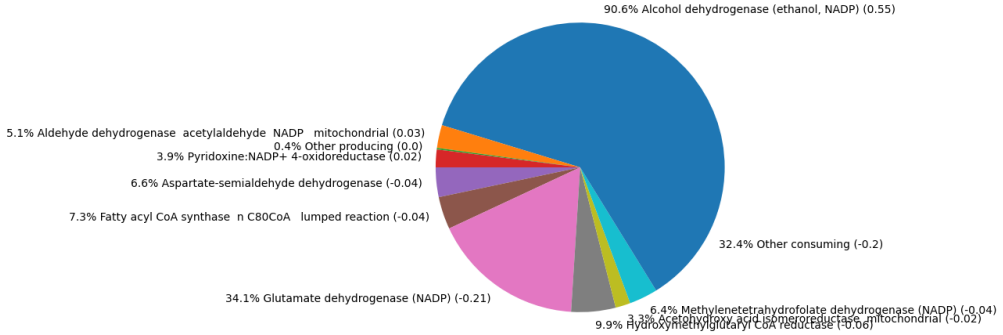

Figure 4.7: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880. The model was optimized for biomass maximization and glucose uptake was constrained.

Rt_IFO0880_LEBp2023

Exchange fluxes align with other models. 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 cytososlic pyruvate. 33% of carbon from mithocondrial citrate is represented by ACL. (4.8) 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.)

On lowest growth rate, NADPH is predicted to be produced mainly ($\sim 85\%$) by aldehyde dehydrogenase 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.

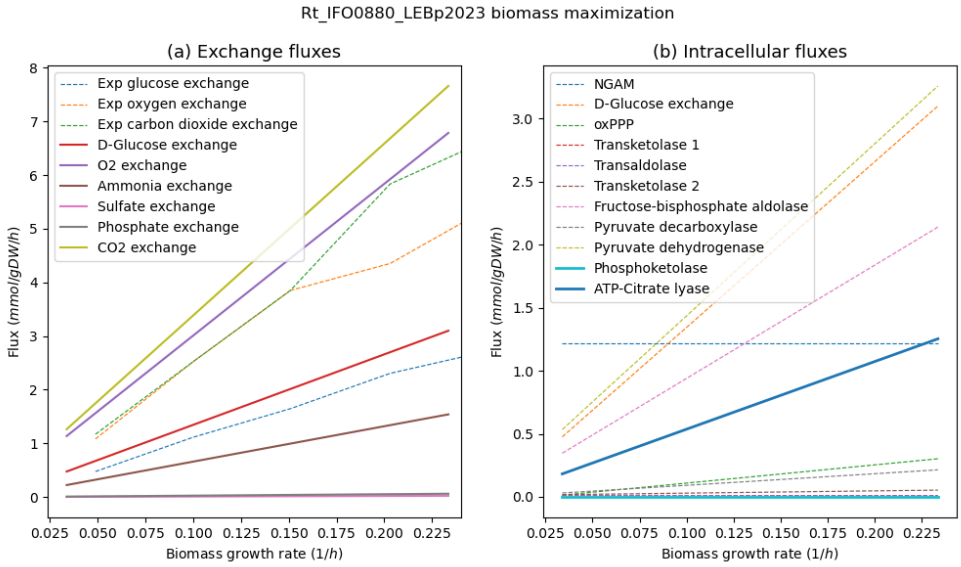

Figure 4.8: Exchange and intracellular fluxes in *R. toruloides* with model Rt_IFO0880_LEBp2023 optimized for biomass maximization and constrained over five glucose uptake rates.

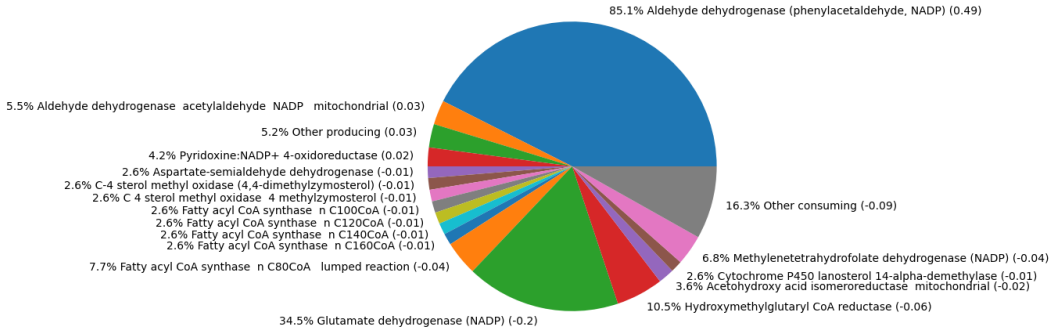

Figure 4.9: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880_LEBp2023. The model was optimized for biomass maximization. Glucose uptake was constrained on the lowest rate.

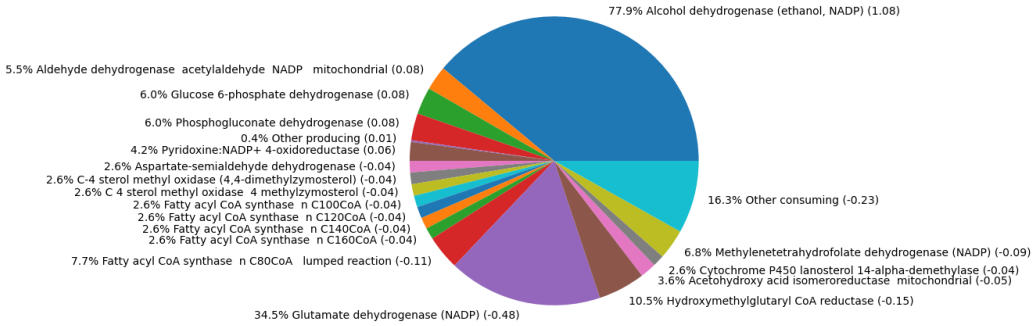

Figure 4.10: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880_LEBp2023. The model was optimized for biomass maximization and glucose uptake constrained to 1.114. The results are very similar when glucose uptake is constrained to $1.114, 1.648, 2.305$ or 3.1.

Simulations with biomass maximisation as objective function under carbon limitation, showed that the models prefer different pathways for production of acetyl-CoA. The models rhto-GEM and Rt_IFO0880 produced acetyl-CoA through phosphoketolase, whereas models iRhtoC and the modified model of Rt_IFO0880 (Rt_IFO0880_LEBp2023) produced acetyl-CoA using ACL. NADPH sources also differ. On all growth rates, models rhto-GEM and iRhtoC predict that most of the NADPH is produced by the oxidative pentose phosphate pathway, but $\sim 6\%$ by methylenetetrahydrofolate 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.

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

To see whether different objctive 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. Optimizing for NGAM minimisation it is assumed that cell strive for satisfying physiological parameters with least energy expenditure as NGAM minimisation decreases the ATP demand.

Same experimentally measured exchange and secretion rates were used throughout the simulations. This objective function also needed constraints on biomass growth rate because

otherwise the simulation chooses zero as its flux. Growth rate was constrained to the values that each model predicted in the simulations optimized for biomass maximisation, because experimental growth rates together with experimental glucose uptakes were infeasible for the models.

rhto-GEM

Model rhto-GEM optimized for NGAM minimisation predicts similar fluxes as with previous objective function.

iRhtoC

iRhtoC optimized for NGAM minimisation predicts same exchange and intracellular fluxes as with previous objective function. As previously, most of the NADPH is produced by oxPPP, but on the lowest growth rate $\sim 6\%$ is produced by isocitrate dehydrogenase and on all other rates by malic enzyme (figures 4.11 and 4.12). NGAM minimisation and biomass maximisation as an objective function predict differences in the use of malic enzyme vs isocitrate dehydrogenase on different growth rates.

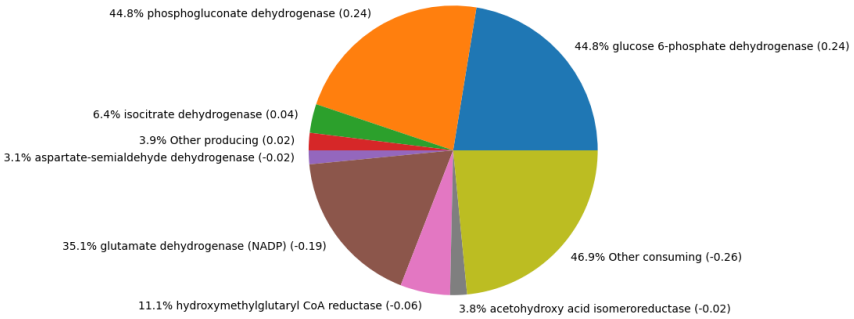

Figure 4.11: NADPH producing and consuming fluxes in *R. toruloides* with model iRhtoC. The model was optimized for NGAM minimization. Glucose uptake and growth rate were constrained on the lowest rate.

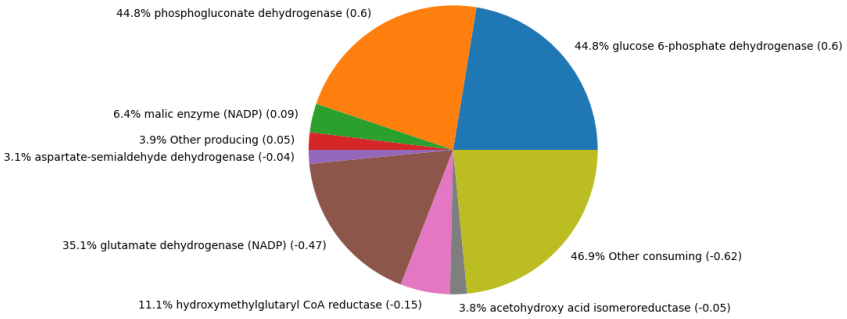

Figure 4.12: NADPH producing and consuming fluxes in *R. toruloides* with model iRhtoC. The model was optimized for NGAM minimization, growth rate and glucose uptake were constrained to 0.08 and 1.114, respectively. The results are the same when glucose uptake is constrained to $1.114, 1.648, 2.305$ or 3.1.

Rt_IFO0880

Model Rt_IFO0880 optimized for NGAM minimisation predicts same exchage and intracellular fluxes as before. But production of NADPH differs - around 90% is produced by homoserine dehydrogenase on rates 0.08 and 0.17 $(1/h)$ and on other rates by alcohol dehydrogenase (figures 4.13 and 4.14).

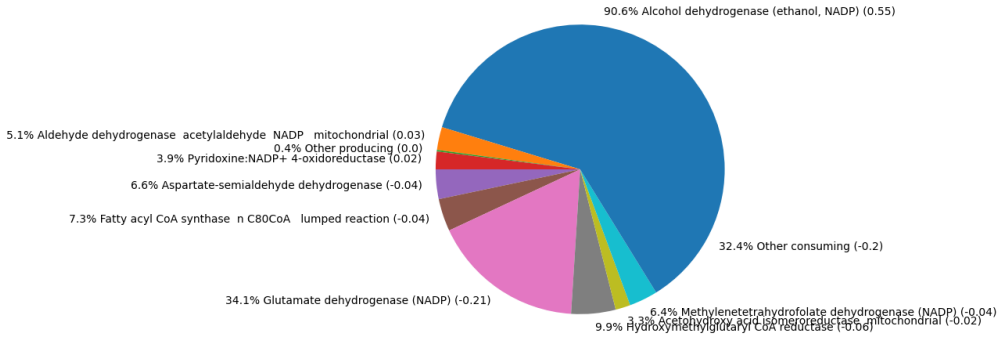

Figure 4.13: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880. The model was optimized for NGAM minimization. Glucose uptake and growth rate were constrained on the lowest rate. The results are the same when glucose uptake is constrained to 1.648 or 3.1.

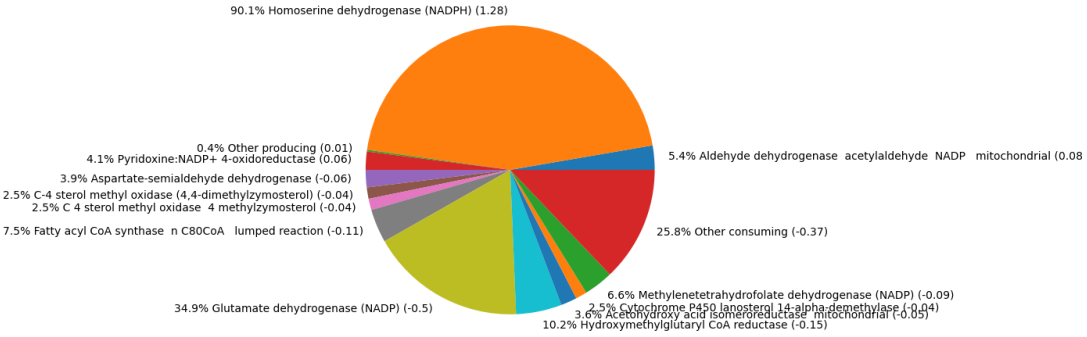

Figure 4.14: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880. The model was optimized for NGAM minimization, growth rate and glucose uptake were constrained to 0.08 and 1.114, respectively. The results are similar when glucose uptake is constrained to 2.305.

Rt_IFO0880_LEBp2023

Rt_IFO0880_LEBp2023 optimized for NGAM minimisation predicts very similar exchange and intracellular fluxes as before. The production of NADPH differs - on rates $0.08, 0.12$ and 0.17 $(1/h)$ use of homoserine dehydrogenase is predicted instead of alcohol dehydrogenase (4.15). On lowest rate aldehyde dehdrogenase and on highest rate alcohol dehydrogenase is predicted, which is the same as previously with biomass maximisation as an objective function.

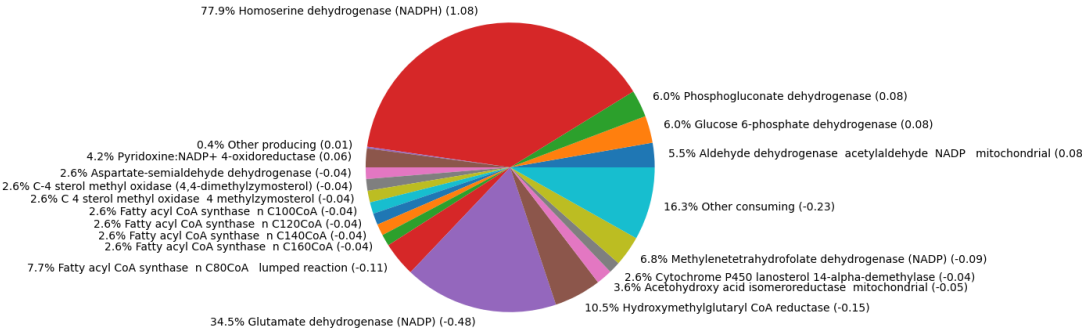

Figure 4.15: NADPH producing and consuming fluxes in *R. toruloides* with model Rt_IFO0880_LEBp2023. The model was optimized for NGAM minimization, growth rate and glucose uptake were constrained to 0.08 and 1.114, respectively. The results are very similar when glucose uptake is constrained to 1.648 or 2.305.

For all simulations, the exchange fluxes are the same as with biomass maximisation as objective function. Also, fluxes of the intracellular metabolites visualized in figures did not change. Meaning that the models that predicted phosphoketolase or ACL in previous simulations,

predicted it again. For models besides rhto-GEM the production of NADPH differs with NGAM as an objective function. With model iRhtoC, most of the NADPH is still produced by oxPPP, but the remaining 6% on all growth rates besides the lowest, is produced by isocitrate dehydrogenase when optimized for NGAM minimisation instead of malic enzyme. Model Rt_IFO0880 predicts that on rates 0.08 and 0.17 $(1/h)$ NADPH is primarily produced by homoserine dehydrogenase instead of homoserine dehydrogenase. Rt_IFO0880_LEBp2023 also predicts the use of homoserine dehydrogenase instead of alcohol dehydrogenase on rates $0.08, 0.12$ and 0.17 $(1/h)$.

5. Conclusion

In this work four genome-scale metabolic models of *Rhodotorula torluoides* were compared and two objective functions, maximization of biomass and NGAM minimization, were compared. The fluxes through phosphoketolase and ACL pathways, that are found to be important for lipid production in all four models, were explored. This study gave more insight about how these model predict central carbon metabolism of *R. torluoides* than [46] that was based on phytoenes simulations.

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

Acknowledgements

References

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

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

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

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

[5] 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).

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

[7] Bonturi, N. et al. "Microbial oil production in sugarcane bagasse hemicellulosic hydrolysate without nutrient supplementation by a Rhodosporidium toruloides adapted strain". In: *Process Biochemistry* 57 (2017), pp. 16–25. DOI: 10.1016/j.procbio.2017.03.007.

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

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

[10] 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.

[11] Park, Y.-K., Nicaud, J.-M., and Ledesma-Amaro, R. "The Engineering Potential of Rhodosporidium 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.

[12] 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.

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

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] *Microbial Growth*. [Online; accessed 2024-05-14]. 2024.

[22] 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.

[23] 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.

[24] Palsson, B. "Metabolic systems biology". In: *FEBS Letters* 583.24 (2009), pp. 3900–3904. DOI: 10.1016/j.febslet.2009.09.031.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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/ (visited on 05/15/2024).

[36] 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.

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

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

[39] Joint Genome Institute, M. F. P. *Rhodosporidium toruloides IFO0880 v4.0*. Accessed: 2024-05-15. 2018. URL: https://mycocosm.jgi.doe.gov/Rhoto_IFO0880_4/Rhoto_IFO0880_4.home.html (visited on 05/15/2024).

[40] 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.

[41] 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.

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

[43] 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.

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

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

[46] 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.

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

[48] 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.

[49] 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.

[50] 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.

[51] Shen, H. et al. "Kinetics of continuous cultivation of the oleaginous yeast Rhodosporidium toruloides". In: *Journal of Biotechnology* 168.1 (2013), pp. 85–89. DOI: 10.1016/j.jbiotec.2013.08.010.

[52] 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.

Abstract

Annotatsioon

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 Alina Rekena,
   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