UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Technology

Alina Rekena

Enzyme-constrained genome-scale metabolic model of *Rhodotorula toruloides*

Master's Thesis (30 ECTS)

Curriculum Bioengineering

Supervisors:
Professor, PhD Petri-Jaan Lahtvee
Senior Scientist, PhD Nemailla Bonturi
Research Scientist, PhD Isma Belouah

Tartu 2021

Enzyme-constrained genome-scale metabolic model of *Rhodotorula toruloides*

Abstract:

Rhodotorula toruloides is a non-conventional, oleaginous yeast able to naturally accumulate high amounts of microbial lipids when grown on various carbon substrates, including from a lignocellulosic origin. Its unique metabolic characteristics, which make lipid synthesis possible, are not fully understood. With genome-scale models (GEMs) it is possible to systematically study cellular metabolism using metabolic flux predictions *in silico*. Enzyme-constrained genome-scale modelling approach has been demonstrated to improve cell phenotype predictions in model organisms, including yeasts. In this work, enzyme-constrained genome-scale metabolic model of *R. toruloides* was developed, incorporating cell physiology and absolute proteomics data on three different carbon substrates (xylose, glucose, acetic acid) under exponential growth and lipid accumulation phases. The generated model could predict experimental rates measured in all conditions, except for the gases on glucose. Further, predicted intracellular flux patterns demonstrated the differences in *R. toruloides* metabolism under different carbon substrates and the importance of cofactor balance (NADPH) during the lipid accumulation. These results and the developed genome-scale model can be further used for the design of efficient microbial cell factories and various metabolic studies.

Keywords:

Rhodotorula toruloides; yeast; non-conventional yeast; oleaginous yeast; metabolism; genome-scale model; GEM; FBA; microbial lipids

CERCS:

T490 Biotechnology

Ensümaatiliste piirangutega ülegenoomne mudel pärmile *Rhodotorula toruloides*

Lühikokkuvõte:

Rhodotorula toruloides on mittetraditsiooniline õlirikas pärm, mis on võimeline akumuleerima suures koguses mikroobseid lipiide, kasvades erinevatel kohalikku päritolu toormetel nagu näiteks lignotselluloosne biomass. Selle mikroorganismi ainulaadsed metaboolsed omadused, mis võimaldavad lipiidide sünteesi, pole tänaseni veel täielikult teada. Ülegenoomsete mudelite (GEM) abil on võimalik süstemaatiliselt uurida raku ainevahetust, kasutades metaboolsete voogude analüüsi *in silico*. Varasemalt on näidatud, et ensüümaatiliste piirangutega kasutamine ülegenoomsetes mudelites parandab märkimisväärselt raku fenotüübi prognoose mudelorganismides, sealhulgas pärmides. Käesolevas töös töötati välja *R. toruloides'*e ensüümaatilite piirangutega ülegenoomne metaboolne mudel, mis sisaldab raku füsioloogiat ja kvantitatiivse proteoomika andmeid kolme erineva süsiniku allika (ksüloosi, glükoosi, äädikhappe) kohta nii eksponentsiaalses kasvu kui ka lipiidide akumulatsiooni faasis. Loodud mudel suutis prognoosida eksperimentaalselt teostatud katsete tulemusi kõikides mõõdetud tingimustes, välja arvatud glükoosil kasvades eralduvaid süsihappegaasi koguseid. Lisaks näitasid ennustatud rakusisesed voogude mustrid erinevusi *R. toruloides* ainevahetuses erinevate süsinikuallikate korral ja kofaktor NADPH tasakaalu hoidmise olulisust lipiidide akumulatsioonil. Neid tulemusi ja väljatöötatud ülegenoomset mudelit saab edaspidi täiendavalt kasutada tõhusamate rakuvabrikute kavandamiseks kui ka erinevate metaboolsete uuringute läbiviimiseks.

Võtmesõnad:

Rhodotorula toruloides; pärm; mitte-traditiooniline pärm; õli-rikas pärm; metabolism; ülegenoomne mudeldamine; voogude analüüs; mikroobsed rasvad

CERCS:

T490 Biotehnoloogia

TABLE OF CONTENTS

ABBREVIATIONS

GEM - genome-scale model

NADPH - nicotinamide adenine dinucleotide phosphate (reduced form)

SCO - single cell oil

AMP - adenosine monophosphate

TCA - Tricarboxylic acid cycle (citric acid cycle)

ACL - ATP citrate lyase

CoA - coenzyme A

HMF - hydroxymethylfurfural

TAG - triacylglycerol

XR - xylose reductase

XDH - xylitol dehydrogenase

FBA - Flux Balance Analysis

k_{cat} - enzymatic turnover number

EC - Number Enzyme Commission Number

BRENDA - Braunschweig Enzyme Database

exp - exponential growth

Nlim - nitrogen limitation

DCW - dry cellular weight

μ - specific growth rate (h^{-1})

r - specific rate (g/(gDCW*h)

Y - yield (g/g)

PPP - pentose phosphate pathway

NADH - nicotinamide adenine dinucleotide (reduced form)

BLAST - Basic Local Alignment Search Tool

XK - xylulokinase

LXR - L-xylulose reductase

DAD-2 - D-arabinitol 2-dehydrogenase

RK - ribulokinase

PK - phosphoketolase

ACS - acetyl-CoA synthetase

ACC - acetyl-CoA carboxylase

ME - malic enzyme

INTRODUCTION

Transition towards biobased, circular economy to reduce the industrial dependence on fossil-based resources requires new technologies. One of the options is to convert available biomass feedstocks into valuable chemicals using microbes as biocatalysts.

Rhodotorula toruloides is a non-pathogenic, aerobic, non-conventional yeast that has recently emerged as one of the most promising yeasts for a sustainable production of chemicals and fuels due to its natural ability to synthesize high amounts of lipids and carotenoids (Park, Nicaud and Ledesma-Amaro, 2018). However, its unique metabolic properties are not yet fully understood. The emergence of systems biology methods, including genome-scale modelling and multi-omics analysis, has enabled a holistic investigation of the metabolism of non-conventional microorganisms, including *R. toruloides* and, in particularly, its oleaginous phenotype.

Genome-scale models (GEMs) have enabled *in silico* prediction of cellular behaviour and the biological discovery through contextualising high-throughput data (Kerkhoven, Lahtvee and Nielsen, 2015). GEM improved with enzymatic constraints of the model yeast *Saccharomyces cerevisiae* demonstrated the ability to predict complex biological phenomenon, such as Crabtree effect (Sánchez *et al.*, 2017).

The first GEM of *R. toruloides*, rhto-GEM, (Tiukova, Prigent, *et al.*, 2019) was able to provide valid growth predictions on various carbon substrates and predict genetic engineering targets, some of which have previously been engineered to successfully increase specialty chemicals production. In this work, the development of an enzyme-constrained GEM of *R. toruloides* ecRhtoGEM, is described. The results of model simulation on three different carbon substrates – xylose, glucose, and acetic acid – showed the differences in metabolic characteristics, in particularly, related to lipid biosynthetic pathways.

1 LITERATURE REVIEW

1.1 The role of yeast cell factories in biobased economy

To ensure a sustainable development of industrialisation in the future, alternatives to the use of fossil fuel-based natural resources are encouraged. Heavy use of fossil fuel resources is associated with the greenhouse gas emissions, political disruptions and military conflicts and price volatility (Panwar, Kaushik and Kothari, 2011; Owusu and Asumadu-Sarkodie, 2016). Among most likely alternatives, biomass is the largest solid renewable natural resource, being less polluting in terms of particulates, like sulphur, lead, and greenhouse gasses that cause global warming (Goldemberg, 2007; Panwar, Kaushik and Kothari, 2011). Biomass feedstock (renewable carbon) includes grasses, wood and forest residues, aquatic biomass (algae and seaweeds), agricultural crops and residues (such as wheat, beet, and straw), as well as animal, urban and industrial residues, including solid waste and liquid effluents that are gradually increasing along with a growing world population (Fernando *et al.*, 2006; Goldemberg, 2007; Takkellapati, Li and Gonzalez, 2018). Biorefinery - a system similar to petroleum refinery – is a process of refining biomass in a commercial context for the production of fuels, chemicals, polymers, materials, food, feed, and value-added products with minimum waste generation (Holladay *et al.*, 2007; Koutinas *et al.*, 2014). Integrated biorefinery concept includes processes, plants and facilities to produce biofuels, energy and chemicals from biomass (Lee, Lee and Lee, 2021). Currently, several biorefinery processes exist allowing the use of a single feedstock to produce two or three products, including energy-driven products (Lee, Lee and Lee, 2021). Successful transition to biobased economy involves the development of new technologies and unit operations, production of new building blocks, conversion of these building blocks into marketable products and significant investments to scale-up new processes (Koutinas *et al.*, 2014).

Microorganisms as biocatalysts have been employed for the production of next generation biofuels and biobased chemicals for the past decades. Biobased production routes have been described to produce: (i) chemicals with 2-6 carbons (C2-C6) that can be further used by the chemical industry or other industrial sectors, (ii) single cell oils (SCOs) with 14-24 carbons, and (iii) various biopolymers (Koutinas *et al.*, 2014; Lee, Lee and Lee, 2021). Typically, for production of the C2-C4 chemicals (lactic acid, propionic acid, 1,3-propanediol (PDO)) and biopolymers (polyhydroxyalkanoate (PHA), cellulose) bacteria are more suitable production hosts. Nevertheless, yeasts are as well used for the production of several chemicals from the C2-C6 category, such as C2 bioethanol (produced approximately 100 billion litres per year

globally, (Bajpai, 2021)), C2 polyglycolic acid (Jem and Tan, 2020), C3 polylactic acid (PLA) (VTT Technical Research Centre of Finland, 2021), C4 L-malic and fumaric acids (Koutinas *et al.*, 2014). Also, for the production of C5 polyol xylitol, yeasts are regarded as most appropriate microbial producers as capable of maintaining redox balance during xylitol accumulation (Koutinas *et al.*, 2014). Yeasts are the best producers of microbial oils and fats, called single cell oils (SCOs) (Koutinas *et al.*, 2014). Yeast SCOs are mainly composed of C14-C18 fatty acids (such as saturated myristic, palmitic, stearic, oleic acids) of neutral fractions (triglycerides and to a lesser extent steryl-esters) and could be a potential substitute for natural oils and fats as feedstock for chemical production, including biodiesel (Koutinas *et al.*, 2014; Bandhu *et al.*, 2020). Compared with extraction from oil crops and plants, microbial production of lipids has many advantages including short production cycle, tailored processes and better accessibility to structural diversity (Wen *et al.*, 2020). Oleochemicals can be used to produce biofuels, surfactants, waxes, surface coatings, lubricants, cosmetics, and plastics, etc. Oleaginous yeasts of *Lipomyces*, *Yarrowia* and *Rhodotorula* genera can natively produce lipid content up to 70% (w/w) and grow on wide range of substrates and thus have been used as model systems to understand genetic control of lipid productivity (Li, Zhao and Bai, 2007; Angerbauer *et al.*, 2008; Ageitos *et al.*, 2011).

Oleaginous yeasts can use a variety of substrates as carbon sources – pure sugars, sugar-enriched wastes (cheese-whey, molasses), vegetable oils, crude industrial saturated fatty acids, waste cooking oils, glycerol, mixtures of hydrophilic fats or oils , and lignocellulosic biomass derived sugars (Papanikolaou and Aggelis, 2011b). Lignocellulosic biomass is an abundant, non-edible organic material in nature. It contains ~40-50% cellulose, a glucose polymer; ~25-35% hemicellulose, a sugar heteropolymer; and ~15-20% lignin, a non-fermentable phenyl-propene unit; plus lesser amount of minerals, oils, soluble sugars, and other components (Holtzapple, 1993). To use it for microbial conversion, lignocellulosic biomass requires pre-treatment to obtain fermentable sugars (Jin *et al.*, 2015). In biorefineries, cellulose fraction of lignocellulosic biomass is used for fuel production, mainly bioethanol. Although hemicellulose and lignin are commonly used for energy production, higher economic efficiency can be reached if all components of lignocellulose will be maximally valorised. Therefore, the integration of hemicellulosic sugars (pentoses, C5) in biorefinery concepts is important to reach higher economical efficiencies. Xylose is the second most abundant sugar in nature after glucose (Sànchez Nogué and Karhumaa, 2015). Xylose and arabinose are part of this renewable and low-cost raw materials interesting alternative to conventional carbon sources (Papanikolaou and Aggelis, 2011b).

Hemicellulosic fraction of lignocellulose biomass could be a promising substrate for the production of microbial oils (SCOs) and oleochemicals.

In the context of biorefineries, it is important that the microorganisms can consume and tolerate the components of the biomass hydrolysate, such as the yeast *Rhodotorula toruloides*. With the advances in biotechnology, genomics and systems biology, microbial production of chemicals can be significantly improved with engineered cell factories in terms of substrate utilization and production of selected chemicals, compared with the metabolic capabilities of native microorganisms.

1.2 Rhodotorula toruloides

Rhodotorula toruloides (previously known as *Rhodosporidium toruloides*) is a non-pathogenic, aerobic, oleaginous red yeast, known as a great producer of microbial (Ageitos *et al.*, 2011; Sampaio, 2011) and carotenoids (Dias *et al.*, 2015). Oleaginous microorganisms are capable of accumulating at least 20% of their dry cell mass as lipids (Thevenieau and Nicaud, 2013), but *R. toruloides* can accumulate up to 76% (Li *et al.*, 2006). The mechanism of SCOs production by oleaginous yeasts has been widely studied.

SCOs accumulation is achieved using either sugars or hydrophobic compounds as carbon sources. Higher lipid accumulation can be achieved under carbon excess and nutrient like nitrogen, sulphur (Wu *et al.*, 2011), or phosphorous (Wu *et al.*, 2010) limitation. Nitrogen limitation leads to rapid decrease in cellular adenosine monophosphate (AMP). Under these conditions, large amounts of tricarboxylic acid (TCA) cycle intermediate citric acid are produced, but not further catalysed via TCA cycle. The latter components are secreted to the cytoplasm and cleaved by ATP citrate lyase (ACL) – a key enzyme characterizing an oleaginous microorganism. ACL cleaves citric acid into acetyl-CoA and oxaloacetate, and acetyl-CoA is used by fatty acid synthetase to generate cellular fatty acids and subsequently triglycerides (Papanikolaou and Aggelis, 2010, 2011a, 2011b). Without ACL (non-oleaginous microorganisms) citric acid is secreted into growth medium that subsequently leads to the accumulation of polysaccharides (Ratledge and Wynn, 2002; Papanikolaou and Aggelis, 2011b).

R. toruloides has been mostly studied in regard to its lipid accumulation capabilities. Production yields and composition of microbial lipids change depending on cultivation conditions. High carbon/nitrogen (C/N) ratio improves lipid production (Ratledge and

Wynn, 2002) and under different carbon sources, the highest lipid content in biomass was achieved when grown on xylose as a sole carbon source (Lopes *et al.*, 2020). *R. toruloides* is a natural producer of other high-value compounds, such as carotenoids (Dias *et al.*, 2015).

R. toruloides can consume 2nd generation carbon sources, such as glucose (Zhang, Ito, *et al.*, 2016), xylose (Wiebe *et al.*, 2012), acetate (Huang *et al.*, 2016), glycerol (Xu *et al.*, 2012), and complex biomass mixtures, such as non-detoxified sugarcane bagasse hydrolysate (Bonturi *et al.*, 2017), lignocellulosic hydrolysates (Fei *et al.*, 2016; Lopes, Bonturi and Miranda, 2020). It can tolerate toxic compounds commonly found in complex biomass substrates (Palmqvist and Hahn-Hägerdal, 2000; Lopes, Bonturi and Miranda, 2020), such as HMF, furfural, acetic acid that can inhibit cell growth in most conventional yeasts (Palmqvist and Hahn-Hägerdal, 2000; Lopes, Bonturi and Miranda, 2020). Among other non-conventional yeasts, *R. toruloides* demonstrated the best growth characteristics when grown on concentrated hemicellulosic hydrolysate from birch (Monteiro de Oliveira *et al.*, 2021).

To facilitate the use of various synthetic and systems biology tools with *R. toruloides*, complete genome sequences have been determined for different strains (Kumar *et al.*, 2012; Zhu *et al.*, 2012; Morin *et al.*, 2014; Hu and Ji, 2016; Zhang, Skerker, *et al.*, 2016; Sambles *et al.*, 2017; Tran *et al.*, 2019). A haploid strain NP11, was the first full genome sequence from *de novo* assembly for *R. toruloides* (Zhu *et al.*, 2012). Strain IFO0880, also a haploid strain, was first annotated by (Zhang, Skerker, *et al.*, 2016), revised by (Coradetti *et al.*, 2018), and now IFO0880 v4.0. is accessible from Joint Genome Institute's Mycocosm genome portal (Nordberg *et al.*, 2014) (available at mycocosm.jgi.doe.gov). (Coradetti *et al.*, 2018) performed fitness analysis of gene deletion or disruption mutants within pooled populations, improving the original genome annotation (Zhang et al. 2016) with PacBio sequencing. Strains NP11 and IFO0880 metabolism is different, but still "closely related" (Zhang, Skerker, *et al.*, 2016). Both strains IFO0880 (Zhang, Ito, *et al.*, 2016; Zhang, Skerker, *et al.*, 2016) and NP11 (Sun *et al.*, 2017; Yang *et al.*, 2018) have been successfully used in metabolic engineering studies to improve lipid production. *R. toruloides* CCT7815 was obtained by the adaptation of the strain CCT0783 (synonym IFO10076) in sugarcane bagasse hydrolysate (Bonturi *et al.*, 2017). The genomic DNA of CCT0783 was sequenced using MiSeq platform and annotated against reference NP11 gene sequences (deposited at DDBJ/ENA/GenBank under the accession JABGON000000000i). Genomic analysis showed that CCT0783 is possibly a multiploid strain, genome sequence with one of its alleles

showing over 98% similarity with NP11, while the other allele shows around 80% of similarity (Bonturi et al., under preparation).

Systems biology methods, including genome-scale modelling and multi-omics analysis, have been used for a holistic understanding of *R. toruloides* metabolism and, in particularly, its oleaginous phenotype. Genome-scale models (GEMs) provide an organism-specific summary of metabolic network that contains all the known metabolic reactions derived from an annotated genome sequence and its gene-protein relationships for the organism of interest (Orth, Thiele and Palsson, 2010; Kerkhoven, Lahtvee and Nielsen, 2015). GEMs are not only valuable for testing different hypothesis *in silico* before initiating *in vivo* laboratory experiments, but also for predicting yields or even suggesting targets for gene expression changes or knockouts (Kerkhoven, Lahtvee and Nielsen, 2015; O'Brien, Monk and Palsson, 2015; Zhang and Hua, 2016). To date, GEMs have been reconstructed for *R. toruloides* strains NP11 (rhto-GEM (Tiukova, Prigent, *et al.*, 2019)) and IFO0880 (iRhto1108 (Dinh *et al.*, 2019; Kim *et al.*, 2021)). These GEMs were used to design metabolic engineering strategies for enhanced triacylgycerol (TAG) (Dinh *et al.*, 2019), linolenic acid and carotenoids production in (Tiukova, Prigent, *et al.*, 2019), and to study the phenotypic changes under nitrogen limitation (Dinh *et al.*, 2019). In the most recent studies, rhto-GEM (Tiukova, Prigent, *et al.*, 2019) was used to study metabolic mechanisms for the utilization of three different carbon sources with the focus on lipid production (Lopes *et al.*, 2020) and to explain metabolic mechanisms involved in the xylose assimilation (Pinheiro *et al.*, 2020). The utilisation of other substrates that can be often found in lignocellulosic biomass hydrolysate was investigated using the GEM by (Kim *et al.*, 2021).

Enabled by high-throughput technologies, such as genomics, transcriptomics, proteomics, metabolomics, and also fluxomics, large-scale biological data sets have been crucial for a better understanding the phenotypic characteristics of *R. toruloides* (Zhu *et al.*, 2012; Coradetti *et al.*, 2018). To improve the lipid yield in *R. toruloides*, metabolic pathways, such as acetyl-CoA production (Zhu *et al.*, 2012), ACL (Evans and Ratledge, 1985), mitochondrial and peroxisomal β oxidation (Hiltunen *et al.*, 2003), and fatty acid synthase (FAS) system, have been investigated using proteomics analysis during growth on glucose (Liu *et al.*, 2009; Shi *et al.*, 2013) and xylose (Tiukova, Brandenburg, *et al.*, 2019). Fatty acid biosynthesis is often limited by the supply of NADPH, and these proteomics studies have shown differences in enzyme levels involved in NADPH regeneration. Canonical fatty acid biosynthesis starts with the conversion of acetyl-CoA into malonyl-CoA by acetyl-CoA

carboxylase (ACC). Fatty acids are synthesized from acetyl-CoA and malonyl-CoA, using latter as the extending unit. Each elongation of two carbon units in fatty acid biosynthesis requires the oxidation of two NADPH molecules (Lian and Zhao, 2015). In oleaginous yeasts, the main candidate enzymes involved in NADPH generation are two enzymes of the pentose phosphate pathway (PPP) (glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase), cytosolic malic enzyme (ME), and NADP+-dependent isocitrate dehydrogenase, if it occurs in the cytosol (Ratledge, 2014). Malic enzyme is known to supplement NADPH for *de novo* lipogenesis from the early biochemical studies (Wynn, Hamid and Ratledge, 1999). In addition to fatty acid biosynthesis, xylose utilisation requires NADPH. Xylose reductase (XR, NADPH-dependent) reduces xylose to xylitol, which is then oxidized by xylitol dehydrogenase (XDH, NADH-dependent) to xylulose (Tiukova, Brandenburg, *et al.*, 2019). Recently, multi-omics analysis suggested an alternative metabolic pathway for xylose utilisation in strain IFO0880 involving D-arabinitol and D-ribulose forming ribulose-5-phosphate instead of the known pathway forming D-xylulose-5-phosphate (Kim *et al.*, 2021).

Despite these findings, there is still a lack of complete understanding of *R. toruloides* metabolism. Current stoichiometric models of *R. toruloides* can capture all routes for conversion of substrates into biomass and globally balance cofactor needs (Dinh *et al.*, 2019). However, they cannot link enzyme levels with metabolite concentrations and metabolic fluxes (Dinh *et al.*, 2019). This study is aimed at development of a genome-scale model with direct integration of enzyme abundances, coupled with the analysis of large-scale proteomics data sets representing protein levels at multiple environmental conditions.

1.3 Overview of modelling methods

The first models of metabolic pathways were based on single enzyme kinetics described by ordinary differential equations (Liao, 1993). In the late 1990s, after the rise of whole-genome sequencing technologies, the field of constraint-based modelling emerged (Kerkhoven, Lahtvee and Nielsen, 2015). Nowadays, functions of most of the metabolic genes in an organism are identified and assigned using high-throughput sequencing technology together with the tools for automated genome annotation. Whole-cell metabolic network is reconstructed based on its genome annotation and curated using databases and strain-specific physiological information available in literature. Once the protein functions are integrated

into a metabolic network, it can be used to analyse, predict and interpret cellular behaviour (**Figure 1**).

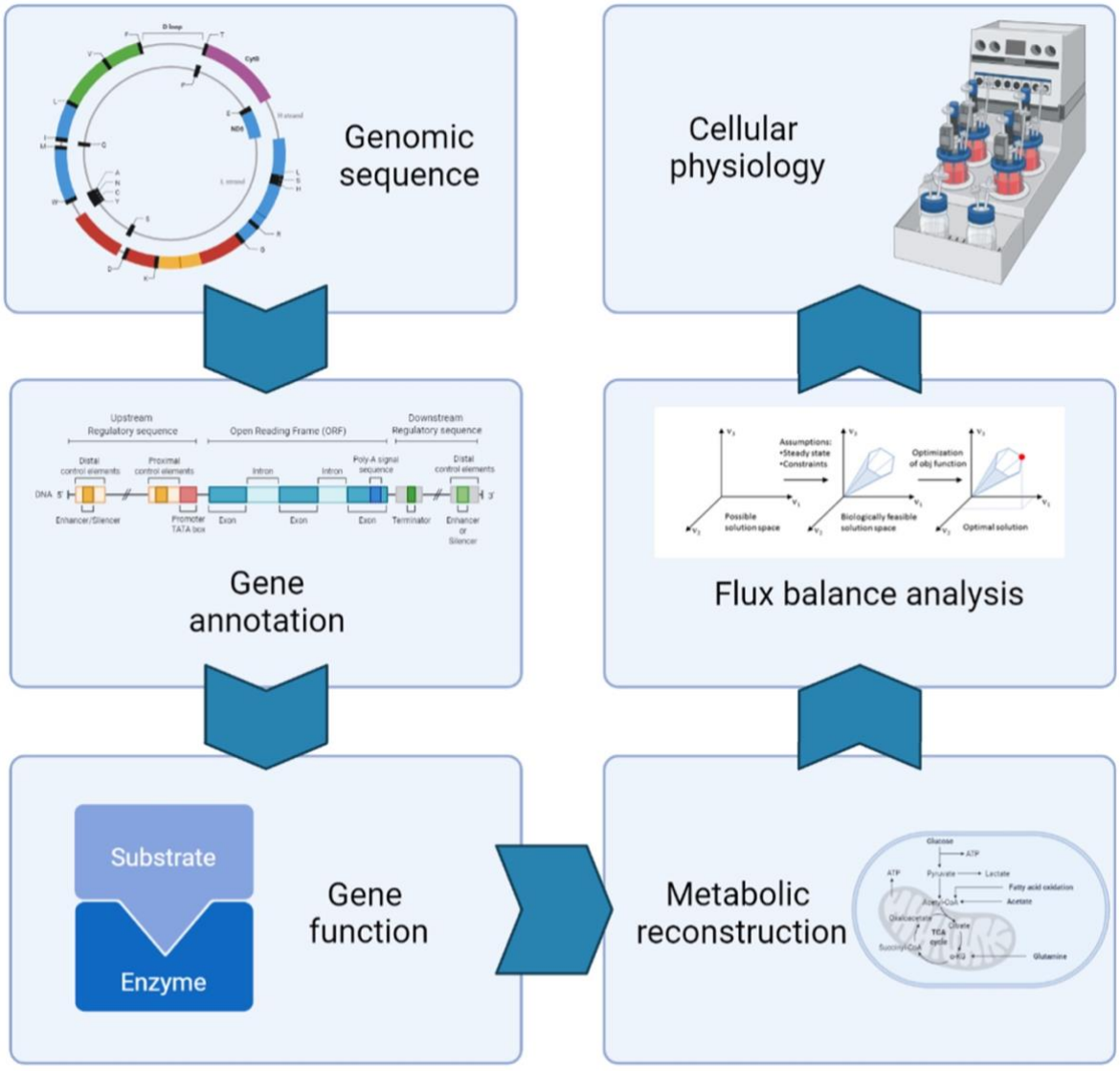

Figure 1. Simplified scheme of a holistic approach in biology. After identification and functional assignment of metabolic genes in organism they are integrated into a metabolic network which can then be subjected to the methods such as flux balance analysis to analyse, interpret and predict cellular behaviour.

Computational modelling is defined in a form of dynamic mass balances – metabolite concentration equals to the difference between the rates at which the metabolite is produced and consumed (Edwards, Covert and Palsson, 2002). Dynamic mass balances are formally

analogous to Kirchhoff's first law for electrical circuits (Edwards, Covert and Palsson, 2002) and can be represented as follows:

$$S \cdot v = b \tag{1}$$

where **S** is a matrix consisting of stoichiometric coefficients of all metabolic reactions where rows represent metabolites and columns represent individual reactions, **v** is a vector of metabolic rates of n reactions, and **b** is a vector containing the net metabolite uptake by the cell (Varma and Palsson, 1994).

If a complete kinetic information is not available, as it happens for most reactions, the scope of dynamic analysis is limited. In that case, metabolic quasi-steady state is assumed, which means that all metabolic fluxes leading to the formation and degradation of any metabolite must be balanced. It is assumed that metabolic transients are typically rapid compared to cellular growth rates and environmental changes (Varma and Palsson, 1994). Therefore, steady state mass balance can be written as:

$$S \cdot v = 0 \tag{2}$$

Steady state can be reached in chemostat cultivation experiments and, in theory, during the controlled batch logarithmic growth phase when organisms grow at their maximum specific growth rate (Kerkhoven, Lahtvee and Nielsen, 2015). As normally the number of fluxes exceeds the number of metabolites, system is underdetermined and cannot solve equations algebraically.

1.3.1 Constraint-based modelling

At the heart of constraint-based modelling are stoichiometric coefficients of each metabolic reaction represented stoichiometric matrix (**S**). Genome-scale models (GEMs) are genome-wide constraint-based models constrained by (i) stoichiometry of network; (ii) substrate uptake rates but can be also thermodynamics; (iii) assumption of a steady state (Kerkhoven, Lahtvee and Nielsen, 2015). Mass balances can be considered "hard-wired" constraints, experimental measurements – constraints to the specific condition. (Edwards, Covert and Palsson, 2002). Due to steady-state assumption, commonly available optimization techniques, such as Flux Balance Analysis (FBA), are applicable to solve the optimization problem of the network. Constraint-based modelling is easier to implement and has become very popular for many applications from biomedicine to industrial biotechnology.

1.3.1.1 Flux Balance Analysis

Reconstruction of a metabolic network is not sufficient to specify the metabolic phenotypes that will be expressed under certain environmental conditions. Metabolic phenotypes can be defined in terms of flux distribution (fluxome) through a metabolic network, obtained using mathematical calculation and computer simulation techniques (Edwards, Covert and Palsson, 2002). FBA is a tool for the analysis of GEMs as a linear programming problem, where by setting an objective function it seeks its maximal value within stoichiometrically defined domain (Varma and Palsson, 1994). FBA can identify a single optimal flux distribution that lies on the edge of allowable solution space (**Figure 2**) (Orth, Thiele and Palsson, 2010):

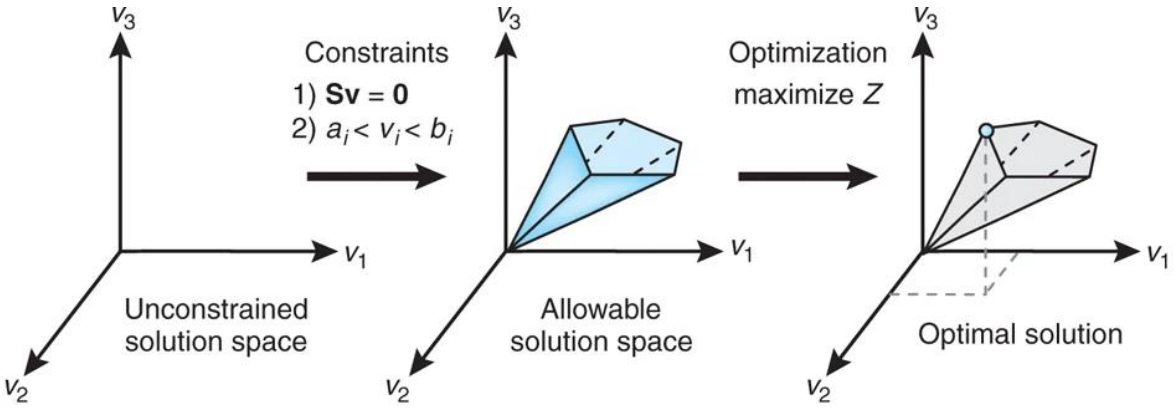

Figure 2. Conceptual basis of constraint-based modelling and FBA. With no constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints imposed by the stoichiometric matrix **S** (1) and capacity constraints imposed by the lower and upper bounds (a_i and b_i) (2) are applied to a network, it defines an allowable solution space denying points outside this space. (Orth, Thiele and Palsson, 2010).

FBA is implemented in software tools for constraint-based modelling that all depend on at least one numerical optimization solver (Heirendt *et al.*, 2019). Commonly used objective function maximizes specific growth rate, ATP generation or product formation (Burgard and Maranas, 2003; Schuetz, Kuepfer and Sauer, 2007):

- biomass reaction is an artificial, lumped reaction representing metabolite consumption in proportion of cell growth, including nucleoside triphosphate (NTP) requirements for mRNA, amino acid requirements for proteins, lipid requirements for the cell wall, and metal ion needs (Salvy and Hatzimanikatis, 2020);
- ATP maintenance requirements are split into growth-associated energy costs (GAEC) that quantify energy costs that are not captured in the biomass equation and non-growth associated maintenance (NGAM) that accounts for (i) shifts in metabolic pathways, (ii) energy spilling reactions, (iii) cell motility, (iv) changes in stored polymeric carbon, (v) osmoregulation, (vi) proofreading, synthesis and turnover of macromolecular compounds, (vii) defence against O_2 stress (van Bodegom, 2007).

1.3.1.2 Enzyme-constrained models

To address the oversimplified assumption associated with FBA-based model predictions that the uptake rate of carbon source limits production and to capture protein level-related limitations, further improvements in GEMs in different directions have been developed. This can be achieved by: (i) global biochemical constraints, or (ii) direct integration of omics data. Models with improved prediction accuracy have mainly been developed for well-characterized microorganisms, such as *Escherichia coli (E. coli)* (Salvy and Hatzimanikatis, 2020) and *Saccharomyces cerevisiae* (*S. cerevisiae*) (Oftadeh *et al.*, 2021). Most advanced metabolic models have been able to simulate biologically complex phenotypes, such as Crabtree effect (Sánchez *et al.*, 2017), maximal growth rate (Niebel, Leupold and Heinemann, 2019). The next challenge, however, is conveying these advanced approaches to non-model organisms, such as *R. toruloides*.

Different approaches have been developed to account for enzymatic limitations in metabolic models. Metabolomics and expression modelling (ME-models) approach includes all processes required for the synthesis of functional proteins starting from the transcription rates of genes (Lloyd *et al.*, 2018). In FBA with molecular crowding (FBAwMC) approach, total proteome capacity constraint is applied without direct integration of proteomics data (Beg *et al.*, 2007; van Hoek and Merks, 2012; Nilsson and Nielsen, 2016). The first approach that allows for a direct incorporation of proteomics data to account for enzyme limitation was the GECKO framework (abbreviated from *G*EM with *e*nzymatic *c*onstraints using *k*inetic and *o*mics data) (Sánchez *et al.*, 2017).

The GECKO method is built on the principle that any metabolic reaction flux has a biologically natural constraint equal to the intracellular enzyme concentration multiplied by the enzyme's turnover number (k_{cat}). In metabolic network, the enzyme constraint is defined as maximum rate of enzymatic reaction (v_{max}) that the metabolic flux cannot exceed calculated as:

$$v \leq k_{cat} \cdot [E] \tag{3}$$

where ***v*** is flux of reaction in mmol (gDCW h)$^{-1}$, k_{cat} is enzyme's turnover number or specific catalytic rate (h)$^{-1}$, and ***E*** is the concentration of an enzyme in mmol gDCW^{-1}.

In GEMs, metabolism is represented in **S**-matrix, where columns represent stoichiometric coefficients, rows indicate mass balances. In GECKO approach, genome-scale modelling is extended by representing enzymes as limited capacity entities in each reaction. New rows are added to **S**-matrix representing enzymes, new columns are added representing enzyme usage (**Figure 3**).

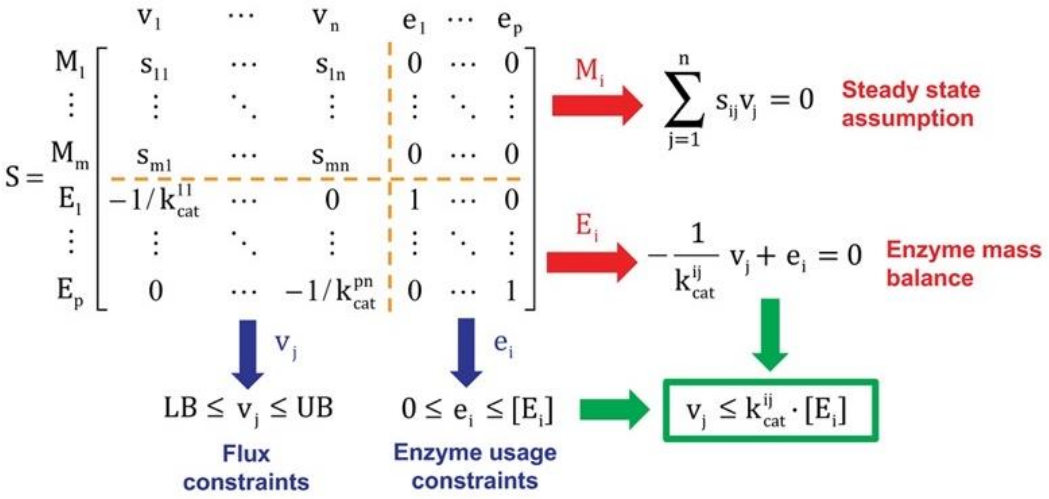

Figure 3. GECKO framework includes 4 submatrices that appear inside a new stoichiometric matrix: upper left equivalent to the original S matrix, upper right has only zeros, lower left with kinetic information, lower right is identity matrix (Sánchez *et al.*, 2017).

With these formalisms, enzymes are not consumed, but rather occupied. Given the steady-state assumption, for a fraction of second, there is a limited amount of enzyme occupied by its substrates to catalyse the corresponding flux. Using the GECKO pipeline requires detailed strain physiological parameters that are applied to (i) automatically modify limiting enzyme

abundances; (ii) rescale biomass equation, and (iii) fit growth (GAEC) and nongrowth-associated energy costs (NGAEC). Energy requirements are estimated by maximizing for ATP production under carbon substrate uptake rate and growth rate constraints. In case of incomplete proteomic data, enzyme constraints are replaced with pseudo-metabolite that acts as an enzyme pool – total mass constraint similar to molecular crowding formalism in FBAwMC approach (Beg *et al.*, 2007). Enzyme turnover numbers are automatically queried from BRENDA (Schomburg *et al.*, 2012).

Using the GECKO toolbox, enzyme-constrained models (ecModels) have been generated for *S. cerevisiae*, *Y. lipolytica*, *Kluyveromyces marxianus (K. marxianus)*, *E. coli* and *Homo sapiens* (Domenzain *et al.*, 2021). In this study, condition-specific enzyme-constrained genome-scale model for *R. toruloides* was developed with an aim to study the differences in metabolism under the consumption of xylose, glucose and acetic acid and to identify the bottlenecks of the conversion of these three carbon substrates into biomass or other valuable products.

2 THE AIMS OF THE THESIS

To design and establish microbial cell factories using non-conventional oleaginous yeast *R. toruloides*, systemic investigation of *R. toruloides* metabolism is necessary. Enzyme-constrained genome-scale metabolic modelling allows for a systemic investigation of cell metabolism by integration of proteomics and kinetic data. The existing metabolic network reconstructions of *R. toruloides* provide for the basis for further development of condition-specific metabolic models.

The aims of this thesis are:

- Generating enzyme-constrained genome-scale model of *R. toruloides*;
- Using model simulations, to understand the main sources of NADPH during lipid accumulation on different carbon substrates.

3 EXPERIMENTAL PART

3.1 MATERIALS AND METHODS

3.1.1 Batch growth experiments

3.1.1.1 Strain, inoculum and media

R. toruloides CCT7815 (Coleção de Culturas Tropicais, Fundação André Tosello, Campinas, Brazil) was used throughout this study due to its increased lipid production and induction of hydrolysate-tolerance and lipid accumulation genes without physiological changes regarding growth and substrate consumption (Bonturi *et al.*, 2017), deeming this strain potentially useful for biorefinery applications.

Pre-inoculum was prepared as described in Pinheiro et al. (2020). Inoculum for the cultivation was prepared in batch cultivation on basal mineral medium supplemented with a single carbon source (either glucose or acetic acid) and one nitrogen source (ammonium sulphate) in duplicate flasks at 200 rpm and 30°C for 24 h. Before inoculation, cells were washed with 0.9% (m/v) NaCl solution. The basal mineral medium contained: 3.0 g/L KH_2PO_4, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 1 mL/L vitamin solution, and 1 mL/L trace metal solution (Lahtvee *et al.*, 2017). This medium was supplemented with sole carbon source of 18.2 g/L glucose or 20.0 g/L acetic acid and 5 g/L $(NH_4)_2SO_4$. The carbon/nitrogen (C/N) molar ratio of the medium was 8.8.

In bioreactor experiments, basal mineral medium was supplemented with either 63.6 g/L glucose and 0.9 g/L urea, or 20.0 g/L acetic acid and 0.6 g/L $(NH_4)_2SO_4$, and 0.1 mL/L antifoam 204 (Sigma-Aldrich, St. Louis, MO, United States). The (C/N) molar ratio of the media were set to 69 and 80, respectively, similarly as in experiments on xylose by (Pinheiro *et al.*, 2020).

3.1.1.2 Yeast cultivation

Cultivation in a batch growth regime was performed in 1-L bioreactors (Applikon Biotechnology, Delft, the Netherlands) with a working volume of 900 mL at pH 6.0 controlled by the addition of 2 mol/L KOH. Dissolved oxygen was maintained not lower than 25% at 1-vvm airflow by regulating stirring speed between 400 and 600 rpm. CO_2 and O_2 outflow gas composition were measured using an online gas analyser (BlueSens gas sensor GmbH, Herten, Germany). Cell turbidity was monitored on-line using Bug Lab BE3000 Biomass Monitor (Bug Lab, Concord, CA, United States) at 1300 nm and off-line

using UV/Vis spectrophotometer at 600 nm (U-1800, Hitachi High-Tech Science, Tokyo, Japan). Data collection and processing was performed with BioXpert_V2 software **v2.95** (Applikon Biotechnology, Delft, the Netherlands).

Cells were inoculated at 0.4 OD600. Samples for dry cell weight were collected every 6 hours during the exponential growth phase and every 24 hours during the nitrogen limitation phase. Samples for the extracellular metabolite, carotenoid, lipid, total cellular protein content and proteomics analyses were collected in every 3 hours during the exponential growth phase and in every 24 or 48 hours during the nitrogen limitation phase. Samples were prepared and stored as previously described (Pinheiro *et al.*, 2020). Samples were taken from bioreactors to 2-mL tubes, centrifugated for 30 s at 4°C and 18000×g. The supernatant was stored at -20°C for extracellular metabolite analyses. Cell pellets were snap-frozen in liquid nitrogen and stored at -80°C for the analyses of biomass composition and proteomics. All experiments were performed in duplicate.

3.1.1.3 Analytical methods

Cell mass was expressed as dry cell weight (DCW) measured gravimetrically and used for the calibration of cell optical density data. Extracellular metabolites were quantified using HPLC (LC-2030C Plus, Shimazu, Kyoto, Japan) equipped with a refractive index detector (RID-20A, Shimadzu, Kyoto, Japan). The concentrations of xylose, glucose, organic acids and glycerol were determined using Rezex ROA Organic Acid column (Phenomenex, Torrance, United States) with isocratic elution of 5 mmol/L H_2SO_4 at a flow rate of 0.6 mL/min and at 45°C. Yields and specific consumption and production rates were calculated considering the consumption and production in each growth phase separately starting from when the carbon source was consumed.

Fatty acid content was determined using quantitative gas chromatography-mass spectrometry (GC-MC) analysis with the internal standard method, similar as described in (Tammekivi *et al.*, 2019). The derivatisation procedure of fatty acids was based on (Tammekivi *et al.*, 2021). Fatty acid composition was profiled by using an Agilent 7890A GC instrument connected to an Agilent 5975C inert XL MSD with a triple-axis detector, an Agilent G4513A autosampler and capillary column Agilent DB-225MS (30 m x 0.25 mm diameter, 0.25 µm film thickness) with a (50%-cyanopropylphenyl)-methylpolysiloxane stationary phase. Commercial standard mixture of fatty acid methyl esters (FAME, C8-C24,

Supelco) was used for fatty acid identification and quantification. Finally, the obtained values were recalculated to represent the concentrations of a TAG molecule, where all the bonded fatty acids are the same. The sum of the quantified fatty acids was presented as the total lipid content. Fatty acids were quantified at late exponential growth phase and at the end of nitrogen limitation phase. The results represent cumulative value for each phase.

Carotenoids were quantified at the late exponential growth phase, mid-nitrogen limitation phase and at the end of nitrogen limitation phase. The results were calculated to represent cumulative value for each phase. The extraction and quantification of carotenoids was performed as described in (Pinheiro *et al.*, 2020) using Acquity Ultra Performance Liquid Chromatography (UPLC) (Waters, Franklin, MA, United States) instrument equipped with a TUV detector (Waters, Franklin, MA, United States) and C18 column (BEH130, 1.7 μm, 2.1 x 100 mm, Waters, Franklin, MA, United States). β-carotene standard (Alfa Aesar, Tewksbury, MA, United States) was used to identify all peaks according to the known carotenoid retention time profile detected at 450 nm (Lee *et al.*, 2014).

Total cellular protein content was quantified at mid-exponential growth phase and mid-nitrogen limitation phase. The results represent cumulative value for each phase. Sample preparation and protein extraction were performed as described in (Kumar and Lahtvee, 2020). Commercially available colorimetric assay (Micro BCA™ Protein Assay Kit, Thermo Fisher Scientific, Waltham, MA, United States) with calibration curve of bovine serum albumin (BSA) standard of linear range dilutions from 0.5 to 200 $\mu g/mL$ was used to measure protein concentration.

Absolute proteome analysis was performed using fully labelled cellular biomass as an internal standard in the measurements. To produce the internal standard, *R. toruloides* was cultivated in a minimal mineral medium containing labelled heavy 15N, 13C-lysine (Silantes, Munich, Germany), resulting in 96.6% heavy labelling of proteogenic lysine ((Pinheiro *et al.*, 2020), data not shown). Proteome quantification was done using nanoscale liquid chromatography with tandem mass spectrometry, as described in (Sánchez *et al.*, 2021).

3.1.2 Quantification of proteome and bioinformatics

Mass-spectrometric raw data were identified with MaxQuant **v1.6.1.0** software package (Tyanova, Temu and Cox, 2016). Data were searched against reference proteome database

R. toruloides NP11 proteome (Zhu *et al.*, 2012) in Uniprot (www.uniprot.org). Normalized total protein approach (TPA) as described in (Sánchez *et al.*, 2021) was used for absolute protein quantification. MS intensities of heavy-labelled internal standard were normalized with the number of theoretically observable peptides using the iBAQ (intensity based absolute quantification) feature in MaxQuant (Schwanhäusser *et al.*, 2011). The resulting "iBAQ" intensities were additionally normalized for the sum of intensities for samples of the same carbon source. For computing absolute protein abundances, the sum of MS intensities of all detected proteins multiplied by corresponding molecular weights was assumed to be proportional to the total amount of protein injected, assuming 80% coverage from the total protein abundance.

3.1.3 Genome-scale modelling

Enzyme-constrained genome-scale metabolic model of *R. toruloides* was generated using the previous metabolic network rhto-GEM **version 1.3.0** (Tiukova, Prigent, *et al.*, 2019). Enzyme-constrained GEM, named ecRhtoGEM, was generated using a semiautomated workflow provided by the GECKO toolbox **version 2.0.2** (Sánchez *et al.*, 2017) on MATLAB (The MathWorks Inc., Natick, MA, Unites States). All scripts and models are available on a dedicated Github repository (https://github.com/alinarekena/ecRhtoGEM/), accessible upon request sent to alina.rekena@ut.ee).

Condition-specific models were generated by incorporating experimental data, including absolute protein quantification obtained from the yeast cultivations in this study (**Section 3.1.1**) and by (Pinheiro *et al.*, 2020). Data were formatted according to GECKO requirements and are provided in the ecRhtoGEM/customGECKO folder. For the manual curation of automatically retrieved enzyme kinetic data, the GECKO toolbox template *manualModifications* was modified for ecRhtoGEM and stored in the ecRhtoGEM/customGECKO folder in the repository. Targets for the curation of enzyme kinetic data were identified using the GECKO toolbox script *topUsedEnzymes*.

The biomass equation from rhtoGEM was fitted to include the measured total cellular protein and lipid content (**Section 3.1.1.3,** ecRhtoGEM/data folder), while the carbohydrate component was scaled accordingly, and nucleotide levels were remained unchanged. To enable simultaneous rescaling of lipid and protein fractions in the biomass, function s*caleLipidProtein*, available in the ecRhtoGEM/code folder, was integrated into GECKO

pipeline. The lipid equation was fitted to the measured fatty acid chain distribution (**Section 3.1.1.3**), while lipid class distributions were left unmodified from rhto-GEM (Tiukova, Brandenburg, *et al.*, 2019). To enable lipid chain adjustment, functions *loadLipidChainData* and *scaleLipidsRhto* were integrated into the GECKO pipeline via the function *scaleLipidProtein* (available in the ecRhtoGEM/code folder). The growth-associated energy costs (GAEC) and non-growth associated maintenance (NGAM) were fit to measured carbon source uptake rates from exponential growth and nitrogen limitation phases and set from 124.4 to 140.0 mmol/gDCW and from 0 to 3.65 mmol/(gDCW h), respectively (**Supplementary Table S1**).

The model was improved by adding translation pseudo-reaction (*reconstruct_ecRhtoGEM* (part III), available in the main page of ecRhtoGEM repository) and an alternative D-xylose utilisation pathway (*edit_rhtoGEM*). With the translation reaction, ribosomal subunits were introduced in the model as pseudo-metabolites and included as a new amino acid pseudo-metabolite to the protein pseudo-reaction. Translation reaction was added to the enzyme-constrained models separately from the GECKO pipeline, all details described in the dedicated script *reconstruct_ecRhtoGEM*, part III. All changes to the metabolic network were introduced using *addRxns*, *addMets*, and *addGenesRaven* functions the RAVEN toolbox **version 2.4.3** (Wang *et al.*, 2018).

Intracellular metabolic flux patterns were predicted using the FBA from the RAVEN toolbox and Gurobi solver (Gurobi Optimization Inc., Houston, Texas, United States). The objective function was set as minimisation of the total usage of unmeasured proteins, assuming that the regulatory machinery for stress tolerance is represented by the condition-specific protein expression profile (Domenzain *et al.*, 2021). Flux variability analysis was done with random sampling (at n=2000) of the solution space (Bordel, Agren and Nielsen, 2010), allowing 10% variability from the maximal protein pool value, 1% variability from maximal growth rate and carbon uptake rate, 1% variability from minimal NGAM rate, 10% variability from predicted carbon dioxide production and oxygen consumption rate, and 10% variability from measured by-product rates. This resulted in average flux with standard deviation, representing the flux variability. Finally, obtained intracellular flux patterns were mapped to the original reactions as they were defined in the non-ecModel to adjust for the flux of the reversible reaction, if any (function *mapRxnsToOriginal*, script *analyze_ecRhtoGEM*).

3.2 RESULTS

3.2.1 Growth characterisation on different carbon substrates

Detailed physiology characterisation of *R. toruloides* CCT7815 and absolute proteomics analysis was carried out with xylose, acetic acid and glucose as sole carbon sources under aerobic batch conditions on a minimal mineral medium. Physiological data with xylose as carbon source were taken from (Pinheiro *et al.*, 2020), repeating the absolute proteome quantification and fatty acid analysis. All results are presented together. Cells were cultivated on glucose and acetic acid, similarly as described in (Pinheiro *et al.*, 2020). Lower acetic acid concentration (20 g/L) was used because in higher concentrations acetic acid is toxic to the cell membrane (Royce *et al.*, 2013). Cell turbidity, CO_2 production and O_2 consumption were monitored on-line. In case of acetic acid and for technical reasons, off-line cell turbidity measurements were additionally taken. Biomass composition (lipids, proteins, carotenoids), metabolites consumption and production were analysed off-line (**Figure 4**). As described in (Pinheiro *et al.*, 2020), a high C/N ratio (80 mol/mol) was used in this study. Regardless of the carbon source, two distinct growth phases were observed from the growth dynamics and substrate consumption data (**Figure 4B**, **4D**, and **4F**, dashed line). The transition of the exponential growth phase (exp) into nitrogen limitation phase (Nlim) was marked by a sharp decrease in a specific growth rate (μ) at the time point when nitrogen was possibly depleted (**Figure 4A**, **4C**, and **4E**). The nitrogen limitation phase lasted until the depletion of the carbon source. Both phases, exponential (exp) and nitrogen limitation (Nlim), were further analysed in this study. The third cultivation phase on xylose, characterised by the consumption of the secreted by-products xylitol and arabinitol, as described in (Pinheiro *et al.*, 2020), was not analysed in this study.

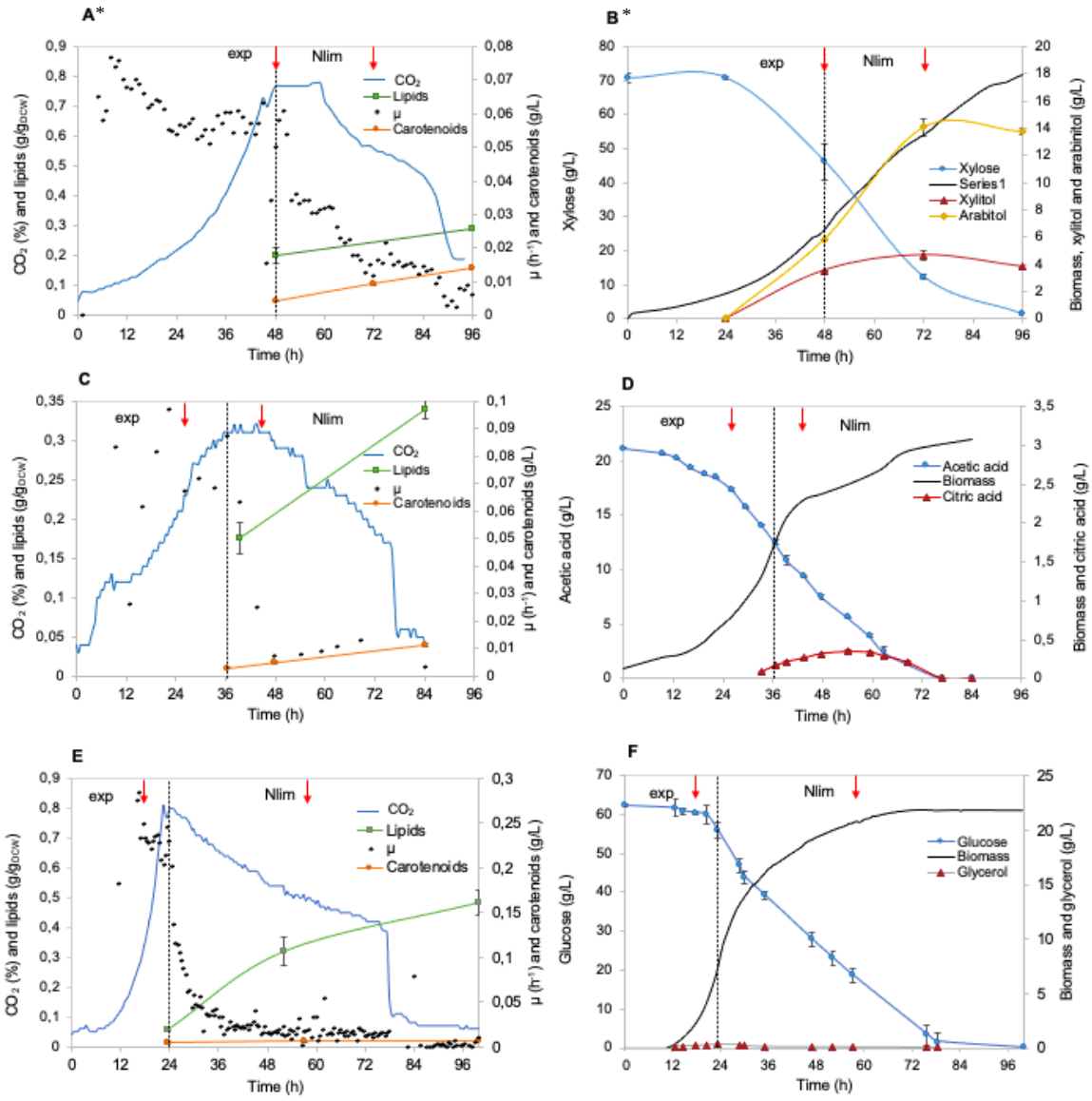

Figure 4. *R. toruloides* CCT7815 characterisation in bioreactor in 70 g/L xylose (A-B*), 20 g/L acetic acid (C-D), and 63 g/L glucose (E-F). A, C, and E: CO_2 production profiles (%), lipid production profiles (g/gDCW), carotenoid production profiles (g/L), and average specific growth rate (μ, h^{-1}). B, D and F: carbon source consumption (g/L), xylitol, D-arabinitol, citric acid and glycerol production profiles (g/L) and growth profile in biomass (gDCW/L). Dashed line marks the transition from exponential growth (exp) to nitrogen limitation (Nlim) phase. Red arrows point mark sampling point for the absolute proteome quantification. * data obtained from (Pinheiro *et al.*, 2020), lipids quantified in this study.

Growth on glucose showed the highest average specific growth rate (μ, 0.190 ± 0.025 h^{-1}), while on xylose and acetic acid similarly low growth rates (μ, 0.054 ± 0.001 h^{-1} and 0.073 ± 0.003 h^{-1}, respectively) were measured (**Supplementary Table S2**). During exponential

phase, the biomass yields on xylose and acetate were 0.19 gDCW/gsub and 0.19 gDCW/gsub (standard deviation (SD): 0.01 and 0.02, respectively), while in glucose 0.52 ± 0.03 gDCW/gsub (**Figure 5A, Supplementary Table S2**). At the end of exponential phase, the neglectable amount of glycerol (0.30 g/L) and a carbon balance of 100% (**Supplementary Table S2**) suggested that all the glucose consumed was directed toward biomass and CO2 production only. During nitrogen limitation phase, the biomass yield on glucose decreased to 0.24 ± 0.03 gDCW/gsub (**Figure 5A, Supplementary Table S2**). However, approximately the same ratio of consumed carbon was secreted as CO_2 as during the exponential growth phase, resulting in carbon imbalance (68%) (**Supplementary Table S2**). Although, high biomass yields are desirable, these observations suggest that results on glucose should be taken with caution. Ammonia was replaced with urea as the nitrogen source in glucose growth media to avoid the previously observed formation of cell aggregates, however, the cells still made aggregates disturbing the measurements. Those aggregates were not observed during cultivations with xylose and acetic acid. In comparison, with xylose as carbon source one third of carbon consumed was secreted as by-products (xylitol and D-arabinitol) and one third was secreted as CO_2 during the exponential phase. During nitrogen limitation phase, the decrease in by-products secretion was balanced by an equivalent increase of biomass and CO_2 production. compensated by equal increase in biomass yield and carbon secreted as CO_2. Nevertheless, based on carbon balance analysis, additional by-products in small amounts would be expected in xylose exponential phase. In acetic acid, carbon balance in exponential growth and nitrogen limitation phases was estimated at 85% and 67% (**Supplementary table S2**), respectively, indicating undetected by-products.

Although growth on glucose showed the highest lipid content in nitrogen limitation phase (0.48 ± 0.04 g/gDCW), the specific production rate was low (r_{LIP}, 0.013 ± 0.002 g/(gDCW*h) due to low growth rate μ (**Figure 5B, Supplementary Table S2**). In all conditions, lipid content increased during nitrogen limitation phase (**Figure 5B**). *R. toruloides* showed 11 times higher carotenoid content during nitrogen limitation phase in acetic acid condition (3.58 ± 0.25 mg/gDCW) than in glucose condition 0.3 ± 0.01 g/gDCW, **Figure 5C**, **Supplementary Table S2**). The specific production rate (r_{CAR}) was the highest in exponential growth phase (0.1 ± 0.01 mg/(gDCW*h) on acetate.

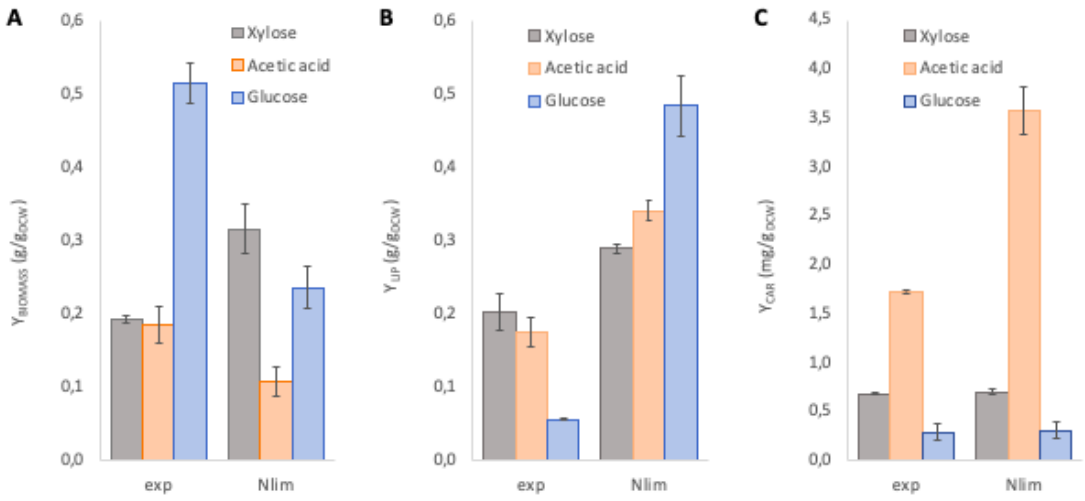

Figure 5. Biomass yield ($Y_{BIOMASS}$) on substrate (A), and lipid (Y_{LIP}) (B) and carotenoid (Y_{CAR}) (C) yields on biomass (gDCW) during exponential growth (exp) and nitrogen limitation (Nlim) phase of *R. toruloides* on different carbon substrates in under aerobic batch cultivation on a minimal mineral medium. Carotenoid and lipid yields were calculated at the late-exp growth phase and the end of Nlim phase. Protein yields were calculated at mid-exp and mid-Nlim phase. Biomass yields were calculated considering the production in each growth phase separately.

3.2.2 ecRhtoGEM: enzyme-constrained model of *R. toruloides*

To facilitate the application of *R. toruloides* for the production of sustainable chemicals and fuels, it is fundamental to improve the understanding of its metabolism. GEMs allow for the prediction intracellular fluxes at specified experimental conditions. In comparison to conventional GEMs, enzyme-constrained models (ecModels) take into account the metabolic limitations due to the enzyme capacity. An ecModel of *R. toruloides* has not been published before.

In this work, an enzyme-constrained GEM of *R. toruloides*, named ecRhtoGEM, was generated from a *R. toruloides* GEM published earlier by (Tiukova, Prigent, *et al.*, 2019), using the GECKO Toolbox (Sánchez *et al.*, 2017). GECKO Toolbox is designed to apply enzymatic constraints to the metabolic network via direct integration of absolute proteomics data. ecRhtoGEM is hosted on a dedicated Github repository (https://github.com/alinarekena/ecRhtoGEM/), where model generation process is tracked,

all scripts and files described in this work, and models in various file formats (mat, xml, yml) are provided, accessible upon request sent to alina.rekena@ut.ee).

Prior integration of enzymatic constraints, the model was improved by adding ribosome into the protein pseudo-reaction and adding an alternative D-xylose utilisation pathway via D-arabinitol to make the model better accommodate the collected metabolomics and proteomics data. Only ribosomal subunits with abundance over 1.e-05 mmol $gDCW^{-1}$ were included, as these are likely essential subunits, while lower abundances are alternative subunits **(Supplementary Figure S3;** ecRhtoGEM/results/ribosome_integration folder). As a result, 74 ribosomal subunits were integrated to the model.

Improvements to the xylose metabolism were motivated by the fact that the last step in xylose metabolism, D-xylulokinase (RHTO_04556, protein ID M7X6R2), has not been detected or has been detected in low abundance in proteomics and transcriptomics analysis in several studies (Pinheiro *et al.*, 2020; Kim *et al.*, 2021), including this work (data available in ecRhtoGEM/customGECKO/abs_proteomics.txt). The conventional xylose utilisation pathway starts with enzyme D-xylose reductase (XR) that reduces xylose to D-xylitol using NADPH as a cofactor (**Figure 6**). Next in the pathway is D-xylitol dehydrogenase (XDH) that converts D-xylitol to D-xylulose. An enzyme responsible for the last step in xylose metabolism is D-xylulokinase (XK) that converts D-xylulose to xylulose 5-phosphate at the cost of one ATP. Xylulose 5-phosphate is an intermediate of the pentose phosphate pathway (PPP). An alternative metabolite to connect xylose metabolism to PPP instead of xylulose 5-phosphate would be D-ribulose 5-phosphate, as recently suggested by (Kim *et al.*, 2021).

The construction of the alternative pathway started by adding D-arabinitol production pathway, as described in (Jagtap and Rao, 2018). From the two annotated *R. toruloides* NP11 D-arabinitol dehydrogenase enzymes (RHTO_07702 and RHTO_07844), only RHTO_07844 was detected by the proteomics analysis in this study. It was chosen to add RHTO_07844 as D-arabinitol 4-dehydrogenase converting D-xylulose to D-arabinitol at cost of one NADH to the metabolic network (**Figure 6**). Similarly it was done in *R. toruloides* GEM iRhto 1108 (Dinh *et al.*, 2019). The clue for adding D-arabinitol 2-dehydrogenase (DAD-2, or alternatively, D-ribulose reductase) to the model was found in the literature that described reversible L-xylulose reductase (LXR) from L-arabinose catabolism of yeast *Ambrosiozyma monospora* overexpressed in *S. cerevisiae* converting not only L-xylulose to xylitol, but also D-ribulose to D-arabinitol, having no activity with L-arabinitol (Verho *et al.*, 2004). In that study, LXR was described as NADH-dependent.

However, in rhtoGEM (Tiukova, Prigent, *et al.*, 2019) based on *R. toruloides* NP11, LXR (RHTO_00373) is NADP-dependent. Previous *in silico* analysis of the amino acid sequence of the LXR enzyme of *R. toruloides* NP11 by the NCBI conserved domain search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) showed that it could use both or either NADH and NADPH (Pinheiro *et al.*, 2020). In order to allow for the NADPH regeneration, a bidirectional NADP-dependent DAD-2 (RHTO_00373) to convert D-arabinitol to D-ribulose was added to the model (**Figure 6**). Finally, to search for D-ribulokinase that converts D-ribulose to D-ribulose 5-phosphate, *R. toruloides* IFO0880 protein 14368 from a GEM by (Kim *et al.*, 2021) was searched against the whole genome annotation of NP11 using the NCBI BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The search found 98.5% match with RHTO_00950, and it was the only hit. According to the Uniprot, the gene corresponds to protein of carboxydrate kinase (FGGY type family) with a protein ID M7WVT6. Consequently, an ATP-dependent D-ribulokinase (RHTO_00950) was added to the model (**Figure 6**). Analogous reactions of D-ribulokinase were present in GEMs of other non-conventional yeasts *Pichia pastoris* (Caspeta *et al.*, 2012) and *K. marxianus* (Marcišauskas, Ji and Nielsen, 2019). As a result, a new xylose metabolic pathway in line with the protein abundances measured in this study and the recent literature was constructed.

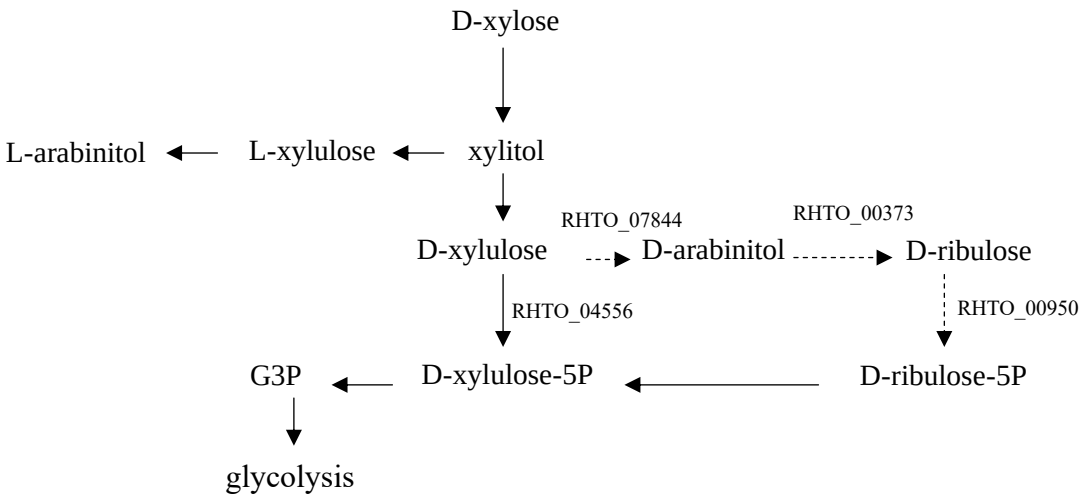

Figure 6. Xylose utilisation pathway (dashed arrows) integrating a new branch going from D-xylulose to D-ribulose-5P to the original pathway (solid arrows) present in *R. toruloides* NP11 (Zhu *et al.*, 2012). RHTO: gene associations of metabolic reactions, G3P: glyceraldehyde 3-phosphate.

The integration of enzymatic constraints required as an input all data from batch cultivation experiments. To test if applying enzymatic constraints allows the model to predict experimentally measured growth rate (i.e. is not growth-limiting), FBA was repeatedly performed on temporary model structures (Sánchez *et al.*, 2017). The integration of enzymatic constraints was initially an error-trial process, whereby the automatically retrieved k_{cat} values reported as growth-limiting were manually curated. Due to lack of data availability, many k_{cat} values used to constrain the model belong to other organisms than *R. toruloides*, even non-microbes (**Supplementary Tables S3, S4**). As a result of manual curation, k_{cat} values were increased for altogether 17 enzymes and their respective isoenzymes (detailed report at ecRhtoGEM/customGECKO/manualModifications.m). In final steps of the GECKO pipeline, protein abundances of growth-limiting enzymes were automatically curated again based on the FBA on temporary model structures (lists available at ecRhtoGEM/results/generate_protModels_pipeline folder *modifiedEnzymes* files). Also, the model was infeasible with the measured growth parameters on glucose. Therefore, the glucose uptake rate for the exponential phase was increased from 2.10 to 2.49 mmol/(gDCW h), leaving the growth rate as measured (0.19 h^{-1}) (ecRhtoGEM/customGECKO/fermentationData.txt). These modifications should be taken into account when interpreting modelling results.

As a result, six different versions of ecRhtoGEM were generated using the GECKO Toolbox to enable condition-specific systematic analysis of exponential growth and nitrogen limitation (where lipid accumulation takes place) phases on three different carbon sources (xylose, acetic acid and glucose). Biomass composition was adjusted to the measured lipid content and fatty acid profiles and total cellular protein content (**Supplementary Table S2**, ecRhtoGEM/data). Exchange fluxes were constrained with measured rates from cultivation experiments (**Supplementary Table S2** and (Pinheiro *et al.*, 2020)). Enzymatic constraints were applied to almost 3000 metabolic reactions, matching 693 different enzymes. 18-28% of the total proteome (mass-wise) were integrated as enzymatic constraints (**Supplementary Table S5**). ecRhtoGEM includes 5768 reactions, 3394 metabolites and 909 genes, corresponding to 111%, 49% and 7%, increase in comparison to rhto-GEM, respectively. High number of reactions in the ecRhtoGEM is associated with many isoenzymes and promiscuous enzymes in rhtoGEM metabolic network.

The power of the GEM in a combination with FBA is to calculate metabolic flux patterns. Model simulations suggested oxaloacetate as the major overflow metabolite in acetate condition (5% and 20% of the total acetate flux, ecRhtoGEM/results/model_simulation). Glucose models did not confirm any major overflow metabolites in glucose nitrogen limitation phase (GNlim), suggesting that the measurements for carbon balance analysis (68%, **Supplementary Table S2**) might have been disturbed by cell aggregates. Model prediction was not in agreement with the measured O_2 consumption and CO_2 production rates in glucose condition. Predicted fluxes were approximately 3-4 times higher compared with the measured ones in glucose exponential phase (GexpUrea).

Metabolic flux distribution illustrated the differences of carbon metabolism under the three carbon sources – xylose, glucose and acetic acid – in the focus of this work. The main difference is that acetate entered central metabolism at the level of acetyl-CoA and activated gluconeogenic reactions towards the glucose 6-phosphate (**Figure 7A, 7B**, orange line), while the main catabolic flux of glucose and xylose enters central carbon metabolism through glycolysis. Differences were noted in regard to metabolic reactions important for lipid synthesis. For example, 21% and 55% of consumed carbon on xylose and glucose during lipid accumulation was metabolised via redox-independent phosphoketolase (PK) - phosphate transacetylase pathway, and the flux was significantly increased during lipid accumulation (**Figure 7A, 7B**). Similar results on metabolic fluxes of PK in *R. toruloides* grown on xylose and glucose have also been described in the literature (Lopes *et al.*, 2020; Pinheiro *et al.*, 2020). PK converts D-xylulose 5-phosphate into acetyl-phosphate and further into acetyl-CoA and enables higher efficiency of carbon metabolism since bypasses the wasteful decarboxylation pyruvate to acetyl-CoA, which causes a loss of one third of the carbon substrate (Tiukova, Brandenburg, *et al.*, 2019). However, on acetic acid, metabolic fluxes via PK were predicted close to zero (t_0081, ecRhtoGEM/results/model_simulation) due to direct conversion of acetate into acetyl-CoA via acetyl-CoA synthetase (ACS) (**Figure 7A, 7B**). On xylose and glucose, fluxes via acetyl-CoA carboxylase (ACC) that converts acetyl-CoA into fatty acid precursor malonyl-CoA were increased on average 2- to 5-fold during lipid accumulation, corresponding to 23% and 55% of consumed carbon, respectively (**Figure 7A, 7B**). While in acetate condition, metabolic flux via ACC during lipid accumulation was downregulated (**Figure 7A, 7B**), which was not supported by previous omics and genome-scale modelling studies in *R. toruloides*, showing an upregulation of key enzymes of lipid production machinery during lipogenesis (Tiukova,

Brandenburg, *et al.*, 2019; Pinheiro *et al.*, 2020). Also, the experimentally measured lipid yield on biomass increased during nitrogen limitation phase (**Figure 5B**).

A unique enzyme of oleaginous yeasts is ATP citrate lyase (ACL) that represents an alternative reaction to produce cytosolic acetyl-CoA from citrate. From the simulations of ecRhtoGEM, 1% or less of consumed carbon was used by ACL (y200003, ecRhtoGEM/results/model_simulation), which is different from previous rhtoGEM predictions (Lopes *et al.*, 2020). This behaviour can be explained in case of acetate, where the key enzyme for acetyl-CoA production is ACS, but on xylose and glucose it demonstrates that the PK is preferred for the acetyl-CoA production. Using classical rhto-GEM, fluxes up to 32% of consumed carbon via ACL have been reported (Lopes *et al.*, 2020).

Ribosomes are essential for achieving faster cell growth and were reported to comprise the largest protein group in xylose-grown cells during exponential phase (almost one third of the total proteome) (Pinheiro *et al.*, 2020). The ecRhtoGEM predicted the highest flux via translation reaction on glucose during exponential growth phase (7.6% of consumed carbon) and during lipid accumulation (4.6% of consumed carbon flux) (translation, ecRhtoGEM/results/model_simulation). The lowest fluxes on translation and the decrease during nitrogen limitation was predicted on acetate (1.1% and 0.6% of consumed carbon flux, respectively). On the contrary, in xylose condition translation flux was higher during lipid accumulation (2.9% and 3.4% of consumed carbon during exp and Nlim, respectively) (translation, ecRhtoGEM/results/model_simulation). This is counterintuitive to the fact that ribosomal protein levels from exponential phase to nitrogen limitation phase decreased in xylose-grown cells (Pinheiro *et al.*, 2020). Downregulation ribosomal proteins as a response to nutrient starvation in *R. toruloides* grown on glucose and xylose has been reported before (Tiukova, Brandenburg, *et al.*, 2019) (Zhu *et al.*, 2012).

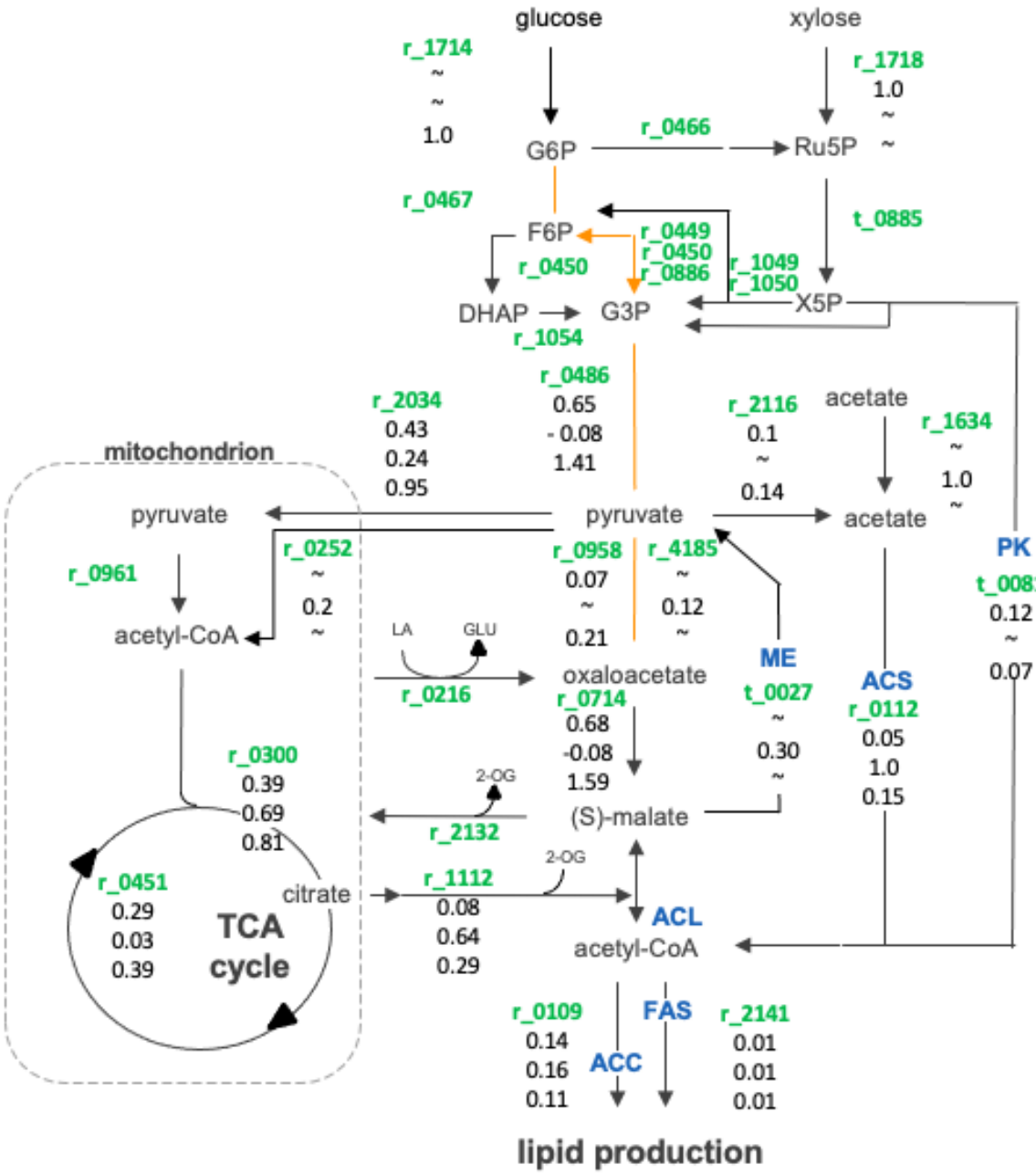

Figure 7A. Intracellular metabolic fluxes of the central carbon metabolism (normalised to the carbon source uptake flux) under exponential growth phase from *R. toruloides* cultivation on xylose (**upper value**), acetic acid (**middle value**), and glucose (**lower value**). Negative values represent fluxes of reverse glycolysis. All fluxes are available at github.com/alinarekena/ecRhtoGEM/results/model_simulation. G3P: glyceraldehyde 3-phosphate, F6P: D-fructose 6-phosphate, X5P: D-xylulose 5-phosphate, Ru5P: D-ribulose 5-phosphate, DHAP: dihydroxyacetone phosphate, LA: L-aspartate, 2-OG: 2-oxoglutarate (α-ketoglutarate), TCA: tricarboxylic cycle, PK: phosphoketolase, ACL: ATP-citrate lyase, FAS: fatty acid synthase, ACC: acetyl-CoA carboxylase.

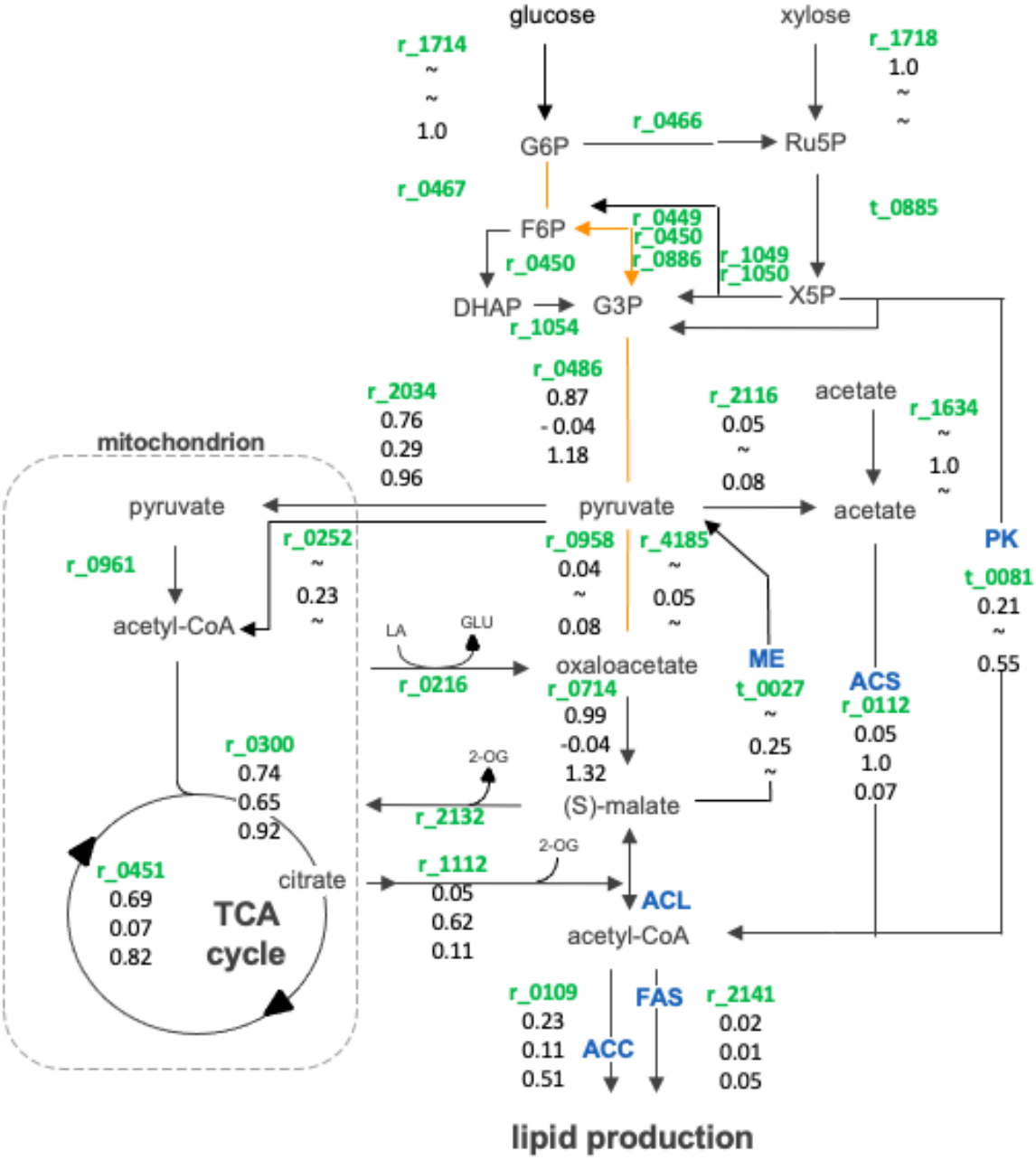

Figure 7B. Intracellular metabolic fluxes of the central carbon metabolism (normalised to the carbon source uptake flux) under nitrogen limitation phase from *R. toruloides* cultivation on xylose (**upper value**), acetic acid (**middle value**), and glucose (**lower value**). Negative values represent fluxes of reverse glycolysis. All fluxes are available at github.com/alinarekena/ecRhtoGEM/results/model_simulation. G3P: glyceraldehyde 3-phosphate, F6P: D-fructose 6-phosphate, X5P: D-xylulose 5-phosphate, Ru5P: D-ribulose 5-phosphate, DHAP: dihydroxyacetone phosphate, LA: L-aspartate, 2-OG: 2-oxoglutarate (α-ketoglutarate), GLU: L-glutamate, TCA: tricarboxylic cycle, PK: phosphoketolase, ACL: ATP-citrate lyase, FAS: fatty acid synthase, ACC: acetyl-CoA carboxylase.

3.2.3 Supply of NADPH during lipid accumulation

Lipid synthesis is NADPH-demanding process in oleaginous yeasts, but only few enzymes in the cytoplasm can generate NADPH. Malic enzyme and pentose phosphate pathway have been proposed as the main candidate enzymes for NADPH recycling in *R. toruloides* (Ratledge, 2014). ecRhtoGEM metabolic fluxes showed that on xylose and glucose NADPH is recycled via the oxidative part of pentose phosphate pathway, namely, glucose 6-phosphate dehydrogenase (G6PDH) and 6-phosphogluconate dehydrogenase (6PGD) (70% of consumed glucose and 42% of consumed xylose during Nlim phase, respectively (**Figure 8**)). In glucose condition PPP was predicted 21 % more active in Nlim phase than exp phase. This result can be explained that during nitrogen limitation the high PPP activity on glucose coupled with high fluxes via PK (carbon saving) and ACC (lipid synthesis) led to the highest measured lipid content in this study (Ypx 0.48 $\pm$ 0.04 g/gDCW, **Figure 5B**, **Supplementary Table S2**).

On xylose, NADPH demand is higher compared to the other substrates studied, as xylose utilisation (XR) is NADPH-dependent. Therefore, the model preferred to use the alternative xylose utilisation pathway via D-ribulokinase instead of D-xylulokinase (**Figure 9**, dashed line). During lipid accumulation, 80% of carbon was metabolised via the NADP-dependent D-arabinitol 2-dehydrogenase (DAD-2) that converts D-arabinitol into D-ribulose, generating one NADPH (**Figure 9**).

On acetic acid, the model predicted NADPH regeneration via the TCA-cycle related malic enzyme, corresponding to 25% of the consumed acetate during lipid accumulation (**Figure 8**). The flux on nitrogen limitation condition was 5% lower than during exponential growth phase.

An alternative option for NADPH generation is via NADP-dependent isocitrate dehydrogenase (IDH), occurring in *R. toruloides* both in cytosol and in mitochondria. From the ecRhtoGEM simulations, only 1-3% of consumed carbon was metabolised using any of the IDH enzymes during lipogenesis in xylose and glucose condition, respectively (**Figure 8**). On acetate, the fluxes via IDH enzymes were below 1% of the consumed carbon (**Figure 8**).

The NADPH balance analysis showed that in addition to fatty acid biosynthesis and xylose utilisation, a high flux involving NADPH-consumption was predicted via NADPH-

dependent glutamate dehydrogenase (GDH). The fluxes via GDH in xylose condition decreased from 18% to 10% of consumed carbon from exponential to nitrogen limitation phase, and from 68% to 18% of consumed carbon on glucose (r_0471, ecRhtoGEM/results/model_simulation). This is on the contrary with proteomics studies where the upregulation of enzymes involved in glutamate biosynthesis in response to nitrogen limitation have been shown (Zhu *et al.*, 2012; Tiukova, Brandenburg, *et al.*, 2019).

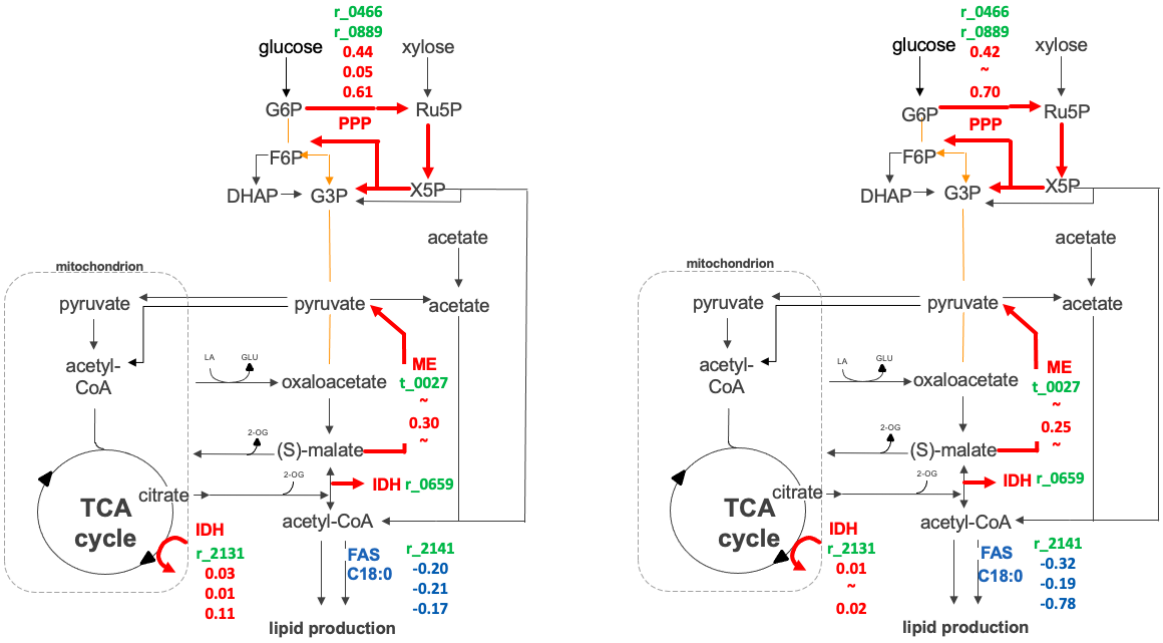

Figure 8. Fluxes of NADPH consumption and regeneration in *R. toruloides* (normalised to the stoichiometric coefficient and carbon uptake flux) under exponential growth (upper value) and nitrogen limitation (lower value) phases from cultivation on xylose (**A**), acetic acid (**B**), and glucose (**C**). Negative values represent fluxes of NADPH consumption, positive values – NADPH regeneration. The upper value represents exponential growth phase, the lower value - nitrogen limitation phase. * adjusted for C18:0 stoichiometry. G3P: glyceraldehyde 3-phosphate, DHAP: dihydroxyacetone phosphate, LA: L-aspartate, 2-OG: 2-oxoglutarate (α-ketoglutarate), TCA: tricarboxylic cycle, IDH: isocitrate dehydrogenase.

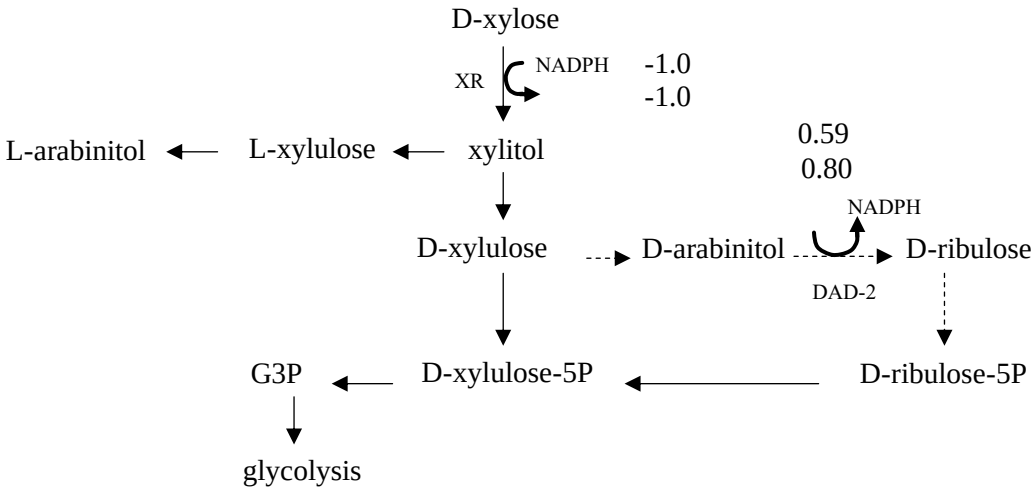

Figure 9. Fluxes of NADPH consumption and regeneration in the xylose utilisation pathway (normalised for carbon source uptake) in *R. toruloides* (normalised to the stoichiometric coefficient and carbon uptake flux) under exponential growth (upper value) and nitrogen limitation (lower value) phases from cultivation on xylose. Negative values represent fluxes of NADPH consumption, positive values – NADPH regeneration. The upper value represents exponential growth phase, the lower value - nitrogen limitation phase. G3P: glyceraldehyde 3-phospahte, XR: xylose reductase, DAD-2: D-arabinitol 2-dehydrogenase.

3.3 DISCUSSION

Enzyme-constrained genome-scale models can improve microbial strain phenotype predictions (Sánchez *et al.*, 2017). In this study, *R. toruloides* growth characterisation in bioreactors on xylose, glucose and acetic acid was carried out and an enzyme-constrained genome-scale model, named ecRhtoGEM, was developed. Except for the gases in glucose condition, the model could predict experimental rates measured in xylose, glucose and acetate condition. Notably, the quality of modelling predictions depends on the quality of experimental data provided as an input. To reduce the cell aggregate formation on glucose, adding 0.2 M NaCl or 0.2 M KCl to the cultivation media shown to reduce aggregates formation without impacting the growth could be suggested (Illarionov, Lahtvee and Kumar, 2021).

Understanding xylose metabolism in *R. toruloides* is important for the application in biorefinery concepts using lignocellulosic biomass, as its hemicellulosic fraction contains also C5 sugars. It has been known about the xylose metabolism that XR reduces xylose to xylitol, which is then oxidised to xylulose (Tiukova, Brandenburg, *et al.*, 2019). Subsequently, phosphorylated xylulose enters the PPP. However, the enzyme responsible for the phosphorylation of xylulose (XK) was not detected by the proteomic analysis in the present study and also previously (Pinheiro *et al.*, 2020). Therefore, it was proposed that xylulose could enter the PPP via another enzyme - ribulokinase. Similar approaches have been described earlier in *P. stipitis* (Jin, Cruz and Jeffries, 2005) and proposed in the most recent GEM of *R. toruloides* (Kim *et al.*, 2021). The ecRhtoGEM simulations in this study demonstrated, once again, that the NADPH plays crucial role in xylose metabolism.

Understanding acetic acid metabolism in *R. toruloides* is important for the biorefinery applications as well, as acetic acid is produced along with C5 and C6 sugars during the pre-treatment steps of the lignocellulosic biomass (Chandel, da Silva and Singh, 2013). Previous studies have suggested that cells growing on acetic acid have a more efficient metabolism and physiology, operating at near maximum efficient levels in terms of biomass production (Lopes *et al.*, 2020). However, ecRhtoGEM simulations demonstrated decreased metabolic fluxes on acetate during lipid accumulation, not typical to what has been mostly described in regard to lipid synthesis in xylose and glucose.

In this work, intracellular flux patterns were used to investigate the sources of NADPH on different carbon sources during lipid accumulation in *R. toruloides*. The main sources of NADPH in *R. toruloides* were predicted to be PPP and malic enzyme, as suggested by

previous theoretical calculations (Ratledge, 2014). The differences depending on the carbon source can be explained by the proximity of NADPH supplier enzymes to the main catabolic flux in the metabolic network. Acetic acid enters central carbon metabolism at the level of acetyl-CoA, assimilated by the key enzymes of glyoxylate cycle, isocitrate lyase and malate synthase, allowing the formation of malate, which serves as a substrate for malic enzyme. While fermentable nutrients, such as glucose and xylose, are assimilated via glycolytic pathways. Therefore, it could be more efficient for the cell to use the PPP in case of fermentable carbon sources and the ME in case of acetic acid and other nonfermentable carbon sources. The results from previous modelling studies with classical rhto-GEM are in the agreement with the PPP being the major NADPH supplier on xylose and glucose (Pinheiro *et al.*, 2020) (Lopes *et al.*, 2020) (Bommareddy *et al.*, 2015).

In the present study, the model predicted an increase in the metabolic flux of PPP in response to nitrogen limitation on glucose, but a decrease on xylose. These results are in agreement with a previous proteomic study on glucose in *R. toruloides*, where a contribution of alternative candidate enzymes (aldehyde dehydrogenases) that in addition to the PPP for the NADPH generation during lipogenesis was also proposed (Tiukova, Brandenburg, *et al.*, 2019). The flux via malic enzyme on acetate was predicted to decrease during nitrogen limitation. It has been shown from omics studies on glucose that the regulation of ME enzyme activity in *R. toruloides* is complicated (Zhu *et al.*, 2012).

Isocitrate dehydrogenase (IDH) has been mentioned as a potential candidate enzyme for NADPH production in oleaginous microorganisms in early biochemical studies (Wynn, Hamid and Ratledge, 1999). It was also predicted as the main source of NADPH on acetate with the classical rhto-GEM (Lopes *et al.*, 2020). However, ecRhtoGEM did not predict significant metabolic fluxes via IDH in any of studied conditions. In *S. cerevisiae*, elevated levels of IDH on nonfermentable carbon sources have been reported to prevent for the endogenous toxicity (Minard and McAlister-Henn, 2005).

The genome scale model developed in this work can be also used to study the fundamental characteristics of *R. toruloides* metabolism. As demonstrated in this study, the model can simulate the rate of translation using ribosomal abundances. It enables to search for the catalytic activities of the ribosome (k_{rib}) that might be governing cell growth laws. The correlation between ribosome abundances and specific growth rate has been studied before on other organisms (Scott *et al.*, 2014).

In conclusion, the model presented in this work can help to understand the metabolism of *R. toruloides*. It can be used to design metabolic engineering strategies, such as improving the utilisation of isocitrate dehydrogenase (increasing NADPH supply), as well as to study the biochemical mechanisms as response to various environmental stresses, such as nutrient starvation. *R. toruloides* is an attractive host for sustainable chemicals and fuels production that has gained a great academic and industrial interest. The genome-scale model developed in this work will enhance the understanding of biology of this yeast.

SUMMARY

Previous genome-scale models (GEMs) of *R. toruloides* have demonstrated the ability to predict measured experimental rates and potential targets for the metabolic engineering. An enzyme-constrained GEM of *R. toruloides*, ecRhtoGEM, developed in this study demonstrated non-seen differences in predicted intracellular metabolic flux patterns among different carbon substrates in metabolic pathways important for lipid synthesis. On acetic acid as a sole carbon source, the main supplier enzyme of the cofactor NADPH was predicted to be cytosolic malic enzyme (ME). In comparison, on fermentable carbon substrates (xylose and glucose) the oxidative part of pentose phosphate pathway (PPP) was predicted to supply the majority of NADPH. These results are open to experimental validation and can be instructive for the design of further metabolic engineering strategies to improve lipid and specialty chemicals production in *R. toruloides*.

ACKNOWLEDGEMENTS

Marina Julio Pinheiro for strain characterisation experiments and analysis on acetic acid, Laura Kibena for HPLC analysis, Proteomics Core Facility at University of Tartu for proteome quantification, Eliise Tammekivi and Koit Herodes from Institute of Chemistry University of Tartu for the fatty acid analysis, and Eduard J. Kerkhoven for help, advice and assistance with using the GECKO Toolbox.

This work has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 668997, and the Estonian Research Council grants PUT1488P and PRG1101.

REFERENCES

Ageitos, J. M. *et al.* (2011) 'Oily yeasts as oleaginous cell factories', *Applied Microbiology and Biotechnology*, 90(4), pp. 1219–1227. doi: 10.1007/s00253-011-3200-z.

Angerbauer, C. *et al.* (2008) 'Conversion of sewage sludge into lipids by Lipomyces starkeyi for biodiesel production', *Bioresource Technology*, 99(8), pp. 3051–3056. doi: 10.1016/j.biortech.2007.06.045.

Bajpai, P. (2021) 'Global Production of Bioethanol', in *Developments in Bioethanol. Green Energy and Technology*. Singapore: Springer, pp. 177–196. doi: 10.1007/978-981-15-8779-5_10.

Bandhu, S. *et al.* (2020) 'Yeast Single Cell Oils from Bioresources: Current Developments in Production and Applications', *Current Sustainable/Renewable Energy Reports*, 7, pp. 109–120. doi: 10.1007/s40518-020-00160-6.

Beg, Q. K. *et al.* (2007) 'Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity', *Proceedings of the National Academy of Sciences*, 104(31), pp. 12663–12668. doi: 10.1073/pnas.0609845104.

van Bodegom, P. (2007) 'Microbial Maintenance: A Critical Review on Its Quantification', *Microbial Ecology*, 53, pp. 513–523. doi: 10.1007/s00248-006-9049-5.

Bommareddy, R. R. *et al.* (2015) 'Metabolic network analysis and experimental study of lipid production in Rhodosporidium toruloides grown on single and mixed substrates', *Microbial Cell Factories*, 14(1), p. 36. doi: 10.1186/s12934-015-0217-5.

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

Bordel, S., Agren, R. and Nielsen, J. (2010) 'Sampling the Solution Space in Genome-Scale Metabolic Networks Reveals Transcriptional Regulation in Key Enzymes', *PLoS Computational Biology*, 6, p. e1000859. doi: 10.1371/journal.pcbi.1000859.

Burgard, A. P. and Maranas, C. D. (2003) 'Optimization-based framework for inferring and testing hypothesized metabolic objective functions', *Biotechnology and Bioengineering*, 82(6), pp. 670–677. doi: 10.1002/bit.10617.

Caspeta, L. *et al.* (2012) 'Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in-silico evaluation of their potentials', *BMC Systems Biology*, 6(1), p.

24. doi: 10.1186/1752-0509-6-24.

Chandel, A. K., da Silva, S. S. and Singh, O. V. (2013) 'Detoxification of Lignocellulose Hydrolysates: Biochemical and Metabolic Engineering Toward White Biotechnology', *BioEnergy Research*, 6, pp. 388–401. doi: 10.1007/s12155-012-9241-z.

Coradetti, S. T. *et al.* (2018) 'Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides', *eLife*, 7, p. e32110. doi: 10.7554/eLife.32110.

Dias, C. *et al.* (2015) 'New dual-stage pH control fed-batch cultivation strategy for the improvement of lipids and carotenoids production by the red yeast Rhodosporidium toruloides NCYC 921', *Bioresource Technology*, 189, pp. 309-318. doi: 10.1016/j.biortech.2015.04.009.

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

Domenzain, I. *et al.* (2021) 'Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0', *preprint on bioRxiv*. doi: 10.1101/2021.03.05.433259.

Edwards, J. S., Covert, M. and Palsson, B. (2002) 'Metabolic modelling of microbes: the flux-balance approach', *Environmental Microbiology*, 4(3), pp. 133–140. doi: 10.1046/j.1462-2920.2002.00282.x.

Evans, C. T. and Ratledge, C. (1985) 'Possible regulatory roles of ATP:citrate lyase, malic enzyme, and AMP deaminase in lipid accumulation by Rhodosporidium toruloides CBS 14', *Canadian Journal of Microbiology*, 31(11), pp. 1000–1005. doi: 10.1139/m85-189.

Fei, Q. *et al.* (2016) 'Enhanced lipid production by Rhodosporidium toruloides using different fed-batch feeding strategies with lignocellulosic hydrolysate as the sole carbon source', *Biotechnology for Biofuels*, 9, p. 130. doi: 10.1186/s13068-016-0542-x.

Fernando, S. *et al.* (2006) 'Biorefineries: Current Status, Challenges, and Future Direction', *Energy & Fuels*, 20(4), pp. 1727–1737. doi: 10.1021/ef060097w.

Goldemberg, J. (2007) 'Ethanol for a Sustainable Energy Future', *Science*, 315(5813), pp. 808–810. doi: 10.1126/science.1137013.

Heirendt, L. *et al.* (2019) 'Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0', *Nature Protocols*, 14, pp. 639–702. doi: 10.1038/s41596-

018-0098-2.

Hiltunen, J. K. *et al.* (2003) 'The biochemistry of peroxisomal β-oxidation in the yeast Saccharomyces cerevisiae', *FEMS Microbiology Reviews*, 27, pp. 35–64. doi: 10.1016/S0168-6445(03)00017-2.

van Hoek, M. J. and Merks, R. M. (2012) 'Redox balance is key to explaining full vs. partial switching to low-yield metabolism', *BMC Systems Biology*, 6, p. 22. doi: 10.1186/1752-0509-6-22.

Holladay, J. E. *et al.* (2007) *Top Value-Added Chemicals from Biomass - Volume II—Results of Screening for Potential Candidates from Biorefinery Lignin*. Richland, WA (United States). doi: 10.2172/921839.

Holtzapple, M. T. (1993) *Encyclopedia of Food Science, Food Technology, and Nutrition*. Edited by R. Macrae, R. K. Robinson, and M. J. Sadler. London: Academic Press.

Hu, J. and Ji, L. (2016) 'Draft Genome Sequences of Rhodosporidium toruloides Strains ATCC 10788 and ATCC 10657 with Compatible Mating Types', *Genome Announcements*, 4(2), pp. e00098-16. doi: 10.1128/genomeA.00098-16.

Huang, X.-F. *et al.* (2016) 'Culture strategies for lipid production using acetic acid as sole carbon source by Rhodosporidium toruloides', *Bioresource Technology*, 206, pp. 141–149. doi: 10.1016/j.biortech.2016.01.073.

Illarionov, A., Lahtvee, P.-J. and Kumar, R. (2021) 'Characterization of potassium and sodium salt stress in yeasts', *Applied and Environmental Microbiology*, p. (accepted manuscript). doi: 10.1128/AEM.03100-20.

Jagtap, S. S. and Rao, C. V. (2018) 'Production of d-arabitol from d-xylose by the oleaginous yeast Rhodosporidium toruloides IFO0880', *Applied Microbiology and Biotechnology*, 102, pp. 143–151. doi: 10.1007/s00253-017-8581-1.

Jem, K. J. and Tan, B. (2020) 'The development and challenges of poly (lactic acid) and poly (glycolic acid)', *Advanced Industrial and Engineering Polymer Research*, 3(2), pp. 60–70. doi: 10.1016/j.aiepr.2020.01.002.

Jin, M. *et al.* (2015) 'Microbial lipid-based lignocellulosic biorefinery: feasibility and challenges', *Trends in Biotechnology*, 33(1), pp. 43–54. doi: 10.1016/j.tibtech.2014.11.005.

Jin, Y.-S., Cruz, J. and Jeffries, T. W. (2005) 'Xylitol production by a Pichia stipitis D-xylulokinase mutant', *Applied Microbiology and Biotechnology*, 68, pp. 42–45. doi:

10.1007/s00253-004-1854-5.

Kerkhoven, E. J., Lahtvee, P.-J. and Nielsen, J. (2015) 'Applications of computational modeling in metabolic engineering of yeast', *FEMS Yeast Research*, 15(1), pp. 1–13. doi: 10.1111/1567-1364.12199.

Kim, J. *et al.* (2021) 'Multi-Omics Driven Metabolic Network Reconstruction and Analysis of Lignocellulosic Carbon Utilization in Rhodosporidium toruloides', *Frontiers in Bioengineering and Biotechnology*, 8, p. 612832. doi: 10.3389/fbioe.2020.612832.

Koutinas, A. A. *et al.* (2014) 'Valorization of industrial waste and by-product streams via fermentation for the production of chemicals and biopolymers', *Chemical Society Reviews*, 43(8), pp. 2587–2627. doi: 10.1039/c3cs60293a.

Kumar, R. and Lahtvee, P.-J. (2020) 'Proteome overabundance enables respiration but limitation onsets carbon overflow', *preprint on bioRxiv*. doi: 10.1101/2020.02.20.957662.

Kumar, S. *et al.* (2012) 'Genome Sequence of the Oleaginous Red Yeast Rhodosporidium toruloides MTCC 457', *Eukaryotic Cell*, 11(8), pp. 1083–1084. doi: 10.1128/EC.00156-12.

Lahtvee, P.-J. *et al.* (2017) 'Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast', *Cell Systems*, 4(5), pp. 495-504.e5. doi: 10.1016/j.cels.2017.03.003.

Lee, J.-Y., Lee, S.-E. and Lee, D.-W. (2021) 'Current status and future prospects of biological routes to bio-based products using raw materials, wastes, and residues as renewable resources', *Critical Reviews in Environmental Science and Technology*. doi: 10.1080/10643389.2021.1880259.

Lee, J. J. L. *et al.* (2014) 'Metabolomic Profiling of Rhodosporidium toruloides Grown on Glycerol for Carotenoid Production during Different Growth Phases', *Journal of Agricultural and Food Chemistry*, 62(41), pp. 10203–10209. doi: 10.1021/jf502987q.

Li, Y. *et al.* (2006) 'Optimization of Culture Conditions for Lipid Production by Rhodosporidium toruloides', *Chinese Journal of Biotechnology*, 22(4), pp. 650–656. doi: 10.1016/S1872-2075(06)60050-2.

Li, Y., Zhao, Z. (Kent) and Bai, F. (2007) 'High-density cultivation of oleaginous yeast Rhodosporidium toruloides Y4 in fed-batch culture', *Enzyme and Microbial Technology*, 41(3), pp. 312–317. doi: 10.1016/j.enzmictec.2007.02.008.

Lian, J. and Zhao, H. (2015) 'Recent advances in biosynthesis of fatty acids derived products

in Saccharomyces cerevisiae via enhanced supply of precursor metabolites', *Journal of Industrial Microbiology and Biotechnology*, 42(3), pp. 437–451. doi: 10.1007/s10295-014-1518-0.

Liao, J. C. (1993) 'Modelling and analysis of metabolic pathways', *Current Opinion in Biotechnology*, 4(2), pp. 211–216. doi: 10.1016/0958-1669(93)90127-I.

Liu, H. *et al.* (2009) 'Comparative proteomic analysis of Rhodosporidium toruloides during lipid accumulation', *Yeast*, 26(10), pp. 553–566. doi: 10.1002/yea.1706.

Lloyd, C. J. *et al.* (2018) 'COBRAme: A computational framework for genome-scale models of metabolism and gene expression', *PLOS Computational Biology*, 14(7), p. e1006302. doi: 10.1371/journal.pcbi.1006302.

Lopes, H. J. S. *et al.* (2020) 'C/N ratio and carbon source-dependent lipid production profiling in Rhodotorula toruloides', *Applied Microbiology and Biotechnology*, 104(6), pp. 2639–2649. doi: 10.1007/s00253-020-10386-5.

Lopes, H. J. S., Bonturi, N. and Miranda, E. A. (2020) 'Rhodotorula toruloides Single Cell Oil Production Using Eucalyptus urograndis Hemicellulose Hydrolysate as a Carbon Source', *Energies*, 13(4), p. 795. doi: 10.3390/en13040795.

Marcišauskas, S., Ji, B. and Nielsen, J. (2019) 'Reconstruction and analysis of a Kluyveromyces marxianus genome-scale metabolic model', *BMC Bioinformatics*, 20, p. 551. doi: 10.1186/s12859-019-3134-5.

Minard, K. I. and McAlister-Henn, L. (2005) 'Sources of NADPH in Yeast Vary with Carbon Source', *Journal of Biological Chemistry*, 280(48), pp. 39890–39896. doi: 10.1074/jbc.M509461200.

Monteiro de Oliveira, P. *et al.* (2021) 'Screening and Growth Characterization of Non-conventional Yeasts in a Hemicellulosic Hydrolysate', *Frontiers in Bioengineering and Biotechnology*, 9, p. 659472. doi: 10.3389/fbioe.2021.659472.

Morin, N. *et al.* (2014) 'Draft Genome Sequence of Rhodosporidium toruloides CECT1137, an Oleaginous Yeast of Biotechnological Interest', *Genome Announcements*, 2(4), pp. e00641-14. doi: 10.1128/genomeA.00641-14.

Niebel, B., Leupold, S. and Heinemann, M. (2019) 'An upper limit on Gibbs energy dissipation governs cellular metabolism', *Nature Metabolism*, 1, pp. 125–132. doi: 10.1038/s42255-018-0006-7.

Nilsson, A. and Nielsen, J. (2016) 'Metabolic Trade-offs in Yeast are Caused by F1F0-ATP synthase', *Scientific Reports*, 6, p. 22264. doi: 10.1038/srep22264.

Nordberg, H. *et al.* (2014) 'The genome portal of the Department of Energy Joint Genome Institute: 2014 updates', *Nucleic Acids Research*, 42(D1), pp. D26–D31. doi: 10.1093/nar/gkt1069.

O'Brien, E. J., Monk, J. M. and Palsson, B. O. (2015) 'Using Genome-scale Models to Predict Biological Capabilities', *Cell*, 161(5), pp. 971–987. doi: 10.1016/j.cell.2015.05.019.

Oftadeh, O. *et al.* (2021) 'A genome-scale metabolic model of Saccharomyces cerevisiae that integrates expression constraints and reaction thermodynamics', *preprint on bioRxiv*. doi: 10.1101/2021.02.17.431671.

Orth, J. D., Thiele, I. and Palsson, B. Ø. (2010) 'What is flux balance analysis?', *Nature Biotechnology*, 28, pp. 245–248. doi: 10.1038/nbt.1614.

Owusu, P. A. and Asumadu-Sarkodie, S. (2016) 'A review of renewable energy sources, sustainability issues and climate change mitigation', *Cogent Engineering*, 3(1), p. 1167990. doi: 10.1080/23311916.2016.1167990.

Palmqvist, E. and Hahn-Hägerdal, B. (2000) 'Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition', *Bioresource Technology*, 74(1), pp. 25–33. doi: 10.1016/S0960-8524(99)00161-3.

Panwar, N. L., Kaushik, S. C. and Kothari, S. (2011) 'Role of renewable energy sources in environmental protection: A review', *Renewable and Sustainable Energy Reviews*, 15(3), pp. 1513–1524. doi: 10.1016/j.rser.2010.11.037.

Papanikolaou, S. and Aggelis, G. (2010) 'Yarrowia lipolytica: A model microorganism used for the production of tailor-made lipids', *European Journal of Lipid Science and Technology*, 112(6), pp. 639–654. doi: 10.1002/ejlt.200900197.

Papanikolaou, S. and Aggelis, G. (2011a) 'Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production', *European Journal of Lipid Science and Technology*, 113(8), pp. 1031–1051. doi: 10.1002/ejlt.201100014.

Papanikolaou, S. and Aggelis, G. (2011b) 'Lipids of oleaginous yeasts. Part II: Technology and potential applications', *European Journal of Lipid Science and Technology*, 113(8), pp. 1052–1073. doi: 10.1002/ejlt.201100015.

Park, Y.-K., Nicaud, J.-M. and Ledesma-Amaro, R. (2018) 'The Engineering Potential of

Rhodosporidium toruloides as a Workhorse for Biotechnological Applications', *Trends in Biotechnology*, 36(3), pp. 304–317. doi: 10.1016/j.tibtech.2017.10.013.

Pinheiro, M. J. *et al.* (2020) 'Xylose Metabolism and the Effect of Oxidative Stress on Lipid and Carotenoid Production in Rhodotorula toruloides: Insights for Future Biorefinery', *Frontiers in Bioengineering and Biotechnology*, 8, p. 1008. doi: 10.3389/fbioe.2020.01008.

Ratledge, C. (2014) 'The role of malic enzyme as the provider of NADPH in oleaginous microorganisms: a reappraisal and unsolved problems', *Biotechnology Letters*, 36(8), pp. 1557–1568. doi: 10.1007/s10529-014-1532-3.

Ratledge, C. and Wynn, J. P. (2002) 'The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms', *Advances in Applied Microbiology*, 51, pp. 1–51. doi: 10.1016/S0065-2164(02)51000-5.

Royce, L. A. *et al.* (2013) 'The damaging effects of short chain fatty acids on Escherichia coli membranes', *Applied Microbiology and Biotechnology*, 97(18), pp. 8317–8327. doi: 10.1007/s00253-013-5113-5.

Salvy, P. and Hatzimanikatis, V. (2020) 'The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models', *Nature Communications*, 11, p. 30. doi: 10.1038/s41467-019-13818-7.

Sambles, C. *et al.* (2017) 'Genome sequence of the oleaginous yeast Rhodotorula toruloides strain CGMCC 2.1609', *Genomics Data*, 13, pp. 1–2. doi: 10.1016/j.gdata.2017.05.009.

Sampaio, J. P. (2011) 'Rhodosporidium Banno (1967)', in *The Yeasts*. 5th edn. Elsevier, pp. 1523–1539. doi: 10.1016/B978-0-444-52149-1.00127-0.

Sánchez, B. J. *et al.* (2017) 'Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints', *Molecular Systems Biology*, 13(8), p. 935. doi: 10.15252/msb.20167411.

Sánchez, B. J. *et al.* (2021) 'Benchmarking accuracy and precision of intensity-based absolute quantification of protein abundances in Saccharomyces cerevisiae', *Proteomics*, 21(6), p. 2000093. doi: 10.1002/pmic.202000093.

Sànchez Nogué, V. and Karhumaa, K. (2015) 'Xylose fermentation as a challenge for commercialization of lignocellulosic fuels and chemicals', *Biotechnology Letters*, 37(4), pp. 761–772. doi: 10.1007/s10529-014-1756-2.

Schomburg, I. *et al.* (2012) 'BRENDA in 2013: integrated reactions, kinetic data, enzyme

function data, improved disease classification: new options and contents in BRENDA', *Nucleic Acids Research*, 41(D1), pp. D764–D772. doi: 10.1093/nar/gks1049.

Schuetz, R., Kuepfer, L. and Sauer, U. (2007) 'Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli', *Molecular Systems Biology*, 3(1), p. 119. doi: 10.1038/msb4100162.

Schwanhäusser, B. *et al.* (2011) 'Global quantification of mammalian gene expression control', *Nature*, 473(7347), pp. 337–342. doi: 10.1038/nature10098.

Scott, M. *et al.* (2014) 'Emergence of robust growth laws from optimal regulation of ribosome synthesis', *Molecular Systems Biology*, 10, p. 747. doi: 10.15252/msb.20145379.

Shi, J. *et al.* (2013) 'Comparative Proteomics Profile of Lipid-Cumulating Oleaginous Yeast: An iTRAQ-Coupled 2-D LC-MS/MS Analysis', *PLoS ONE*, 8(12), p. e85532. doi: 10.1371/journal.pone.0085532.

Sun, W. *et al.* (2017) 'Homologous gene targeting of a carotenoids biosynthetic gene in Rhodosporidium toruloides by Agrobacterium-mediated transformation', *Biotechnology Letters*, 39, pp. 1001–1007. doi: 10.1007/s10529-017-2324-3.

Takkellapati, S., Li, T. and Gonzalez, M. A. (2018) 'An overview of biorefinery-derived platform chemicals from a cellulose and hemicellulose biorefinery', *Clean Technologies and Environmental Policy*, 20, pp. 1615–1630. doi: 10.1007/s10098-018-1568-5.

Tammekivi, E. *et al.* (2019) 'Comparison of derivatization methods for the quantitative gas chromatographic analysis of oils', *Analytical Methods*, 11(28), pp. 3514–3522. doi: 10.1039/C9AY00954J.

Tammekivi, E. *et al.* (2021) 'Quantitative GC–MS Analysis of Artificially Aged Paints with Variable Pigment and Linseed Oil Ratios', *Molecules*, 26(8), p. 2218. doi: 10.3390/molecules26082218.

Thevenieau, F. and Nicaud, J.-M. (2013) 'Microorganisms as sources of oils', *OCL*, 20(6), p. D603. doi: 10.1051/ocl/2013034.

Tiukova, I. A., Prigent, S., *et al.* (2019) 'Genome-scale model of *Rhodotorula toruloides* metabolism', *Biotechnology and Bioengineering*, 116(12), pp. 3396–3408. doi: 10.1002/bit.27162.

Tiukova, I. A., Brandenburg, J., *et al.* (2019) 'Proteome analysis of xylose metabolism in Rhodotorula toruloides during lipid production', *Biotechnology for Biofuels*, 12(1), p. 137.

doi: 10.1186/s13068-019-1478-8.

Tran, T. N. *et al.* (2019) 'Draft genome sequence data of Rhodosporidium toruloides VN1, a strain capable of producing natural astaxanthin', *Data in Brief*, 26, p. 104443. doi: 10.1016/j.dib.2019.104443.

Tyanova, S., Temu, T. and Cox, J. (2016) 'The MaxQuant computational platform for mass spectrometry-based shotgun proteomics', *Nature Protocols*, 11(12), pp. 2301–2319. doi: 10.1038/nprot.2016.136.

Varma, A. and Palsson, B. O. (1994) 'Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use', *Nature Biotechnology*, 12(10), pp. 994–998. doi: 10.1038/nbt1094-994.

Verho, R. *et al.* (2004) 'A Novel NADH-linked L-Xylulose Reductase in the L-Arabinose Catabolic Pathway of Yeast', *Journal of Biological Chemistry*, 279(15), pp. 14746–14751. doi: 10.1074/jbc.M312533200.

VTT Technical Research Centre of Finland (2021) *Bioplastic from soy residues: Pilot plant coming to Finland*. Available at: https://www.vttresearch.com/en/news-and-ideas/bioplastic-soy-residues-pilot-plant-coming-finland (Accessed: 15 May 2021).

Wang, H. *et al.* (2018) 'RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor', *PLOS Computational Biology*, 14(10), p. e1006541. doi: 10.1371/journal.pcbi.1006541.

Wen, Z. *et al.* (2020) 'Rhodosporidium toruloides - A potential red yeast chassis for lipids and beyond', *FEMS Yeast Research*. Oxford University Press (OUP), 20(5), p. foaa038. doi: 10.1093/femsyr/foaa038.

Wiebe, M. G. *et al.* (2012) 'Lipid production in batch and fed-batch cultures of Rhodosporidium toruloidesfrom 5 and 6 carbon carbohydrates', *BMC Biotechnology*, 12, p. 26. doi: 10.1186/1472-6750-12-26.

Wu, S. *et al.* (2010) 'Phosphate-limitation mediated lipid production by Rhodosporidium toruloides', *Bioresource Technology*, 101(15), pp. 6124–6129. doi: 10.1016/j.biortech.2010.02.111.

Wu, S. *et al.* (2011) 'Microbial lipid production by Rhodosporidium toruloides under sulfate-limited conditions', *Bioresource Technology*, 102(2), pp. 1803–1807. doi: 10.1016/j.biortech.2010.09.033.

Wynn, J. P., Hamid, A. bin A. and Ratledge, C. (1999) 'The role of malic enzyme in the

regulation of lipid accumulation in filamentous fungi', *Microbiology*, 145(8), pp. 1911–1917. doi: 10.1099/13500872-145-8-1911.

Xu, J. *et al.* (2012) 'Microbial conversion of biodiesel byproduct glycerol to triacylglycerols by oleaginous yeast Rhodosporidium toruloides and the individual effect of some impurities on lipid production', *Biochemical Engineering Journal*, 65, pp. 30–36. doi: 10.1016/j.bej.2012.04.003.

Yang, X. *et al.* (2018) 'Expression of phosphotransacetylase in Rhodosporidium toruloides leading to improved cell growth and lipid production', *RSC Advances*, 8(43), pp. 24673–24678. doi: 10.1039/C8RA03028F.

Zhang, C. and Hua, Q. (2016) 'Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine', *Frontiers in Physiology*, 6, p. 413. doi: 10.3389/fphys.2015.00413.

Zhang, S., Skerker, J. M., *et al.* (2016) 'Engineering Rhodosporidium toruloides for increased lipid production', *Biotechnology and Bioengineering*, 113(5), pp. 1056–1066. doi: 10.1002/bit.25864.

Zhang, S., Ito, M., *et al.* (2016) 'Metabolic engineering of the oleaginous yeast Rhodosporidium toruloides IFO0880 for lipid overproduction during high-density fermentation', *Applied Microbiology and Biotechnology*, 100, pp. 9393–9405. doi: 10.1007/s00253-016-7815-y.

Zhu, Z. *et al.* (2012) 'A multi-omic map of the lipid-producing yeast Rhodosporidium toruloides', *Nature Communications*, 3, p. 1112. doi: 10.1038/ncomms2112.

SUPPLEMENTARY

Table S1. Growth- (GAM) and nongrowth-associated (NGAM) energy costs fitted to measured carbon uptake rates from exponential growth and nitrogen limitation phases of *R. toruloides* growth under conditions studied in this work.

model version	GAM (mmol/gDCW)	NGAM (mmol/(gDCW*h))
ecRhtoGEM_Xexp	133.217	3.65
ecRhtoGEM_XNlim	126.423	3.15
ecRhtoGEM_Aexp	132.256	3.65
ecRhtoGEM_ANlim	125.923	3.45
ecRhtoGEM_GexpUrea	139.991	0
ecRhtoGEM_GNimUrea	124.354	3.3

Table S2. Physiological parameters of *R. toruloides* growth in batch cultivation for exponential growth (exp) and nitrogen limitation (Nlim) phase in xylose (X), acetic acid (A) and glucose (G). μ: specific growth rate, Y: yield, r: specific rate of productivity, rCO_2: rate of CO_2 production, rO_2: rate of O_2 consumption, C balance: carbon balance, SD: standard deviation of two replicates.

Condition	μ (h-1)	SD	rcarbon (mmol/ gDCW*h)	SD	rCO_2 (mmol/ gDCW*h)	SD	rO_2 (mmol/ gDCW*h)	SD
Xexp	0.054	0.001	-1.859	0.022	2.845	0.325	-2.678	0.012
XNlim	0.021	0.001	-0.434	0.026	1.277	0.139	-1.028	0.012
Aexp	0.073	0.003	-6.627	0.557	5.224	0.724	-7.144	-
ANlim	0.012	0.000	-1.971	0.429	2.229	0.198	-2.265	-
GexpUrea	0.191	0.025	-2.096	0.557	2.685	0.584	-1.423	0.779
GNlimUrea	0.021	0.007	-0.410	0.007	1.547	0.519	-0.704	0.338

Condition	rxylitol (mmol/ gDCW*h)	SD	rarabinitol (mmol/ gDCW*h)	SD	rcitrate (mmol/ gDCW*h)	SD	rglycerol (mmol/ gDCW*h)	SD
Xexp	0.223	0.026	0.367	0.0433	-		-	
XNlim	0.004	0.004	0.077	0.0118	-		-	
Aexp	-		-		0.122	0.016	-	
ANlim	-		-		-0.043	0.001	-	
GexpUrea	-		-		-		0.049	0.016
GNlimUrea	-		-		-		-0.007	0.000

Condition	C balance	SD	rcarotenoid (mg/gDCW*h)	SD	rlipid (g/gDCW*h)	SD
Xexp	0.893	0.056	0.037	0.001	0.011	0.002
XNlim	1.086	0.183	0.014	0.001	0.006	0.000
Aexp	0.849	0.046	0.126	0.010	0.013	0.001
ANlim	0.665	0.029	0.044	0.001	0.004	0.000
GexpUrea	1.061	0.155	0.051	0.009	0.011	0.001
GNlimUrea	0.685	0.056	0.005	0.001	0.013	0.002

Condition	Yxs (gDCW/ gsub)	SD	Ypx (mgcarot/ gDCW)	SD	Ypx (glipid/ gDCW)	SD	Ypx (gprotein/ gDCW)	SD
Xexp	0.193	0.006	0.681	0.006	0.202	0.025	0.438	0.013
XNlim	0.316	0.034	0.700	0.027	0.290	0.006	0.197	0.002
Aexp	0.185	0.024	1.723	0.025	0.175	0.019	0.386	0.012
ANlim	0.109	0.020	3.575	0.250	0.341	0.013	0.217	0.012
GexpUrea	0.515	0.027	0.290	0.086	0.057	0.002	0.636	0.024
GNlimUrea	0.236	0.029	0.310	0.088	0.483	0.041	0.227	0.031

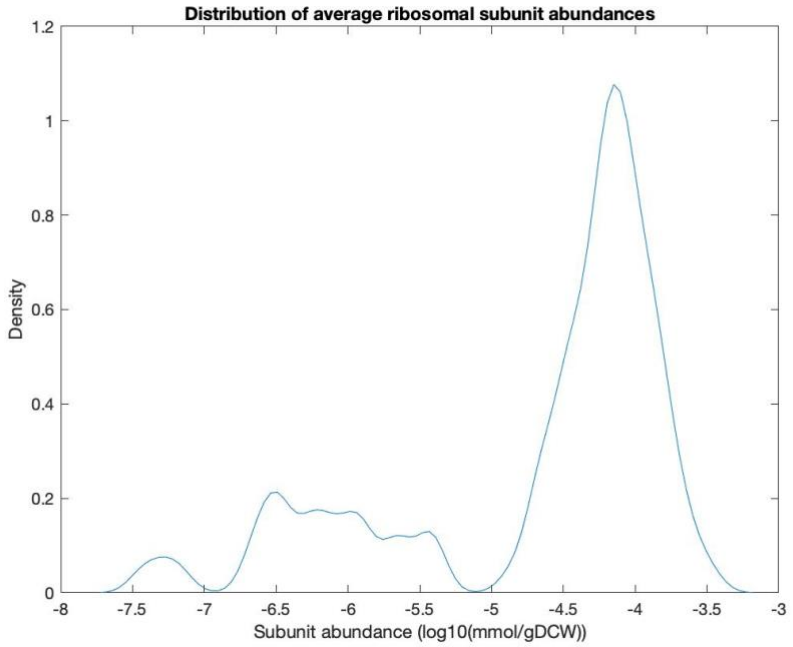

Figure S3. Distribution of average ribosomal subunit abundance obtained from absolute proteome quantification in all studied conditions.

Table S3. Kinetic data of enzymes from the central carbon metabolism and sugar utilisation pathways.

Uniprot ID	k_{cat}	Organism	Name
M7X8C7	324000	*Rasamsonia emersonii*	xylose reductase
M7WT79	558.999	*Ovis aries*	xylitol dehydrogenase
M7X6R2	6599.9	*Escherichia coli*	xylulokinase
M7XGH5	950.8379		D-arabinitol dehydrogenase
M7WGA7	39.2336	*various*	phosphoketolase
M7WSJ0	360000	*various*	acetyl-CoA carboxylase
M7XLR4	360000		acetyl-CoA carboxylase
M7XNL9	121.8003		transketolase
M7WY13	121.8003		transketolase_2
M7X0R7	121.8004		transketolase
M7XNL9	121.8004		transketolase_2
M7WY13	12.9997	*Escherichia coli*	transaldolase
M7WNZ9	650	*Homo sapiens*	glucose-6-phosphate isomerase
M7WR01	3500	*Thermotoga maritima*	glucose 6-phosphate dehydrogenase
M7WWW0	3376.5		6-phosphogluconolactonase

M7X3Z4	72.9995	*Homo sapiens*	phosphogluconate dehydrogenase
M7WUP8	39530		ribose-5-phosphate isomerase
M7XEA2	7099.9	*Spinacia oleracea*	ribulose 5-phosphate 3-epimerase
M7WNF8	441.9975	*Trypanosoma cruzi*	glyceraldehyde-3-phosphate dehydrogenase
M7WI96	441.9975		glyceraldehyde-3-phosphate dehydrogenase
M7X689	2633	*Homo sapiens*	phosphoglycerate kinase
M7XSI3	3200	*Lactococcus lactis subsp. lactis*	phosphoglycerate mutase
M7XRQ3	3200		phosphoglycerate mutase
M7X749	230.0056	*Saccharomyces cerevisiae*	enolase
M7WUI5	3204	*Geobacillus stearothermophilus*	pyruvate kinase
M7XNE0	31530	*S.cerevisiae* and others	pyruvate carboxylase
M7WZC3	529.999	*Escherichia coli*	aspartate transaminase
M7WQ86	121600	*Triticum aestivum*	malate dehydrogenase, cytoplasmic
M7X2B5	28650	*Symbiobacterium toebii*	glutamate dehydrogenase (NADP)
M7WIG9	485.9993	*Zymomonas mobilis*	pyruvate dehydrogenase

Table S4. Kinetic data of enzymes from the TCA cycle.

Uniprot ID	k_{cat}	Organism	Name
M7XE29	450.0029	*Sus scrofa*	citrate syntase
M7WHC9	179.6635		ATP-citrate lyase (ACL)
M7X6X3	200	*Mycobacterium tuberculosis*	citrate to cis-aconitate
M7X6X3	200		cis-aconitate(3-) to isocitrate
M7WQ73	200		citrate to cis-aconitate
M7WQ73	200		cis-aconitate(3-) to isocitrate
M7WN97	255.0058	*Archaeoglobus fulgidus*	isocitrate dehydrogenase
M7WW42	29.9999	*Homo sapiens*	isocitrate dehydrogenase (NAD+)
M7XE28	29.9999		isocitrate dehydrogenase (NAD+)
M7XGI7	648.9983		oxoglutarate dehydrogenase (lipoamide)
M7XGI7	898.9863		oxoglutarate dehydrogenase (dihydrolipoamide S-succinyltransferase)
M7XGI7	898.9863		glycine-cleavage complex (lipoamide)

M7WR40	648.9983	*Homo sapiens*	oxoglutarate dehydrogenase (lipoamide)
M7WR40	898.9863		oxoglutarate dehydrogenase (dihydrolipoamide S-succinyltransferase)
M7WR40	898.9863	*Homo sapiens*	glycine-cleavage complex (lipoamide)
M7WKF8	648.9983		oxoglutarate dehydrogenase (lipoamide)
M7WKF8	898.9863		oxoglutarate dehydrogenase (dihydrolipoamide S-succinyltransferase)
M7WKF8	898.9863		glycine-cleavage complex (lipoamide)
M7WPA9	898.9863		glycine-cleavage complex (lipoamide)
M7X413	898.9863		glycine-cleavage complex (lipoamide)
M7XCY6	898.9863		glycine-cleavage complex (lipoamide)
M7X0P4	898.9863		glycine-cleavage complex (lipoamide)
M7WVW2	200.9969	*Acetobacter aceti*	succinate-CoA ligase (ADP-forming)
M7WM30	200.9969		succinate-CoA ligase (ADP-forming)
M7XJJ4	259.9942	*Paracoccus denitrificans*	succinate dehydrogenase (ubiquinone-6)
M7XF32	259.9942		succinate dehydrogenase (ubiquinone-6)
M7X6W8	259.9942		succinate dehydrogenase (ubiquinone-6)
M7X560	259.9942		succinate dehydrogenase (ubiquinone-6)
M7XEU6	1150	*Escherichia coli*	fumarase
M7XHV2	134.3999	*Escherichia coli K-12*	malic enzyme (NAD)
M7XHF8	1140	*Methylosinus trichosporium*	malate dehydrogenase
M7XQ01	670.006	*Escherichia coli*	aspartate transaminase
M7XFR0	9400	*Salmonella enterica*	acetyl-CoA synthetase
M7WMR9	161	*Colwellia maris*	malate synthase
M7XLR4	360000		acetyl-CoA carboxylase
M7WLQ0	97	*Mus musculus*	carnithine o-acyltransferase
M7WSW5	120	specific activity of *S.cerevisiae*	fatty acid synthase
M7XM89	120	specific activity of *S.cerevisiae*	fatty acid synthase

Table S5. Integration of proteome to the model.

model version	Total enzymes measured	Total protein in model	Total protein amount measured (g $gDCW^{-1}$)	Total protein amount in model (g $gDCW^{-1}$)	Coverage from total proteome (%)
Xexp	402	693	0.101	0.439	23
XNlim	375	693	0.0422	0.197	21
Aexp	447	693	0.103	0.386	27
ANlim	443	693	0.0597	0.217	28
GexpUrea	386	693	0.113	0.636	18
GNlimUrea	439	693	0.0459	0.227	20

NON-EXCLUSIVE LICENCE TO REPRODUCE THESIS AND MAKE THESIS PUBLIC

I, Alina Rekena,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, Enzyme-constrained genome-scale metabolic model of *Rhodotorula toruloides*, supervised by Petri-Jaan Lahtvee, Nemailla Bonturi and Isma Belouah.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Alina Rekena

20/05/2021