

DOCTORAL THESIS

Computational Design of
Yeast-Based Cell Factories

Alīna Rekēna

TALLINNA TEHNKAÜLIKOO
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2026

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
12/2026

Computational Design of Yeast-Based Cell Factories

ALĪNA REĶĒNA



TALLINN UNIVERSITY OF TECHNOLOGY

School of Science

Department of Chemistry and Biotechnology

This dissertation was accepted for the defence of the degree Doctor of Philosophy in Chemistry and Biotechnology on 07/01/2026.

Supervisor:

Associate Professor Petri-Jaan Lahtvee
Department of Chemistry and Biotechnology
Tallinn University of Technology
Tallinn, Estonia

Opponents:

Professor Merja Penttilä
VTT Technical Research Centre of Finland
Espoo, Finland

Professor Marija Cvijovic
Department of Mathematical Sciences
Chalmers University of Technology
Gothenburg, Sweden

Defence of the thesis: 20/02/2026, Tallinn

Declaration:

Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.

Alīna Reķēna

signature



Copyright: Alīna Reķēna, 2026

ISSN 2585-6898 (publication)

ISBN 978-9916-80-453-7 (publication)

ISSN 2585-6901 (PDF)

ISBN 978-9916-80-454-4 (PDF)

DOI <https://doi.org/10.23658/taltech.12/2026>

TALLINNA TEHNIKAÜLIKOO
DOKTORITÖÖ
12/2026

**Pärmipõhiste rakuvalbrikute
in silico disain**

ALINA REKENA



Contents

List of publications	7
Author's contribution to the publications	8
Introduction	9
1 Theoretical background	11
1.1 Yeast-based cell factories	11
1.2 Cellular metabolism	13
1.3 Gene - protein relationship.....	16
1.4 Methodological approaches of computational design	18
1.5 Genome-scale metabolic models and flux balance analysis.....	20
1.5.1 Enzyme-constrained models	23
1.6 Aims and significance.....	23
2 Materials and methods	24
2.1 Strains and media	25
2.2 Plasmid construction and strain engineering	26
2.3 Laboratory evolution	27
2.4 Cultivations	28
2.4.1 Batch cultivations in actively controlled bioreactors	28
2.4.2 Batch cultivations in sensor-monitored Falcon tube bioreactors	28
2.5 Analytical methods	28
2.6 Absolute proteomics.....	29
2.6.1 Sample preparation for proteomics	29
2.6.2 LC-MS/MS analysis	29
2.6.3 Raw MS data analysis.....	29
2.6.4 Quantitative and statistical analysis.....	29
2.7 Lipidomics	30
2.8 Enzyme-constrained genome-scale models	31
2.9 Statistical analysis	32
3 Results and discussion.....	33
3.1 Study 1: Evaluating and validating enzyme-constrained genome-scale modeling of anaerobic cultivation in <i>S. cerevisiae</i> (Publication I).....	33
3.1.1 Calibration of a total-protein ecGEM enables prediction of 2,3-butanediol and glycerol co-production	34
3.1.2 Predicted and <i>in vivo</i> fluxes align in 23BD-glycerol co-producing <i>S. cerevisiae</i>	36

3.1.3	Proteome allocation: alignment and divergence between ecGEM predictions and <i>in vivo</i> data	37
3.1.4	Proteomics-constrained modeling exposes limitations of <i>in vitro</i> enzyme data.....	40
3.2	Study 2: Predicting and validating central carbon metabolism in the oleaginous yeast <i>R. toruloides</i> (Publications II and III)	42
3.2.1	Global proteome allocation analysis reveals metabolic strategies and growth constraints in <i>R. toruloides</i>	43
3.2.2	Construction and curation of the first ecGEM of <i>R. toruloides</i> reveals limits of enzyme-constrained modeling in lipogenesis.....	46
3.2.3	Predicted intracellular fluxes reveal substrate-dependent strategies for acetyl-CoA and NADPH supply during lipogenesis.....	50
3.2.4	Knockout engineering demonstrates ACK, PK, and cMAE are dispensable for growth but differ in metabolic impact	54
3.2.5	Knockout phenotyping validates and challenges ecGEM predictions in <i>R. toruloides</i>	55
4	Conclusions	63
	Broader conclusions and future perspectives.....	63
	Study 1: Anaerobic 23BD-glycerol co-production in <i>S. cerevisiae</i>	64
	Study 2: Lipogenesis in <i>R. toruloides</i>	64
	References	67
	Disclosure of Copyrighted Materials.....	80
	Acknowledgements.....	81
	Abstract.....	82
	Lühikokkuvõte.....	83
	Appendix 1	85
	Supplementary materials.....	85
	Supplementary Tables.....	87
	Supplementary figures	102
	Appendix 2	113
	Publication I	113
	Publication II	127
	Publication III	161
	Curriculum vitae.....	181
	Elulookirjeldus.....	184

List of publications

The list of author's publications, on the basis of which the thesis has been prepared:

- I Sjöberg, G., **Rekēna, A.**, Fornstad, M., Lahtvee, P.-J., van Maris, A. J. A. (2024). Evaluation of enzyme-constrained genome-scale model through metabolic engineering of anaerobic co-production of 2,3-butanediol and glycerol by *Saccharomyces cerevisiae*. *Metabolic Engineering*, 82, 49–59.
- II **Rekēna, A.**, Pinheiro, M. J., Bonturi, N., Belouah, I., Tammekivi, E., Herodes, K., Kerkhoven, E. J., Lahtvee, P.-J. (2023). Genome-scale metabolic modelling reveals metabolic trade-offs associated with lipid production in *Rhodotorula toruloides*. *PLOS Computational Biology*, 19(4), e1011009.
- III **Rekēna, A.**, Pals, K., Gavrilović, S., Lahtvee, P.-J. (2025). The role of ATP citrate lyase, phosphoketolase and malic enzyme in oleaginous *Rhodotorula toruloides*. *Applied Microbiology and Biotechnology*, 109, 77.

Additional publications not included in this thesis:

- IV Senatore, V. G.[†], **Rekēna, A.**[†], Mapelli, V., Lahtvee, P.-J., Branduardi, P. (2025). Ethylene glycol metabolism in the oleaginous yeast *Rhodotorula toruloides*. *Applied Microbiology and Biotechnology*, 109, 114.
- V Sepúlveda Del Rio Hamacek, H., Tingajeva, O., Ostertag, K., **Rekēna, A.**, Illarionov, A., Jõul, P., Monteiro de Oliveira, P., de La Caridad Martin-Hernandez, G., Muller, B., Bonturi, N., Passoth, V., Lahtvee, P.-J., Kumar, R. (2025). Acidification by nitrogen metabolism triggers extracellular biopolymer production in an oleaginous yeast. *Applied and Environmental Microbiology*, e00947-25.

Author's contribution to the publications

Contribution to the papers described in this thesis according to the CRediT (Contribution Roles Taxonomy) classification system are:

- I Methodology, Software, Formal analysis, Writing – original draft.
- II Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization.
- III Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization.

Introduction

The development of sustainable manufacturing technologies using biology increasingly relies on the ability to design and optimize microbial cell factories *in silico*. Yeasts, in particular, represent versatile eukaryotic hosts that combine robust growth with advanced genetic tools and a long history of industrial application in fermentation, food, and biotechnology. Rational design of novel yeast-based production platforms, however, requires a predictive understanding of how metabolism, energy, and protein resources are distributed under varying environmental and genetic conditions.

The rapid progress of high-throughput DNA sequencing in the mid-2000s and protein profiling has enabled the quantitative characterization of cellular components at unprecedented scale. These advances laid the foundation for computational frameworks that help to understand how genetic information (genotype) gives rise to observable cellular behavior (phenotype) through mechanistic modeling. Among such approaches, genome-scale metabolic models (GEMs) allow the systematic study of cellular metabolism using metabolic rate (flux) predictions *in silico*. Since the first GEM of baker's yeast *Saccharomyces cerevisiae* was developed in 2003 numerous nonconventional yeast species, newly isolated for their unique metabolic capabilities, have been introduced into biotechnological applications.

Over the past two decades, GEMs have been expanded and refined into community-curated versions, culminating in the Yeast9 model released in 2024. These models have been successfully applied to guide the production of fuel precursors, organic acids, medicinal compounds, flavours, fragrances, and other industrially relevant metabolites, enzymes and proteins. Yet, despite these advances, their predictive power remains fundamentally dependent on accurate experimental parameterization and validation. In particular, conventional GEMs neglect the finite capacity of the proteome and lack representation of enzyme kinetics and regulation. Moreover, not always experimental yeast data are readily available across different conditions. This limitation becomes critical in nonconventional yeast species, such as the lipid-producing yeast *Rhodotorula toruloides*, where experimental data remain sparse and standard assumptions derived from *S. cerevisiae* no longer hold.

To address these challenges, in this thesis I used a combination of genetic, biochemical, and computational approaches to develop and validate predictive models of yeast metabolism. Enzyme-constrained GEMs (ecGEMs) were employed as the central modeling framework, linking metabolic activity to the amount and efficiency of enzymes measured in the cell. Quantitative proteomics, physiological data from controlled growth experiments, and genetic perturbations were integrated to parameterize and evaluate ecGEMs of *S. cerevisiae* and *R. toruloides*. *S. cerevisiae*, the canonical model organism for yeast biology, was used as a benchmark to assess model accuracy under newly tested environmental conditions. *R. toruloides*, by contrast, served as a case study for extending enzyme-constrained modeling to a nonconventional, lipid-accumulating yeast, enabling the identification of biological and methodological factors that limit current model predictiveness.

The results of this work reveal both the capabilities and the boundaries of current computational design frameworks. The models accurately captured proteome limitations and substrate-dependent metabolic allocation, yet they exposed systematic deviations arising from strict models' assumptions. These findings not only provide new insight into the regulatory mechanisms underlying energy and redox balance in yeast and clarify the

data requirements for extending quantitative models to unconventional species, but also demonstrate the potential of such models to bridge quantitative proteomics and genome-scale modeling, and provides perspectives on how future integration of regulation and kinetics can advance *in silico* biomanufacturing design.

1 Theoretical background

This chapter aims to provide a scope of the studies conducted in this thesis. Although the topic of this thesis spans broad and rapidly evolving areas of science, such as systems biology and metabolic engineering, I will focus on genome-scale metabolic models and how we use them in studying yeast. While acknowledging how quantitative modeling frameworks, such as genome-scale models, are used to extract biological insights, in this chapter I will also discuss how these models can be experimentally validated.

This thesis is written from a very practical standpoint in mind, emphasizing the computational design of microbial cell factories. Therefore, throughout the theoretical background, I want to present it from a curiosity driven perspective: on what principles are these models built, and how accurately do they represent biological reality? In the final two subchapters, I will present core methodological concepts underlying the presented studies, but I will begin the chapter with the biochemical foundations necessary to understand the basis of computational design.

1.1 Yeast-based cell factories

Yeasts are unicellular organisms that naturally occur on plant surfaces, soil and aquatic environments. In nature, they live in diverse places that are rich in biomass, and they play a key role in carbon cycling. In the laboratory, they are foundational systems for eukaryotic cell biology and biotechnology (Walker, 2000). On one hand, yeast cellular functionalities, such as well defined nucleus, that stores genetic material, and other cellular compartments follow by definition the same rules as plants, insects, animals and humans. On the other hand, their unicellular nature allows easier studying, compared to bacteria and other unicellular life forms. All in all, that is why yeasts are used from fundamental studies understanding the rules of life, to industrial applications in microbial catalysis – a discipline known as biotechnology. From an evolutionary perspective, yeasts belong to the Fungi kingdom, primarily phylum Ascomycota and to a lesser extent Basidiomycota (Kurtzman, 2011). The two phyla diverged over 500 million years ago, their known differences are often grouped into different ecological roles, cellular organization, and organisation of metabolism. Yeasts can reproduce both asexually and sexually. As most yeasts that have been historically used by humans entail asexual reproduction, forming a bud of an original cell, they are popular by the term budding yeasts.

Yeasts must be great consumers of simple sugars, such as monosaccharides such as glucose and fructose, and disaccharides such as sucrose and maltose, as the name of most and widely known yeast in Latinized Greek means “sugar fungus” – *Saccharomyces* (Chambers & Pretorius, 2010). The most well known species of yeast is ascomycetous *Saccharomyces cerevisiae*. Due to its Eukaryotic nature, it has been used as a model organism to study cellular signaling, division, aging, death, and other cellular processes. At least ten Nobel prizes have been awarded for the work in yeasts (Hohmann, 2016), mainly in the category of physiology and medicine. And I want to highlight the 2001 award in Medicine to Hartwell, Hunt and Nurse for unraveling the genetic logic of the eukaryotic cell cycle, showing how genotype determines phenotype through molecular logic.

Given the extensive knowledge of yeast biology, these microorganisms have long been engineered for industrial use – not only in the production of bread, beer, and wine, but also as biocatalysts for ethanol and other biofuels (Nielsen & Keasling, 2016; Walker &

Stewart, 2016). Although ethanol is by far the most widely produced biofuel using yeast biotechnology, ethanol formation is actually an exception in the context of yeast metabolism. Yeasts are inherently efficient at synthesizing highly reduced, long-chain molecules (C₂₂–C₂₄), including fatty acids, sterols, and long-chain alcohols (Koutinas et al., 2014). Thus, the extensive use of yeasts for ethanol production reflects historical and industrial choices rather than their metabolic predisposition. Due to their robustness, metabolic diversity, and substrate flexibility, non-conventional yeasts have gained increasing prominence as emerging hosts for industrial biomanufacturing. For instance, non-conventional yeasts, such as *Yarrowia lipolytica* (Ledesma-Amaro & Nicaud, 2016) and *Rhodotorula toruloides* (Park et al., 2018) are recognized for lipid and carotenoid biosynthesis, *Kluyveromyces marxianus* (Nonklang et al., 2008) for rapid growth and high temperature tolerance, *P. pastoris* (now *Komagataella phaffii*) for protein synthesis (Cregg et al., 2000), *Z. bailii* for tolerance to organic acids (Branduardi et al., 2004). Chemical building blocks produced with microbial cell factories are more and more used in the production of pharmaceuticals, flavors, fragrances, cosmetics, materials (Nielsen, 2019), and the opportunities are theoretically endless. The global bio-based chemicals market is estimated to exceed USD 125 billion by 2030, with microbial processes contributing a rapidly increasing share (Grand View Research, n.d.; Korosuo et al., 2024).

The choice of carbon source is a critical factor in designing cell factories, as it directly influences not only central carbon metabolism, redox balance, and biosynthetic capacity, but also their economic feasibility. Glucose, the most abundant sugar in nature, supports rapid growth and high glycolytic fluxes, but often leads to fermentative metabolism and catabolite repression that limits the utilization of alternative substrates (Gancedo, 1998). In contrast, non-fermentable carbon sources such as acetate, ethanol, or fatty acids promote respiratory metabolism and can alter cellular energy efficiency and redox status (Papanikolaou & Aggelis, 2011b). Glycerol, a byproduct of biodiesel production, represents an inexpensive and sustainable substrate. Xylose and other pentoses derived from lignocellulosic hydrolysates are of growing interest for second-generation bioprocesses; however, their efficient assimilation often requires metabolic engineering to overcome limited native transport and conversion capacities (Kwak et al., 2019; Sàncchez Nogué & Karhumaa, 2015). Consequently, optimizing carbon source utilization not only enhances process sustainability by integrating waste-derived substrates but also enables the tuning of cellular physiology toward improved yield of fuels, lipids, and biochemicals.

The term microbial cell factory emphasizes the analogy to traditional production plants – yet these “factories” operate at the cellular level, converting renewable substrates into valuable products through metabolic pathways. With advances in systems biology and synthetic biology, the molecular arsenal and efficiency of yeast-based cell factories are rapidly improving, supporting the development of biorefineries that transform biomass and waste into sustainable biochemicals and biofuels, and move away from fossil fuel-based refineries.

Yeasts not only serve as models for understanding eukaryotic cell biology, but also as powerful platforms for sustainable production, bridging fundamental biology and industrial biotechnology.

1.2 Cellular metabolism

Metabolism inside the cell is an integrated network of biochemical reactions that sustains life. The metabolic network in a yeast cell operates in a manner that nutrients, taken up from the environment as substrates, are converted into products that help the cell to grow, and byproducts, such as CO_2 , ethanol or others. A yeast cell needs a carbon source, energy source, nitrogen source, minerals and vitamins. Glucose is the most common carbon source, which is also simultaneously an energy source (Villadsen et al., 2011). Yeasts (like animals and most microorganisms) are heterotrophs, they cannot use carbon dioxide from the atmosphere as their sole carbon source.

The cell's primary goals via metabolism are to extract energy from substrate(s) and to synthesize molecules essential for self-maintenance, growth and reproduction. The series of enzyme-catalyzed reactions constitute metabolic pathways. Catabolic pathways, such as glycolysis and tricarboxylic acid cycle, break down substrates into smaller, simpler end products to release energy, while anabolic pathways, also called biosynthesis, synthesize macromolecules, such as proteins, lipids, and nucleic acids, from small, simple precursors (Nelson et al., 2021). Anabolic reactions require energy as input. In the metabolic network, the Gibbs free energy, which cells can and must use to drive chemical reactions, is acquired from nutrient molecules and stored as an energy currency molecule (ie. adenosine triphosphate, ATP) and other cofactors (ie. NADPH). Biological energy transformations obey the laws of thermodynamics, so evolution has constructed metabolic pathways that are thermodynamically feasible, such as carbon-carbon bond break, isomerizations, group transfers, oxidation-reductions, and others (Nelson et al., 2021). In catabolism, ATP provides energy by group transfers, not by simple hydrolysis. In anabolic pathways, ATP hydrolysis drives energetically unfavourable reactions forward. All small molecules that participate as substrates, intermediates, or products in the metabolism, are called metabolites.

The central metabolic pathways, which are very few and remarkably similar in all forms of life, are associated with glucose oxidation but each fulfills a different purpose. Through glycolysis, glucose can be oxidized to a three-carbon compound pyruvate to provide ATP and metabolic intermediates. *S. cerevisiae* and most other yeasts rely on the Embden-Meyerhof-Parnas (EMP) pathway for glycolysis, as they lack alternative glycolytic routes. It is a series of 10 reactions that yields 2 molecules of ATP, 2 molecules of nicotinamide adenine dinucleotide (NADH) and supplies glyceraldehyde-3-phosphate and phosphoenolpyruvate as precursors for biosynthetic pathways (Teusink et al., 1998). NADH is one of the few prominent electron carriers in central metabolism, which temporarily stores and transports reducing equivalents to other enzymes or pathways where they can be used. Through the pentose phosphate (phosphogluconate) pathway (PPP), glucose can be oxidized to yield ribose 5-phosphate for nucleic acid synthesis plus a molecule of CO_2 (oxidative PPP, oxPPP), NADPH for reductive biosynthetic processes, and xylulose-5-phosphate. Non-oxidative phase interconverts pentose phosphates with glycolytic intermediates (glyceraldehyde-3-phosphate and fructose-6-phosphate), or transforms into erythrose-4-phosphate (for amino acid synthesis). PPP runs parallel to glycolysis. The branching point of PPP and glycolysis is glucose-6-phosphate, which is the first metabolic intermediate of glycolysis. These pathways in eukaryotes are cytosolic (Kruger & Von Schaewen, 2003).

Pyruvate can either be fermented, or undergo the tricarboxylic acid cycle (TCA). Fermentation is ubiquitous among microorganisms and even some higher organisms because it provides a simple, robust, and rapid mechanism for sugar utilization. Cells

obtain ATP solely by breaking down glucose into pyruvate and, like ethanol fermentation in yeast, the pyruvate is converted in two molecules of ethanol and CO₂, as end products. It is a two step process: decarboxylation by pyruvate decarboxylase and reduction through alcohol dehydrogenase, using NADH generated during the dehydrogenation of glyceraldehyde 3-phosphate in glycolysis (Nelson et al., 2021). The CO₂ produced by pyruvate decarboxylation in brewer's yeast is responsible for the characteristic carbonation of champagne, but in baker's yeast – causes dough to rise.

The complete oxidation of glucose, when pyruvate produced by glycolysis is further oxidized to water and CO₂, is termed cellular respiration. Unlike fermentation, it involves an external O₂ as the final electron acceptor. The amount of released energy, conserved in the form of theoretically up to 36 ATP per glucose by a process called oxidative phosphorylation (oxPhos), is far more than in fermentation (in process called substrate-level phosphorylation). Nevertheless, many aerobic organisms, including *S. cerevisiae*, retain the fermentative capacity. The Crabtree effect is a hallmark of *S. cerevisiae* metabolism, describing its tendency to ferment glucose to ethanol even in the presence of oxygen when sugar is abundant. First described by Herbert Crabtree (Crabtree, 1929) and later quantified by van Dijken and Scheffers (Dijken & Scheffers, 1986), the effect reflects a metabolic shift from respiration to fermentation when the glucose uptake rate exceeds the respiratory capacity of the mitochondria, typically above a critical dilution rate of 0.25–0.3 h⁻¹ in chemostat cultures. Physiologically, it manifests as a biphasic growth pattern: during glucose excess, yeast cells produce ethanol aerobically, and once glucose is depleted, they undergo the diauxic shift, oxidizing ethanol to CO₂ and water. The effect is not universal to all yeasts but characteristic of “Crabtree-positive” species, including *S. cerevisiae* and *Kluyveromyces marxianus*, which evolved in sugar-rich environments where rapid substrate consumption provided a competitive advantage (Hagman & Piškur, 2015; Pfeiffer & Morley, 2014). In contrast, “Crabtree-negative” yeasts such as *Pichia stipitis* (now *Scheffersomyces stipitis*) and *Candida utilis* maintain fully respiratory metabolism in the presence of oxygen regardless of glucose concentration (Dijken & Scheffers, 1986; Merico et al., 2007). The Crabtree effect has a regulatory advantage that supports high glycolytic fluxes and rapid ATP turnover (Malina et al., 2021; Pfeiffer & Morley, 2014).

To undergo TCA cycle, pyruvate is converted to acetyl groups and activated by an irreversible oxidative decarboxylation reaction by pyruvate dehydrogenase (PDH) complex, producing 1 molecule of CO₂ and 1 molecule of NADH. Acetyl-CoA is condensed with oxaloacetate (OAA) to form citrate (CIT), which is subsequently rearranged and oxidized, regenerating OAA for another cycle run. Therefore, TCA cycle is also called the citric acid cycle, or the Krebs cycle (after its discoverer, Hans Krebs, who received a Nobel Prize in Physiology or Medicine for this discovery in 1953). In eukaryotes, it takes place entirely in mitochondria. In this cycle, four of the eight steps are oxidations, conserving the energy in the form of 3 reduced coenzymes NADH, 1 flavin adenine dinucleotide (FADH₂), and 1 energy currency molecule (GTP or ATP). The chemical energy extraction releases 2 molecules of CO₂ per 1 pyruvate molecule. Because glucose molecule yields 2 pyruvate molecules, full oxidation of glucose via the TCA cycle releases total of 6 CO₂ (2 from PDH and 4 from the TCA cycle). Intermediates of the TCA cycle are siphoned off as biosynthetic precursors (2-oxoglutarate, succinate, oxaloacetate); it is a hub of metabolism, closely regulated (Nelson et al., 2021).

Respiration proceeds with NADH and FADH₂ in mitochondria transferring electrons to the electron transport chain. In case oxygen as the final electron acceptor is not available,

the buildup of reduced cofactors inhibits TCA enzymes (dehydrogenases), and the cycle stalls or shifts to branched operation to supply succinate, oxaloacetate, and 2-oxoglutarate, but not as a full oxidative loop. In case glucose is unavailable, cells shift metabolism and the TCA cycle is fed with acetyl-CoA and 2-oxoglutarate from other carbon sources (ethanol, acetate, fatty acids, amino acids), facilitating the production of phosphoenolpyruvate (PEP), a glycolytic intermediate – precursor of gluconeogenesis (essentially reverse glycolysis, but with bypass reactions to irreversible phosphofructokinase and hexokinase reactions). In such a scenario, ATP production can continue, if O_2 is available. There are 12 precursors – building blocks – for biomass synthesis required. They are generated by central metabolic pathways: glucose-6-phosphate, fructose-6-phosphate, glyceraldehyde-3-phosphate, 3-phosphoglycerate, phosphoenolpyruvate, pyruvate (all glycolysis), ribose-5-phosphate, erythrose-4-phosphate (both PPP), acetyl-CoA, oxaloacetate, 2-oxoglutarate, and succinyl-CoA (all TCA cycle) (Neidhardt et al., 1990).

Electron transport chain (ETC) in yeasts is a series of inner mitochondrial membrane proteins (Complexes I–IV, ubiquinone (coenzyme Q), and cytochrome c) releasing electrons and protons from their reduced cofactors NADH and $FADH_2$ and using the free energy to pump protons from mitochondrial matrix to the intermembrane space, generating an electrochemical proton gradient (the proton motive force). Flow of protons in yeasts, animals and many bacteria is accepted at the end by oxygen. No ATP is produced directly at this stage. Proton flow through the last protein complex, ATP synthase, drives the rotary motion of its rotor subunits, which induces conformational changes in the catalytic F_1 domain that enable phosphorylation of ADP to ATP (thus, the name oxidative phosphorylation) on the matrix side of the inner mitochondrial membrane (Boyer, 1997). In analogy, the proton gradient across the inner mitochondrial membrane is like water stored behind a dam – it holds potential energy. The stoichiometry of ATP production depends on (i) the number of subunits in the c ring of the F_0 complex and (ii) the number of protons per NADH pumped in the intermembrane space (widely accepted experimental value is 10 protons). Full rotation of the F_1 domain produces 3 ATP. Atomic force microscopy has shown yeast mitochondria have 10 c subunits. Adding 3 protons to bring 3 P_i into the matrix brings a total cost of 13 protons (~ 4 protons/ATP). Oxidizing 1 NADH produces (10 protons/4 protons ATP $^{-1}$) 2.5 ATP, which due to transport costs, potential proton leak and less efficient NADH shuttles is lower than the theoretical maximum. If the cell has a different electron transport chain, like *S. cerevisiae* who lack one of the ETC proteins (Complex I), the ATP yield per NADH drops even more to (6 protons/4 protons ATP $^{-1}$) 1.5 (Nelson et al., 2021).

Fatty acid biosynthesis is one of the cytosolic biosynthetic pathways – reactions that provide precursors for macromolecules and secondary metabolites. It proceeds via a cyclic sequence of four main enzymatic reactions – condensation, reduction, dehydration, and a second reduction – catalyzed by the multifunctional fatty acid synthase (FAS) complex. Each cycle elongates the acyl chain by two carbons, contributed from malonyl-CoA, which itself is generated by carboxylation of acetyl-CoA through the action of acetyl-CoA carboxylase (ACC). Thus, for every elongation step, one molecule of acetyl-CoA (as the starter unit) and one molecule of malonyl-CoA (as the extender unit) are consumed, along with 2 molecules of NADPH as reducing equivalents. The process continues iteratively until the full-length fatty acid, typically palmitoyl-CoA (C16:0), is synthesized. In eukaryotic cells such as *S. cerevisiae*, this process occurs in the cytosol, followed by either elongation and desaturation by dedicated enzymes in the endoplasmic

reticulum, or acylation of glycerol backbone to synthesize neutral lipids (triacylglycerols, TAGs) (Ratledge & Wynn, 2002a; Tehlivets et al., 2007).

Xylose metabolism is crucial in the context of biorefineries of lignocellulosic biomass – one of the most abundant biomass sources on earth (Sánchez Nogué & Karhumaa, 2015). The xylose assimilation pathway converts xylose into xylulose-5-phosphate through a series of enzymatic reactions catalyzed by xylose reductase (XR), xylitol dehydrogenase (XDH), and xylulokinase (XK). The resulting xylulose-5-phosphate enters the pentose phosphate pathway. In yeasts, this process often causes a redox imbalance due to the differing cofactor specificities of XR and XDH, whereas bacteria use a simpler isomerase pathway that avoids this limitation (Bhosale et al., 1996).

Central metabolism is versatile, regulated and robust to produce biomass in a wide variety of environmental conditions. While the general metabolic architecture is conserved across all domains of life, yeast metabolism exhibits adaptations (such as Crabtree effect) that make yeasts powerful model systems and hosts for metabolic engineering compared with both prokaryotic and multicellular eukaryotic systems.

1.3 Gene - protein relationship

Almost all chemical reactions in living systems are catalyzed by enzymes, and all known enzymes are proteins (F. H. Crick, 1958). Enzymes are proteins with catalytic activity. Proteins are sequences of covalently bonded amino acids which form three-dimensional structures. In nature, there are 20 different amino acids that can be combined in sequences to form proteins. Polypeptides with molecular masses greater than approximately 10 kDa (corresponding to ~100 amino acid residues) are conventionally termed proteins, whereas smaller chains are referred to as oligopeptides or peptides (Nelson et al., 2021). This distinction reflects the minimal chain length required for a polypeptide to fold into a stable tertiary structure and exhibit independent biological function. Proteins also perform a broad range of other cellular functions, including regulation, structural support, but it is not their main role. Proteins are at the centre of a living system because to produce a new small molecule, the cell needs to produce a new protein to catalyse the reaction.

Protein synthesis requires a signal and information on the sequence of amino acids. This is controlled by genes, but not directly. Genes are linear (one-dimensional) sequences of deoxyribonucleic acids (DNAs) that carry this information inside the cells. The signal for protein synthesis originates from the activation of gene transcription (via some transcription factor (TF), which is a protein too) in response to cellular and environmental cues (nutrient levels, stress, cell cycle, external signals). In yeast, genes are densely packed into structures called chromosomes (16 in *S. cerevisiae*) primarily located in the nucleus, with a small number present in mitochondrial DNA. Protein synthesis proceeds through two main stages: transcription in the nucleus, where DNA is unpacked, transcribed into messenger ribonucleic acid (mRNA) by RNA polymerase II (and TFs) and exported to the cytoplasm through nuclear pores, and translation in the cytoplasm, where ribosomes decode the mRNA to assemble linear amino acid chains corresponding to sequence of DNA in a gene. While the classic view of “one gene – one protein” provides a useful simplification, the relationship between genes and proteins is more complex. In eukaryotes, a single gene can give rise to multiple protein isoforms through mechanisms such as alternative splicing, alternative start sites, or post-translational processing. Alternatively, a protein complex (such as ATP synthase) is a functional assembly composed of multiple distinct polypeptide subunits, each

encoded by a separate gene, illustrating cooperative interactions among several gene products rather than from single proteins acting alone. Conversely, multiple genes may encode identical or functionally redundant proteins, as seen in gene duplicates or isoenzymes. In yeast, alternative splicing is rare, but alternative promoter usage and gene duplication are common, providing metabolic and regulatory flexibility. Translation occurs either on free cytosolic ribosomes or on those bound to the endoplasmic reticulum, depending on the destination of the protein. Mitochondria also possess their own ribosomes for the synthesis of a few organelle-encoded proteins. Following translation, proteins fold and may undergo post-translational modifications before reaching their functional location. Although the final shape of the folded protein is dictated by amino acid sequence, the precise three-dimensional structure (native conformation) is crucial to its function and requires the right cellular environment – pH, ionic strength, metal ion concentrations, etc. Thus, DNA sequence alone is not enough to form and maintain a fully functioning cell (Alberts et al., 2022; Nelson et al., 2021).

Synthesis of a protein true to the information specified in its mRNA requires energy. One ATP for amino-acid activation by aminoacyl-tRNA synthetase, two GTP for elongation and ribosomal translocation, and a fractional ATP for initiation and chaperone-mediated folding. Protein biosynthesis is one of the most energy-demanding cellular processes. RNA and DNA synthesis each require one nucleoside or deoxynucleoside triphosphate per monomer addition, together with auxiliary ATP usage by helicases, ligases, and topoisomerases. Polysaccharide and cell-wall polymerization involve UTP or ATP hydrolysis to generate activated sugar donors (e.g., UDP-glucose, GDP-mannose), while lipid biosynthesis requires ATP for acyl-CoA activation and malonyl-CoA formation, in addition to NADPH for reductive steps (Alberts et al., 2022; Nelson et al., 2021). This *polymerization cost* represents the sum of all high-energy bond hydrolyses directly required to convert metabolic precursors into macromolecules. Based on chemostat energy-balance calculations for *S. cerevisiae*, Verduyn et al. (Verduyn, 1991; Verduyn et al., 1990) quantified this demand as approximately 24 mmol ATP per g cell dry weight (gDW).

Mechanisms of protein biosynthetic machinery are remarkably well conserved across all life-forms because a typical cell requires thousands of different proteins to respond to a cell's needs at any given moment. The cellular resources that are devoted to this process are remarkable; up to 90% of the chemical energy and more than 35% of the cell's dry weight (CDW). Despite complexity, a polypeptide of 100 residues in an *Escherichia coli* cell (at 37 °C) is synthesized in about 5 seconds (Nelson et al., 2021). That is because each of 20 amino acids is encoded by a triplet combination of 4 DNA nucleotide bases (adenine (A), thymine (T), guanine (G), and cytosine(C)), known as the genetic code. The code is nearly universal, non-overlapping, and degenerate, meaning that most amino acids are encoded by multiple codons (F. H. C. Crick, 1968). Several codons, such as initiation codon AUG and termination codons UAA, UAG, and UGA) serve special functions and do not code for any known amino acids. Some codons for a particular amino acid are used more frequently than others, called codon (usage) bias, refers to the unequal usage of synonymous codons in an organism's genome (Nelson et al., 2021). In genetic and metabolic engineering, codon optimization is crucial when expressing heterologous genes, as mismatched codon usage can severely limit protein production and metabolic fluxes. Rational adjustment of codon usage

allows fine-tuning of enzyme expression levels and pathway balance, improving the performance of synthetic or engineered cell factories (Sharp & Li, 1987).

The central dogma of molecular biology describes the unidirectional flow of genetic information from DNA to RNA to protein, and not the other way around (F. Crick, 1970). DNA encodes hereditary information, but proteins are the molecular instruments, through which genetic information is expressed. While the core framework remains valid, discoveries such as reverse transcription, RNA-based regulation, and epigenetic control have expanded the concept to encompass multiple feedback and regulatory layers. Nevertheless, these concepts revealed how genotype determines phenotype. Together, the discovery of genetic code unified genetics and biochemistry, laid the foundation for molecular genetics, and opened the way to biotechnology, synthetic biology, and modern computational modeling of cellular systems. The understanding of gene-protein relationships enables to transform biology into a rule-based information system that can be modeled, simulated, and engineered.

1.4 Methodological approaches of computational design

Gene sequence that is linked to a protein sequence of known function is a logical rule that can be stored as an information bit. The first entries linking genes to proteins came from manual knockout and biochemical characterization, followed by sequencing of genes (Sanger et al., 1977) and proteins (Edman degradation) and curated entry into early databases, such as SWISS-PROT (Bairoch, 1996). In the 1980s sequencing accelerated, but annotation was still manual. Experts compared sequences and literature to assign enzyme functions. From 1956, Enzyme Commission (EC) numbers were used to standardize reaction types those enzymes catalyze (Dixon & Webb, 1958). Gene-protein-reaction (GPR) associations were assigned by human curation, not algorithms. Even when sequencing was semi-automated (Sanger-based in the 1980s), their functional mapping was not automatic yet.

With the advent of whole genome sequencing, including single-cell technologies, in the 1990s powered by next generation sequencing technologies (NGS) (Goodwin et al., 2016; Levy & Myers, 2016; Shapiro et al., 2013), gene sequencing became automated. Mass spectrometry methods developed for detection and quantification of thousands of proteins inside the cell (Aebersold & Mann, 2003; Millán-Oropeza et al., 2022; Sánchez et al., 2021). The study of these datasets is called -omics (proteomics, genomics, and so forth) (Cammack, 2006). Altogether, thousands of new gene and protein sequence data were generated, but there was no time for manual annotation. Therefore, functional annotation became semi-automated to: (i) compare new sequences to known databases (BLAST), (ii) transfer functional annotation from homologs (ortholog mapping OrthoFinder (Emms & Kelly, 2019)), and (iii) assign functional categories such as Gene Ontology (GO) terms (The Gene Ontology Consortium et al., 2021), enzyme commission (EC) numbers, and pathway membership automatically via databases including KEGG (Kanehisa et al., 2021), BioCyc (Karp et al., 2019), Swiss-Prot, Uniprot (Bairoch, 1996), BRENDA (Schomburg, 2002).

Modern functional annotation pipelines use gene sequencing (Illumina, Nanopore, PacBio) and proteomic identification (Mascot, MaxQuant), bioinformatics and database integration to infer GPRs automatically, providing the *molecular link* for systems biology models that links genome content to metabolic capability (The UniProt Consortium et al., 2023). Bioinformatics and machine learning models trained on millions of known

sequences to predict enzyme functions from sequence features (Y. Li et al., 2018; Ryu et al., 2019). Only the quality control and GPR curation is refined by expert curation for accuracy.

Diverse uses of the gene-protein relationship – from stochastic fluctuations to flux predictions – illustrate how molecular biology information becomes a quantitative element of different types of computational design strategies (**Table 1**). Modeling frameworks that directly quantitatively link genotype to phenotype for describing and predicting cellular behaviour are constraint-based models; they provide a mechanistic mapping between genes, enzymes, and reactions that can be expressed quantitatively through flux constraints and by extension enzyme capacity terms (Lu et al., 2024). Beyond these, the gene-protein relationship plays distinct roles across other computational design strategies. In boolean / logic models, it defines qualitative activation rules for genes and reactions, forming the foundation of regulatory network simulation (Kauffman, 1969; Thomas, 1973). In stochastic models, it underlies probabilistic gene expression events that generate variability in protein abundance (Elowitz & Leibler, 2000; McAdams & Arkin, 1997). In regulatory network models, genes produce proteins that act as transcription factors regulating other genes, forming dynamic feedback circuits that quantitatively define gene-protein causality (Gardner et al., 2000; Karlebach & Shamir, 2008). In kinetic models, protein concentrations appear explicitly in rate equations as catalysts determining reaction velocities, linking gene expression to metabolic dynamics (Teusink et al., 2000). At the protein level, structural enzyme design uses the gene-protein relationship directly to manipulate amino acid sequences and catalytic properties, thereby improving reaction kinetics (Huang et al., 2016; Jumper et al., 2021). Machine learning-based prediction tools exploit the vast landscape of gene-protein data to infer enzyme functions, kinetics, and phenotypes from sequence or omics features (Heckmann et al., 2018; F. Li et al., 2022). Beyond constraint-based approaches, emerging multi-scale models extend predictive power toward dynamic and regulatory phenomena, such as kinetic and expression-coupled models (ME-models), constraint-based modeling combined with transcriptional regulatory networks, multi-scale and whole-cell models (WCMs), as reviewed elsewhere (Lu et al., 2024).

Despite enormous demand for building strong chassis strains for metabolic engineering, advanced computational models for rational strain design are still in their infancy (Lu et al., 2024). Nevertheless beyond their role in validation, computational design tools are increasingly used to accelerate rational metabolic engineering. Computational design also facilitates multi-omics integration, as described in the next section, helping to disentangle regulatory from capacity-based constraints and to contextualize proteomic or transcriptomic shifts (O'Brien et al., 2015). Moreover, these models assist in strain comparison, hypothesis generation, and resource allocation analysis, providing a quantitative basis for enzyme economy and proteome investment. Together, computational modeling is a cornerstone of predictive and data-driven strain design, bridging molecular mechanisms with system-level optimization and enabling faster, more sustainable engineering of cell factories.

Table 1. Computational design strategies for cell factories.

Design framework	Scope and scale	Mathematical formulation	Role of gene-protein relationship
Constraint-based modeling	Steady-state fluxes in metabolism	Linear programming	Quantitative, stoichiometric
Boolean / logic modeling	Regulatory or metabolic activation states	Boolean logic	Qualitative (on/off)
Stochastic modeling	Randomness in molecular events	Gillespie / stochastic differential equations	Noise in gene-protein expression, indirect
Regulatory (gene regulatory networks)	Control of gene expression	Boolean / differential equations	Causal, with quantitative link
Kinetic modeling	Dynamic metabolic reaction rates	Ordinary differential equations	Often parametrized indirectly
Structural enzyme design	Optimize catalytic properties, cofactor use, stability	Physics-based	Direct sequence
Machine learning based prediction tools	Predict enzyme function, kinetics or phenotype	Data driven / supervised or deep learning	Implicit (learned from data)

1.5 Genome-scale metabolic models and flux balance analysis

The gene-protein relationship forms the structural foundation of genome-scale metabolic models (GEMs). Early metabolic reconstructions were primarily stoichiometric representations of biochemical reactions, disconnected from the genome. With the advent of complete genome sequencing and functional annotation, each reaction could be formalized as gene-protein-reaction (GPR) rules. These associations, expressed as Boolean logic (“AND” for multi-subunit complexes, “OR” for isoenzymes), allow creating a mechanistic map from genotype to metabolic phenotype (Thiele & Palsson, 2010).

GEM is constructed using the organism's genome information and it encompasses all biochemical reactions inside the cell, including reactions of enzymes and transporters encoded in the DNA (Thiele & Palsson, 2010). Model reconstruction is semi-automatic – ModelSEED, CarveMe, RAVEN, KBase, using curated databases, such as SGD (Hellerstedt et al., 2017), KEGG, BioCyc, Reactome, Uniprot, and experimental data on substrate

usage (Domenzain et al., 2021). The scale of a genome makes it a valuable source of information about the organism. The model's quantitative conception is based on ordinary differential equations of change of metabolite concentrations in time. This means that the production of a product with rate v is accompanied by the decay of a substrate with rate $-v$, with stoichiometric coefficients denoting the proportion of substrate and product molecules involved in a reaction. For a metabolic network consisting of m substances and r reactions, the system dynamics is described by the system equations (**Equation 1.1**), or the mass balance equations assuming reactions are the only cause for mass flow:

$$\frac{dS_i}{dt} = \sum_{j=1}^r n_{ij} v_j \text{ for } i = 1, \dots, m \quad (1.1)$$

Where quantities n_{ij} are the stoichiometric coefficients of the i th metabolite in the j th reaction. The stoichiometric coefficients n_{ij} assigned to the compounds S_i and the reactions v_j can be combined into the stoichiometric matrix N , where each column belongs to a reaction and each row to a compound.

Altogether, the mathematical description of the metabolic system in steady state consists of a vector S of concentration values, a vector v of reaction rates, a parameter vector p , and the stoichiometric matrix N (**Equation 1.2**).

$$\frac{dS}{dt} = Nv = 0 \quad (1.2)$$

The stoichiometric matrix N does not contain information about pre-assumed thermodynamics (reaction reversibility), but the sign in front of rate v describes it. If complete kinetic information is not available, as is the true situation for many of reactions, it is possible to assume a metabolic quasi-steady state, in which rates of metabolite formation and degradation inside the cell are equal ($Nv = 0$), ie. the mass of metabolites does not accumulate internally. Being of a scale of a genome, the N matrix is very large and underdetermined; metabolism is very interconnected and the number of reactions exceeds the number of metabolites (Varma & Palsson, 1994). From linear algebra, this equation has no trivial solutions (Klipp, 2016).

To simulate and analyze the flow of metabolites with GEM, the powerful Flux Balance Analysis (FBA) method is used (Lewis et al., 2012). FBA uses linear programming to find a distribution of reaction rates (fluxes) by optimizing for an objective function (z) while accounting for mass conservation and thermodynamic constraints but satisfying the constraints of the metabolic network (lb , ub) under steady state assumption (Orth et al., 2010) (**Equation 1.3**).

$$\begin{aligned} Nv &= 0 \\ lb < v_i < ub, \text{ optimization of objective } z \end{aligned} \quad (1.3)$$

Underdetermined systems are characterized by a solution space of infinite solutions, rather than just one. The constraints of a metabolic network, which calibrate model behavior, represent substrate uptake and secretion rates, called bounds b , and these are experimentally determined values. Through the constraints as the lower bound (lb) and upper bound (ub) of metabolic reactions as described in **Eq. 1.3**, the model integrates both experimental data and a pre-assumed reaction thermodynamics (Edwards et al., 2002). A commonly used objective function z in FBA is maximization of cellular growth rate (Feist & Palsson, 2010). The growth rate is implemented as a biomass synthesis pseudo-reaction, similarly as external metabolites that are not included in the initial mass

balance equations based on GPR. Biomass synthesis reaction is a simplification of complex biosynthetic processes, such as protein, RNA, DNA, lipid, cell wall formation, by combining all metabolic precursors, cofactors, and energy requirements in the stoichiometric proportions necessary to produce one gram of CDW with a flux rate corresponding to the specific growth rate μ of the organism. The stoichiometric proportions are derived from experimental biomass composition data (eg., protein, carbohydrate, lipid fractions), elemental analysis, polymer composition (amino acid, nucleotide, fatty acid distribution. This reaction also includes the energetic costs of biosynthesis, captured through ATP, GTP, UTP, and NAD(P)H usage, obtained from chemostat experiments, which together define the growth-associated maintenance (GAM) term (Feist & Palsson, 2010). The maintenance processes that drive metabolic processes of the cell different than biosynthetic, polymerization, or anabolic metabolic reactions - cellular homeostasis (turgor pressure, pH regulation), maintenance of ion gradients, and others - are included in non-growth associated maintenance (NGAM) (Thiele & Palsson, 2010). NGAM is an ATP hydrolysis reaction ($1 \text{ ATP} + 1 \text{ H}_2\text{O} \rightarrow 1 \text{ ADP} + 1 \text{ P}_i + 1 \text{ H}^+$), which in yeast models was set between 0.5-1 mmol (gCDW h) $^{-1}$ (Sánchez et al., 2017). The ATP yield per NADH (P/O ratio) is not defined as a fixed parameter but emerges from the stoichiometric coupling of the electron transport chain and ATP synthase reactions. In modern GEMs such as Yeast8 and ecYeast8, the stoichiometry of ATP synthase was changed to correctly represent the effective P/O ratio that applies to organisms that lack proton-translocating Complex I (Ferguson, 2010; Sánchez et al., 2017). On a final note about constraints, metabolic transients are typically rapid compared to growth rates (Stephanopoulos et al., 1998), making steady-state assumption reasonable.

By predicting fluxes, researchers can find out how nutrients are utilized, how byproducts are generated, which metabolic pathways are active inside the cell. With GEMs, it is possible to quantitatively predict targets for metabolic engineering, manipulate reactions *in silico*, allowing for prediction of a knockout phenotype or heterologous byproducts, guide medium and process optimization through scenario testing under varying oxygen, carbon, or nutrient regimes (Kerkhoven et al., 2014). Due to relative ease of implementation, GEMs are one of the core quantitative tools enabling computational design. Since the first yeast GEM reconstruction of *S. cerevisiae* in 2003, more than 45 metabolic models have been developed for a wide variety of yeast-based cell factories. A landscape of yeast GEMs, their accessibility and usability is available from The FEMS Yeast Research review (Domenzain et al., 2021). However, their accuracy and predictive capacity depend on parametrization, constraints, and validation.

Because genome-scale models are quantitative frameworks, their predictive power relies on accurate experimental parameterization and validation. Experimental data are used to constrain and calibrate (fitting bounds/parameters) model behavior. The FEMS Yeast Research review (Domenzain et al., 2021) discusses that many yeast GEMs lack consistent, reproducible validation (testing on held-out observables/conditions) across species and indicates the type of biological and computational methods used when available. Genetic perturbations, including gene deletions, overexpression, and knockdowns, isotope labeling (^{13}C) and comparing predicted and measured phenotypes, including growth rates, yields, flux distributions, remain as fundamental experimental strategies for validating flux predictions from genome-scale metabolic models. By systematically engineering cells and measuring resulting phenotypes – such as growth rate, substrate uptake, product yield, or intracellular fluxes – researchers can test whether model predictions reflect real metabolic responses. Modern tools like CRISPR/Cas9 and

CRISPRi enable precise and multiplex genome editing, while omics-based measurements provide quantitative validation of flux distribution and proteome allocation. Together, these validation steps establish the quantitative reliability of the ecGEMs across multiple observable levels - from global phenotypes to proteome allocation. Such comprehensive evaluation forms the basis for subsequent computational design and interpretation of experimental results presented in the following chapters.

1.5.1 Enzyme-constrained models

The prediction of biologically meaningful metabolic flux values is the major challenge of GEMs, as a result of all assumptions made upon the construction. Also, not always experimental yeast data are readily available for a wide variety of organisms across different conditions. Enzyme-constrained genome-scale models (ecGEMs) are built on the principle that any metabolic flux, apart from the constraints already discussed earlier, has a natural biological constraint equal to the enzyme's concentration [E] multiplied by the enzyme's turnover number (k_{cat}). In the ecGEMs, this constraint, derived through Michaelis-Menten equation, is defined and implemented as the maximum rate of enzymatic reaction (v_{max}) that the metabolic flux (v) cannot exceed (Sánchez et al., 2017) (**Equation 1.4**):

$$v \leq k_{cat}[E] \quad (1.4)$$

This way, GEMs are used as scaffolds for proteomic data integration and interpretation (Sánchez & Nielsen, 2015), and can execute simulations without experimental data on substrate usage. The very first enzyme-constrained model was developed in yeast *Saccharomyces cerevisiae* in 2017 (Sánchez et al., 2017) and was capable of capturing the Crabtree effect.

1.6 Aims and significance

A central challenge in modern biotechnology is to understand and predict how cellular metabolism responds to genetic and environmental perturbations. Achieving this requires computational models capable of quantitatively describing enzyme- and resource-limited metabolism. Enzyme-constrained genome-scale models (ecGEMs) provide a powerful framework for this purpose, linking enzyme capacity to metabolic flux and thus improving prediction accuracy beyond traditional constraint-based approaches. Such models are critical not only for advancing fundamental understanding of metabolic organization but also for accelerating design-build-test-learn cycles in metabolic engineering, reducing development costs and time-to-market for bioprocess innovations.

The overarching aim of this thesis was to develop, apply, and critically evaluate enzyme-constrained genome-scale models for predicting yeast metabolism across both model and non-model yeasts. Specifically, the aims were:

- (i) to parameterize and validate ecGEMs of *Saccharomyces cerevisiae* and *Rhodotorula toruloides* as tools for quantitative metabolism studies using quantitative proteomics, exchange flux measurements, and targeted genetic perturbations;
- (ii) to assess their predictive capacity in reproducing experimental phenotypes and proteome allocation patterns;
- (iii) to identify biological and methodological factors that limit their accuracy.

2 Materials and methods

To parametrize and validate ecGEM of a model yeast *S. cerevisiae* (**Study 1**), growth experiments accompanied with absolute proteome quantification were performed in both engineered (R,R)-2,3-butanediol (23BD) producer strain and the wild type strain. Similarly, growth characterisation and absolute proteome quantification were performed in the wild type strain of *R. toruloides* (**Study 2**). Moreover, these experiments were performed on three various carbon sources - glucose, xylose and acetate. After parametrization of the ecGEM, its predictions were validated by engineering gene knockout strains and characterizing their growth on the same carbon sources. The corresponding publications to these two studies are outlined in **Table 2**. The materials and methods refer to Study I or Study II.

Table 2. Studies and related publications included in this thesis.

Study	Topic	Deliverables	Reference
1	Evaluating and validating enzyme-constrained genome-scale modeling of anaerobic cultivation in <i>S. cerevisiae</i>	(i) Engineering of a 23BD producer strain; (ii) Growth characterisation and absolute proteome quantification of the engineered and wild type strains; (iii) <i>In silico</i> flux and proteome prediction using total-protein ecGEM; (iv) Integration of absolute proteome abundances to the ecGEM and <i>in silico</i> flux prediction.	Publication I
2	Predicting and validating central carbon metabolism in the oleaginous yeast <i>R. toruloides</i>	(i) Growth characterization and absolute proteome quantification of a wild type strain on different carbon substrates; (ii) Development of condition-specific proteomics-constrained ecGEMs; (iii) <i>In silico</i> flux predictions.	Publication II
2	Predicting and validating central carbon metabolism in the oleaginous yeast <i>R. toruloides</i>	(i) Engineering of different gene knockout strains; (ii) Growth characterization of gene knockout strains on different carbon sources.	Publication III

2.1 Strains and media

The *S. cerevisiae* and *R. toruloides* strains relevant to this thesis are listed in **Table 3**. Of the *S. cerevisiae* strains, only GSY013 and GSY014 were used to generate results presented here. All yeast strains were cultivated in chemically defined media. The base medium was prepared according to Verduyn et al., previously optimized for yeast biomass formation under glucose-limited conditions (Verduyn et al., 1992). Auxotrophic markers and a Tween-ergosterol-ethanol supplement were added to support growth of auxotrophic strains under anaerobic conditions (**Study 1**). In **Study 1**, standard starting glucose concentrations were used as a carbon source. In **Study 2**, non-standard starting concentrations of glucose (63.6 g/L), xylose (70 g/L), and acetate (20 g/L) were used in actively controlled bioreactor cultivations. For lipid production in *R. toruloides*, the nitrogen concentration was reduced to achieve a starting carbon-to-nitrogen (C/N) ratio of 69 or 80 (mol/mol) (**Study 2**). Only *R. toruloides* CCT 7815 cultures on glucose used urea instead of ammonium sulfate as the nitrogen source (**Study 2**). Routine yeast growth was performed in YPD medium (10 g L⁻¹ yeast extract, 20 g L⁻¹ peptone, 20 g L⁻¹ glucose).

For plasmid propagation and routine cloning, *Escherichia coli* strains DH5 α (**Study 1**) and TOP10 (**Study 2**) were grown in LB medium containing 5 g L⁻¹ yeast extract, 10 g L⁻¹ peptone, and 10 g L⁻¹ NaCl. For plasmid propagation intended for *R. toruloides*, kanamycin was added to LB medium (**Study 2**).

Table 3. *S. cerevisiae* and *R. toruloides* strains relevant to this thesis.

Study	Strain name (accession number)	Description of relevant genotype	Parental strain	Reference
Study 1	<i>S. cerevisiae</i> CEN.PK113-7D			(Entian & Kötter, 2007)
Study 1	IMX672 (Y40595)	<i>MATa ura3-52 trp1-289 leu2-3112 his3Δ can1Δ::cas9-natNT2</i>	<i>S. cerevisiae</i> CEN.PK2-1C	(Mans et al., 2015)
Study 1	GSY006	pROSU-mth1-pdc5+6 <i>trp1-289 leu2-3112 his3Δ pdc6::pTEF1-ACH1; MTH1ΔT; pdc5</i>	IMX672	Publication I
Study 1	GSY008	pROSH-pdc1; <i>trp1-289 leu2-3112 pdc6::pTEF1-ACH1; MTH1ΔT; pdc5Δ; pdc1::(alsD-tADH2 pTDH3-BDH1 pFBA1-budA)</i>	GSY006	Publication I
Study 1	GSY013	pTHUL	IMX672	Publication I

Study 1	GSY014	pTHUL; <i>pdc6::pTEF1-ACH1; MTH1ΔT; pdc5Δ; pdc1::(alsD-tADH2 pTDH3-BDH pFBA1-budA)</i>	GSY008	Publication I
Study 2	<i>R. toruloides</i> CCT 7815		<i>R. toruloides</i> CCT 0783	(Bonturi et al., 2017)
Study 2	<i>R. toruloides</i> NBRC 0880			(Zhang et al., 2016)
Study 2	ΔACL	pPBO.202-9725(1) (random integration)	<i>R. toruloides</i> NBRC 0880	Publication III
Study 2	ΔPK	pPBO.202-13382(3) (random integration)	<i>R. toruloides</i> NBRC 0880	Publication III
Study 2	ΔcMAE	pPBO.202-12761(1) (random integration)	<i>R. toruloides</i> NBRC 0880	Publication III

2.2 Plasmid construction and strain engineering

Yeast transformation was performed using lithium acetate/PEG-mediated chemical transformation method as described by Mans et al. (Mans et al., 2018) (*S. cerevisiae*) and **Publication III** (*R. toruloides*). Colonies of the resultant strains were screened by PCR amplification of relevant sites and Sanger sequencing (full details are available from **Publications I and III**).

In **Publication I**, plasmids were constructed by Gibson assembly (Gibson et al., 2009) and transformed into *E. coli* DH5 α for propagation. In **Publication III**, custom guide RNA sequences replaced the GFP cassette in plasmid pPBO.202 via a Golden Gate – style reaction, using annealed oligonucleotides with Bsal-compatible overhangs in a one-pot digestion – ligation with T4 DNA ligase, followed by transformation into *E. coli* TOP10. Selected colonies were screened by PCR and confirmed by Sanger sequencing of the relevant regions.

Plasmids used in this thesis are listed in **Table 4**. Details of plasmid and PCR DNA purification kits, polymerases, as well as sequences of primers and gRNAs are provided in the respective publications (**Table 2**).

Table 4. Plasmids relevant to this thesis. The number in the brackets refers to which gRNA variant in a row was successful to generate the gene knockout. Gene symbols in *italics* denote coding sequences; heterologous origins (e.g., *B. subtilis* *alsS*, *K. pneumoniae* *budA*) are specified in text. *SpCas9*, derived from *Streptococcus pyogenes*, is indicated explicitly.

Study	Plasmid name	Relevant characteristics	Reference
Study 1	pROS10	2µm ampR <i>URA3</i> gRNA-CAN1 gRNA-ADE2	(Mans et al., 2015)
Study 1	pROS16	2µm ampR <i>HIS3</i> gRNA-CAN1 gRNA-ADE2	(Mans et al., 2015)
Study 1	pBC414	<i>CEN6/ARS4</i> chlR <i>TRP1</i>	(Frazer & O'Keefe, 2007)
Study 1	pROSU-mth1-pdc5+6	2µm ampR <i>URA3</i> gRNA- <i>MTH1</i> gRNA-[<i>PDC5</i> and <i>PDC6</i>]	Publication I
Study 1	pROSH-pdc1	2µm ampR <i>HIS3</i> gRNA- <i>PDC1</i> [2X]	Publication I
Study 1	pTHUL	<i>CEN6/ARS4</i> chlR <i>TRP1</i> <i>LEU2</i> <i>HIS3</i> <i>URA3</i>	Publication I
Study 1	pYGS011	ampR <i>budA</i> -tADH2	Publication I
Study 1	pYGS012	ampR <i>pTDH3-BDH1</i>	Publication I
Study 1	pYGS013	ampR <i>pFBA1-alsS</i>	Publication I
Study 1	pYGS017	ampR <i>pTEF1-ACH1</i>	Publication I
Study 2	pPBO.202	KanR G418R eGFP <i>SpCas9</i>	(Otoupal et al., 2019)
Study 2	pPBO.202-9725(1)	KanR G418R gRNA- <i>ACL</i> <i>SpCas9</i>	Publication III
Study 2	pPBO.202-13382(3)	KanR G418R gRNA- <i>PK</i> <i>SpCas9</i>	Publication III
Study 2	pPBO.202-12761(1)	KanR G418R gRNA- <i>cMAE</i> <i>SpCas9</i>	Publication III

2.3 Laboratory evolution

Laboratory evolution was conducted by serially culturing strain GSY014 in 5 mL of chemically defined medium under anaerobic conditions for 35 inoculation cycles (≈200 generations). The detailed procedure is described in **Publication I**.

2.4 Cultivations

2.4.1 Batch cultivations in actively controlled bioreactors

Batch cultivations were performed in 1 L stainless-steel (SS) bioreactor system (**Study 1**) and 1-L glass bioreactors (**Study 2**) equipped with automated feedback control of aeration, temperature, stirring, and pH, ensuring tightly regulated growth conditions. Outlet gas composition was continuously monitored using online gas analyzers, and antifoam was added as needed to prevent foam formation. Anaerobic conditions in **Study 1** were maintained by applying a continuous nitrogen stream in the reactor headspace. In **Study 2**, cell turbidity was monitored online using a biomass sensor. Full cultivation parameters are provided in the respective publications.

Seed cultures were prepared in standard chemically defined media in shaker flasks. For anaerobic cultivations (**Study 1**), serum bottles were deoxygenated by alternating cycles of vacuum and nitrogen gas. For *R. toruloides* aerobic cultures (**Study 2**), pre-seed cultures were grown in YPD, followed by seed cultures in defined media containing non-standard concentrations glucose (18.2 g L⁻¹), xylose (20 g L⁻¹), or acetate (20 g L⁻¹) at a C/N ratio of 8.8.

Samples for HPLC analysis were taken at each sampling point. In **Study 1**, optical density (OD) was measured at each point, and cell dry weight (CDW) was determined twice per batch. In **Study 2**, CDW was measured three times in xylose and up to six times in glucose cultures. Proteomic samples were collected at the penultimate sampling point during exponential growth – prior to carbon (**Study 1**) or nitrogen (**Study 2**) depletion – and an additional sample was taken during nitrogen limitation in **Study 2**. Samples for lipidomics were collected at the final sampling point, when the respective nutrient was depleted. Detailed sampling protocols and time points are described in the original publications.

2.4.2 Batch cultivations in sensor-monitored Falcon tube bioreactors

In **Study 2**, batch cultivations were also performed in a multi-channel Falcon tube bioreactor system equipped with a proprietary reverse-spin mixing mechanism and non-invasive optical sensors for real-time monitoring of optical density (OD), pH, and dissolved oxygen (DO). Non-standard initial concentrations of glucose (20 g L⁻¹), xylose (20 g L⁻¹), and acetate (10 g L⁻¹) were used, while maintaining a starting C/N ratio of 80. The pH was adjusted using K₂HPO₄ (acted as a buffer system in glucose/xylose media) or KOH (to increase the starting pH in the presence of acetate) but not actively controlled during cultivation. Full cultivation parameters are available in the respective publication.

Seed cultures were prepared by overnight growth in liquid YPD medium in shaker flasks. Samples for OD and HPLC analysis were taken at each sample point. Real-time OD, pH, and DO data were continuously logged throughout cultivation, and OD readings were converted to CDW (g L⁻¹) for data presentation.

2.5 Analytical methods

Extracellular metabolites were quantified using high-performance liquid chromatography (HPLC) systems equipped with an ion-exclusion column and a refractive index detector (RID). In **Study 1**, trace amounts of organic acids were detected at 210 nm using a photodiode array detector (PDA); since these represented < 1 % of the total carbon and degree-of-reduction balance, they were excluded from modeling and from the presented results. In **Study 2**, stereoselective analysis of polyols was performed using a chiral

stationary phase column with RID detection. Detailed chromatographic conditions are provided in the respective publications.

Optical density was measured spectrophotometrically at 600 nm (OD_{600}). Cell dry weight (CDW) was determined gravimetrically, and a calibration curve was established by linear regression between OD_{600} and CDW to estimate biomass across all sampling points.

Total protein content was quantified from the same frozen cell pellets used for proteomic analysis, using a modified Lowry method (**Study 1**) or the bicinchoninic acid (BCA) assay (**Study 2**). Bovine serum albumin (BSA) served as the standard in both cases. Absorbance was measured at 510 nm, and protein concentrations were determined by linear interpolation from the BSA standard curve.

2.6 Absolute proteomics

2.6.1 Sample preparation for proteomics

Sampling of cell pellets for proteomic analysis was performed as described in **Publication I** and completed within 3 min to minimize physiological changes. Samples were shipped on dry ice to the Proteomics Core Facility, University of Tartu, for processing and analysis. Cell pellets were lysed, and total protein yield on biomass ($mg\ g^{-1}$ CDW) was determined using the bicinchoninic acid (BCA) assay. The lysates were then processed into peptides for LC-MS/MS analysis as described previously (Hughes et al., 2019; Sánchez et al., 2021). In **Study 2**, proteome samples were mixed with a heavy-labeled *Rhodotorula toruloides* internal standard grown in defined medium supplemented with ^{15}N - and ^{13}C -labeled lysine (also known as SILAC, *Stable Isotope Labeling by Amino acids in Cell culture*).

2.6.2 LC-MS/MS analysis

Peptides were analyzed on an Orbitrap mass spectrometer (Thermo Fisher Scientific). In **Study 1**, the instrument was operated in data-independent acquisition (DIA) mode, whereas in **Study 2**, data were acquired using a data-dependent acquisition (DDA) setup. Detailed descriptions of the mass spectrometer components and key data acquisition parameters are provided in the respective publications and in Sanchez *et al.* (Sánchez et al., 2021).

2.6.3 Raw MS data analysis

Mass spectrometric raw data were processed using DIA-NN (Demichev et al., 2020) for **Study 1** and MaxQuant (Tyanova et al., 2016) for **Study 2**. Database searches were performed against the UniProt (The UniProt Consortium et al., 2023) reference proteomes of *Saccharomyces cerevisiae* S288C, supplemented with *Bacillus subtilis* alsS and *Klebsiella aerogenes* budA sequences (**Study 1**), or *Rhodotorula toruloides* NP11 (**Study 2**). Details of search settings, parameters, and data processing methods that deviated from software defaults are provided in **Publication I** and in Sanchez *et al.* (Sánchez et al., 2021).

2.6.4 Quantitative and statistical analysis

Absolute protein quantification was performed using the total protein approach (Sánchez et al., 2021), which assumes proportionality between measured MS signal intensity and total protein yield on biomass. All quantitative analyses were performed with proteome-normalized proteomics data ($\mu g\ g_{protein}^{-1}$). For differential expression

analysis, *p*-values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995a). Additional analyses included principal component analysis (PCA) using the ClustVisweb tool (Metsalu & Vilo, 2015) and gene set analysis with the Piano package (Väremo et al., 2013). In **Study 2**, the rate of protein synthesis per ribosome (also referred to as ribosome efficiency or protein translation rate) was calculated as described in **Table 5**.

Table 5. Translation rate calculation example in **Study 2** using absolute proteomics data of batch cultivations in chemically defined media containing glucose (G), xylose (X), or acetate (A) as the carbon source during exponential growth (exp) and nitrogen-limited (Nlim) phases. Reproduced from Rekēna et al., *PLOS Comput. Biol.* 19(4):e1011009, 2023, under a CC BY licence.

Parameter	Gexp	GNlim	Xexp	XNlim	Aexp	ANlim
ribo_mmol_gDCW (median)	4.21E-05	1.30E-05	2.76E-05	9.82E-06	2.25E-05	9.87E-06
all_protein_mmol_gDCW (recov-adj)	7.65E-03	3.21E-03	5.38E-03	2.88E-03	5.41E-03	3.32E-03
av. AA in euk. protein				472		
all_prot_mmolAA_gDCW ^a	3.61E+00	1.52E+00	2.54E+00	1.36E+00	2.55E+00	1.57E+00
ratio(mmolAA/ribo_mmol) ^b	8.59E+04	1.17E+05	9.20E+04	1.38E+05	1.14E+05	1.59E+05
doubling time (h) ^c	3.63	36.48	12.84	38.51	9.50	57.76
doubling time (s) ^d	13064.55	131333.15	46209.81	138629.43	34182.60	207944.15
translation rate (1/s) ^e	6.57	0.89	1.99	1.00	3.32	0.76
μ (1/h)	0.19	0.021	0.054	0.021	0.073	0.012

^aall_protein_mmol_gDCW * av. AA in euk. protein

^ball_prot_mmolAA_gDCW / ribo_mmol_gDCW (median)

^cln(2) / μ (1/h)

^ddoubling time (h) * 3600

^eratio(mmolAA/ribo_mmol) / doubling time (s)

2.7 Lipidomics

Sampling of cell pellets for lipidomics analysis was performed as described in **Publication III**. Samples were shipped by regular post to the University of Tartu (for GC-MS analysis) and to the Estonian University of Life Sciences (for GC-FID analysis). Lipid quantification via fatty acid methyl ester (FAME) analysis was carried out using gas chromatography (GC)-based methods.

In **Publication II**, FAMEs were analyzed by gas chromatography–mass spectrometry (GC-MS). After chromatographic separation, compounds were identified by comparing retention times and mass spectra with reference standards and entries from the National Institute of Standards and Technology (NIST, USA) Mass Spectral Library and quantified from extracted ion chromatograms using calibration curves constructed for each methylated fatty acid from seven calibration standards prepared with reference compounds and the internal standard (hexadecane). Fatty acid extraction and derivatization were performed according to Tammekivi *et al.* (Tammekivi et al., 2021); key methodological details and GC-MS parameters are provided in **Publication II**.

In **Publication III**, FAMEs were analyzed by GC-FID. Peaks were identified based on retention times of reference standards and quantified using flame ionization detection. A one-step fatty acid extraction and derivatization procedure was applied using

heptanoic acid as the internal standard, as described by (Sukhija & Palmquist, 1988). Key instrument operating parameters are provided in **Publication III**.

Total lipid yield on biomass was calculated as the sum of individual fatty acid masses, expressed as milligram of lipid per gram of cell dry weight ($\text{mg g}^{-1} \text{CDW}$), as described in **Publication III**.

2.8 Enzyme-constrained genome-scale models

In **Study 1**, the recently published total-protein *Saccharomyces cerevisiae* ecGEM (F. Li et al., 2022) was chosen as a starting point for modeling GSY014. The model was adapted to anaerobic growth conditions, fitted to experimentally measured exchange fluxes of GSY013 and adapted to the genotype of GSY014 as described in **Publication I**.

In **Study 2**, the conventional GEM of *Rhodotorula toruloides* (Tiukova, Prigent, et al., 2019) served as the starting point for modeling CCT 7815, as no enzyme-constrained version (ecGEM) for this species was available. The base model was converted into a total-protein ecGEM fitted to experimentally measured exchange fluxes of CCT 7815 through the GECKO toolbox version 2.0.2 (Domenzain et al., 2022), using *R. toruloides* NP11 enzyme identifiers from UniProt. Individual protein constraints that limited growth rate predictions were flexibilized by adjusting their k_{cat} values (and therefore the upper bounds of enzyme usage), initially retrieved from the BRENDA database via the algorithm described by Domenzain *et al.* (Domenzain et al., 2022). The detailed procedure for flexibilization is described in **Publication II**.

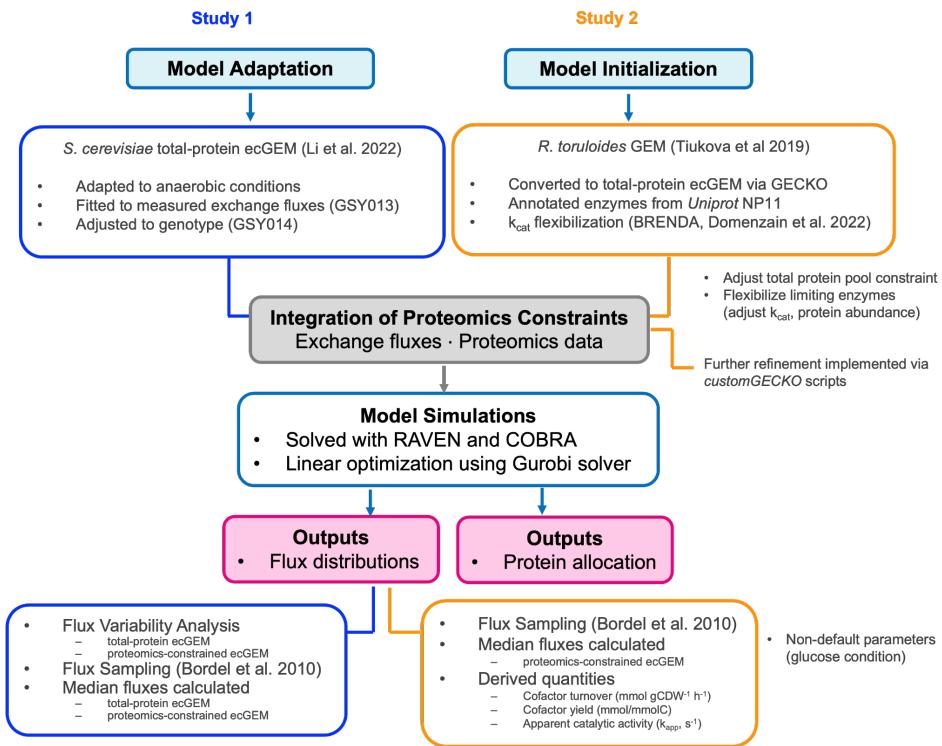
The total-protein ecGEMs were subsequently constrained by measured exchange fluxes and biomass-normalized proteomics data ($\mu\text{g gCDW}^{-1}$) using the GECKO pipeline, resulting in proteomics-constrained ecGEMs. Where necessary, the total protein pool pseudoreaction was adjusted to match measured fluxes. In addition, individual enzyme constraints that prevented the model from reaching the observed growth rate were flexibilized by adjusting their protein abundances (upper bound, ub). In **Study 2**, further modifications were implemented as described in **Publication II**, via user-defined scripts integrated into the GECKO pipeline.

Model simulations and alterations were carried out using the RAVEN toolbox version 2.0 (Wang et al., 2018) (both **studies**) and COBRA toolbox version 3.0 (Heirendt et al., 2019) (**Study 1**). Both frameworks solved the linear optimization problem underlying Flux Balance Analysis (FBA) using the Gurobi solver. In total-protein ecGEMs (**Study 1**), solving this linear problem provided not only predicted distribution of flux values but also protein allocation patterns, which can be directly obtained from the model's enzyme pool constraints. Thus, FBA on a total-protein or proteomics-constrained ecGEM inherently predicts individual enzyme allocation alongside metabolic fluxes.

To assess the variability of predicted flux values, flux variability analysis (FVA) and flux sampling (Bordel et al., 2010) were performed on the total-protein ecGEM (**Study 1**) and proteomics-constrained ecGEMs (both **studies**). In **Study 1**, default simulation parameters were used. In **Study 2**, only flux sampling was performed. For the glucose condition, simulated gas-exchange values were used instead of measured values to constrain sampling; other parameters followed those described in **Publication II**.

Flux sampling, as implemented by Bordel et al. (Bordel et al., 2010), explores the convex feasible flux space defined by the stoichiometric constraints. Median flux values were calculated from the sampling results as described in **Publication II**. Cofactor turnover

rates ($\text{mmol gCDW}^{-1} \text{h}^{-1}$), cofactor yields ($\text{mmol cofactor mmol carbon}^{-1}$), and apparent catalytic activities (k_{app}) (s^{-1}) were also calculated following the procedures in **Publication II**. A schematic overview of the modeling workflow is presented in **Figure 1**.



Study 2

Model Initialization

R. toruloides GEM (Tiukova et al 2019)

- Converted to total-protein ecGEM via GECKO
- Annotated enzymes from *Uniprot* NP11
- k_{cat} flexibilization (BRENDA, Domenzain et al. 2022)

- Adjust total protein pool constraint
- Flexibilize limiting enzymes (adjust k_{cat} , protein abundance)

Further refinement implemented via *customGECKO* scripts

Figure 1. Schematic overview of model reconstruction and simulation workflow. In **Study 1**, the total-protein *S. cerevisiae* ecGEM was adapted to anaerobic conditions and fitted to experimental data. In **Study 2**, the *R. toruloides* GEM was reconstructed into a total-protein ecGEM via the GECKO pipeline. Both models were subsequently constrained with proteomics and exchange flux data, simulated using FBA (Gurobi solver) to obtain flux variability (by flux variability analysis and flux sampling), protein allocation, and derived kinetic parameters.

2.9 Statistical analysis

Statistical analysis of cultivation parameters was performed using GraphPad Prism 9.5.1 (GraphPad Software Inc., San Diego, Ca, USA). Statistical significance was calculated using one-way ANOVA (Analysis of Variance) at 0.05 significance level. *p*-values were adjusted (apval) for multiplicity following Dunnett multiple comparison testing against the wild type strain.

3 Results and discussion

Although all results have been published in the respective publications, in this chapter I will focus on the predictive power of constraint-based modeling by interpreting the predicted fluxes and derived quantities presented in these studies. Some of the discussion points have been addressed briefly in the original publications but are elaborated here in greater depth. In the following sections, I revisit the relevant quantitative results and methodological aspects – some overlapping with the published material and others presented here for the first time – to provide an integrated, modeling-centered interpretation.

3.1 Study 1: Evaluating and validating enzyme-constrained genome-scale modeling of anaerobic cultivation in *S. cerevisiae* (Publication I)

Anaerobic redox-neutral, ATP-yielding pathways from substrate to product represent the stoichiometrically most efficient bioconversion routes, maximizing carbon yield while conserving energy. This principle underlies classical fermentations – such as alcoholic or homolactic – but also applies to engineered heterologous pathways when thermodynamically feasible (Cueto-Rojas et al., 2015). Such pathways offer major bioprocessing advantages: they minimize substrate costs, reduce capital investment by eliminating aeration, lower cooling requirements, and enable coupling of product formation to growth. This growth coupling facilitates adaptive laboratory evolution (ALE) for improved performance and allows efficient cell recycling under production conditions. Given these properties, anaerobic, ATP-yielding, redox-neutral product formation provides a stringent and informative test case for enzyme-constrained genome-scale models (ecGEMs), which explicitly link catalytic capacity to metabolic flux. In this work, we focused on the heterologous production of (R,R)-2,3-butanediol (23BD) in *Saccharomyces cerevisiae* as an industrially relevant model system. 23BD is a versatile platform chemical used in solvents, synthetic rubber, plastics, and fuels (Celińska & Grajek, 2009; Garg & Jain, 1995). The pathway – from two molecules of pyruvate via acetolactate synthase (AlsS), acetolactate decarboxylase (BudA), and 2,3-butanediol dehydrogenase (Bdh1) – has been established previously under aerobic conditions (Kim et al., 2017; Ng et al., 2012), demonstrating high titers at high rate, meaning that confounding effects from product toxicity or export should not be a problem. To enable anaerobic operation, all pyruvate decarboxylase (PDC1, PDC5, and PDC6) genes were deleted to block competing ethanol fermentation, while MTH1 Δ T was introduced to restore glucose uptake regulation (Oud et al., 2012), and ACH1 was overexpressed to alleviate cytosolic acetyl-CoA deficiency (Y. Chen et al., 2015). Because glycolytic 23BD formation generates one surplus NADH per glucose, and *S. cerevisiae* reoxidizes NADH mainly through glycerol production under anaerobic conditions (van Dijken & Scheffers, 1986), the pathway intrinsically couples 23BD and glycerol formation in a redox-balanced, near 1:1 ratio.

In **Study I**, we experimentally measured exchange fluxes, growth rates in the reference strain, and parameterized and validated the enzyme-constrained GEM (ecGEM) to assess how catalytic constraints redistribute proteome investment and limit fluxes. The evaluation proceeded in two stages. First, the ecGEM was parameterized and fitted using data from the reference strain to capture *aerobic* growth physiology. The calibrated model was then applied to predict *anaerobic* growth and metabolite exchange rates of the engineered 2,3-butanediol–glycerol co-producing strain, with predictions benchmarked

against experimental measurements obtained before and after 200 generations of laboratory evolution. Then model predictions were compared to quantitative proteomics. The analysis proceeds from coarse simulations using only a total-protein constraint to proteomics-constrained models integrating 750 individual enzyme abundances, revealing how enzyme constraints reduce flux variability, reshape glycolytic investment, and expose discrepancies between *in vitro* kinetics and *in vivo* proteome behavior. Together, these results illuminate the energetic and proteomic trade-offs underlying engineered anaerobic fermentation and evaluate the predictive scope of ecGEMs for rational strain design.

3.1.1 Calibration of a total-protein ecGEM enables prediction of 2,3-butanediol and glycerol co-production

To fit an enzyme-constrained model (ecGEM) of *S. cerevisiae* engineered for (R,R)-2,3-butanediol (23BD) production, we first parameterized the reference strain GSY013 in actively controlled anaerobic batch cultures (**Figure S1A**). The strain produced biomass (89 ± 6 mg/g glucose), carbon dioxide (380 ± 100 mg/g glucose), ethanol (340 ± 1 mg/g glucose), and glycerol (101 ± 2 mg/g glucose) (**Table 7**), values consistent with previously reported reference strains under similar conditions (Papapetridis et al., 2018). The maximum specific growth rate was 0.36 ± 0.02 h⁻¹, supported by a specific glucose uptake rate of -23 ± 2 mmol/gDCW/h, with other uptake and secretion rates given in **Table 7**.

As a starting point, we used the most recently published ecGEM of *S. cerevisiae*, which constrains fluxes by the total enzyme pool (F. Li et al., 2022). To align model predictions with experimentally observed data, two adjustments were required: (i) the degree of reduction of biomass was lowered by 3 mmol/gDCW, and (ii) the upper bound of the protein pool pseudoreaction was reduced from the experimentally determined 450 mg/gDCW to an *effective value* of 123 mg/gDCW. This adjustment is a common feature of GECKO-based models, reflecting the fact that not all measured protein mass can be functionally assigned to catalytic activity in the model. Reasons include incomplete enzyme annotation, missing or inaccurate k_{cat} values, and allocation to non-modeled proteins (e.g., regulatory, structural, or stress-related proteins). Thus, the effective pool of usable protein *in silico* is smaller than the total measured proteome. These adjustments ensured that simulated fluxes reproduced the experimentally observed growth rate of the control strain (GSY013). The resulting curated model provided a calibrated baseline for adapting the genotype to the engineered 23BD pathway (for full details see **Publication I**).

Using this adapted coarse total-protein ecGEM, we predicted growth and metabolism of the anaerobic 23BD-glycerol co-producing strain GSY014. With measured boundary conditions (extracellular fluxes and total protein abundance) applied, the model optimized protein allocation and predicted almost equimolar production of 23BD and glycerol (16.8 vs. 19.0 mmol/gDCW/h), accompanied by 34.9 mmol/gDCW/h of CO₂ and no ethanol formation (**Table 7**). Compared to the reference strain, glycerol production increased 4.2-fold. The predicted maximum specific growth rate was 0.175 h⁻¹ with a specific glucose uptake rate of -27.8 mmol/gDCW/h. While the ATP yield of 23BD/glycerol fermentation is \sim 3-fold lower than alcoholic fermentation, the ecGEM predicted an increased glucose uptake rate that partially offset this energy deficit, resulting in a \sim 2-fold, rather than 3-fold, reduction in specific growth (0.175 h⁻¹ vs. 0.36 h⁻¹). Biomass

yield was reduced almost 3-fold (35 vs. 89 ± 6 mg/g glucose), consistent with experimental observations (**Table 7**).

Overall, the need to reduce the protein pool constraint from 450 to 123 mg/gDCW illustrates a central limitation of ecGEMs: while they capture enzyme allocation principles, incomplete annotation and kinetic data necessitate downscaling of the effective proteome to reconcile simulations with reality. Nevertheless, this calibration step ensured a well-fitted reference state, enabling meaningful predictions of the engineered 23BD-glycerol strain.

Table 6: Yields and carbon- and degree of reduction balances observed in *S. cerevisiae* anaerobic bioreactor batch cultivations of the reference strain GSY013 and 23BD-glycerol co-producing strain GSY014. Reproduced from Sjöberg et al., *Metabolic Engineering* 82, 49-59, 2024, under a CC BY licence.

Strain	Yields on glucose (mg (g glucose) $^{-1}$)						Recovery of consumed substrate, based on:	
	CDW	Glycerol	Ethanol	23BD ^a	Acetoin	CO ₂	Carbon	DR ^b
GSY013 ^c	89 \pm 6	101 \pm 2	340 \pm 1	ND ^e	ND ^e	380 \pm 100	93 \pm 7%	91 \pm 1%
GSY014 ^c	30 \pm 2	346 \pm 1	ND ^e	274 \pm 1	7 \pm 1	300 \pm 10	95 \pm 1%	94.5 \pm 0.1%
GSY014E ^d	33 \pm 1	350 \pm 3	ND ^e	273 \pm 1	9 \pm 1	300 \pm 90	97 \pm 7%	95.7 \pm 0.4%
Predicted	35	349	0	302	0	307	NA ^f	NA ^f

a – 2,3-butanediol

b – Degree of reduction as defined by (Heijnen, 1994)

c – Average of 6 independent cultivations \pm standard deviation

d – Average of 2 independent cultivations \pm mean deviation

e – Not detected

f – Not applicable

Table 7: Biomass specific rates observed in *S. cerevisiae* anaerobic bioreactor batch cultivations of the reference strain GSY013 and 23BD-glycerol co-producing strain GSY014. Reproduced from Sjöberg et al., *Metabolic Engineering* 82, 49-59, 2024, under a CC BY licence.

Strain	μ (h $^{-1}$)	Rate of consumption or production (mmol (g CDW) $^{-1}$ h $^{-1}$)					
		Glucose	Glycerol	Ethanol	23BD ^a	Acetoin	CO ₂
GSY013 ^b	0.36 \pm 0.02	-23 \pm 2	4.5 \pm 0.4	31 \pm 2	ND ^d	ND ^d	38 \pm 10
Predicted GSY014	0.175	-27.8	19.0	0	16.8	0	34.9
GSY014 ^b	0.15 \pm 0.01	-29 \pm 1	19.6 \pm 0.5	ND ^d	15.8 \pm 0.4	0.39 \pm 0.03	36 \pm 2
GSY014E ^c	0.16 \pm 0.01	-27 \pm 1	18.6 \pm 0.8	ND ^d	14.8 \pm 0.7	0.48 \pm 0.01	33 \pm 9

a – 2,3-butanediol

b – Average of 6 independent cultivations \pm standard deviation

c – Average of 2 independent cultivations \pm mean deviation

d – Not detected

3.1.2 Predicted and *in vivo* fluxes align in 23BD-glycerol co-producing *S. cerevisiae*

To compare model predictions with experimental physiology, we characterized the engineered (R,R)-2,3-butanediol-glycerol co-producing strain GSY014 in anaerobic batch cultures (**Figure S1B**). This strain was constructed by overexpressing the native *S. cerevisiae* 2,3-butanediol dehydrogenase *Bdh1*, together with heterologous acetolactate decarboxylase *budA* (*Klebsiella aerogenes*) and acetolactate synthase *alsS* (*Bacillus subtilis*) (Choi et al., 2016). To enforce redirection of carbon and redox fluxes into the engineered 23BD pathway, several additional modifications were introduced. All functional alleles of pyruvate decarboxylase (PDC1, PDC5, PDC6) were deleted to abolish ethanol fermentation, thereby preventing competition for pyruvate and NADH. Because Pdc-deficient strains display impaired glucose uptake due to loss of pyruvate-acetaldehyde-ethanol cycling, the MTH1 Δ T allele was introduced as previously described (Oud et al., 2012) to alleviate glucose sensitivity and restore efficient hexose uptake. Finally, ACH1 (acetyl-CoA hydrolase) was overexpressed to mitigate cytosolic acetyl-CoA deficiency expected in the absence of PDC-mediated acetaldehyde supply (Y. Chen et al., 2015). Together, these modifications created a metabolic background in which pyruvate could be efficiently redirected into the 23BD pathway, while maintaining redox and acetyl-CoA balance.

The carbon balance of GSY014 cultures closed well ($95\% \pm 1\%$) when calculated by both degree of reduction and carbon balances (**Table 6**). Ethanol production was fully abolished and replaced by near-equimolar production of 23BD and glycerol (15.8 ± 0.4 vs. 19.6 ± 0.5 mmol gDCW $^{-1}$ h $^{-1}$), consistent with ecGEM predictions (**Table 6**). The increase in glycerol production matched predictions closely, while a small but measurable acetoin byproduct (0.39 ± 0.03 mmol gDCW $^{-1}$ h $^{-1}$) was observed experimentally but not predicted by the model. The biomass yield on glucose decreased ~ 3 -fold compared to the reference strain (30 ± 2 vs. 89 ± 6 mg g $^{-1}$ glucose), in line with the predicted yield (35 mg g $^{-1}$ glucose) (**Table 6**). Growth rate was 0.15 ± 0.01 h $^{-1}$, slightly below the prediction (0.175 h $^{-1}$) but not as strongly reduced as expected from the threefold lower ATP yield ($\frac{1}{3}$ ATP mol $^{-1}$ glucose vs. 2 ATP mol $^{-1}$ glucose in ethanol fermentation). Instead, the higher specific glucose uptake rate (-29 ± 1 mmol gDCW $^{-1}$ h $^{-1}$, slightly above predictions) compensated for reduced ATP yield, illustrating how metabolic capacity constraints limit the full translation of thermodynamic penalties into growth rate reductions.

Overall, the ecGEM successfully predicted the key trade-offs in 23BD-glycerol production: (i) elimination of ethanol production, (ii) redirection of flux into 23BD and glycerol, and (iii) reduced biomass yield due to ATP limitation. Importantly, this study highlighted the thermodynamic and energetic limits of this anaerobic pathway, with the Gibbs free energy change shifting from -215 to -147 kJ mol $^{-1}$ glucose and ATP yield dropping threefold.

To probe whether the predicted maximum growth rate could be achieved or improved through natural adaptation, GSY014 was evolved for ~ 200 generations in anaerobic batch serum cultures. However, no significant differences were observed between the evolved and parental strains, except for a modest increase in acetoin production (**Table 6**). This outcome suggests that the engineered pathway is already close to its physiological optimum under the given conditions, with ATP yield and redox constraints posing hard biophysical limits that cannot be easily overcome by short-term laboratory evolution. Excretion of acetoin suggests that there is room for (evolutionary) improvement but

potentially requires much longer time spans. In this context, ecGEMs provide crucial value by distinguishing between hard limits, such as thermodynamics and ATP yield that cannot be overcome by evolution, and soft limits, such as enzyme allocation or regulation, that evolution may eventually optimize. This allows more realistic expectations about when ALE is a viable strategy and when it will likely fail.

3.1.3 Proteome allocation: alignment and divergence between ecGEM predictions and *in vivo* data

To study proteome allocation, we quantified 3797 absolute protein abundances using DIA (Demichev et al., 2020) and total protein quantification (Sánchez et al., 2021). Both engineered and reference strains contained 450 mg protein/gDCW. Principal component analysis of three replicates confirmed clear separation between strains (**Figure S2B**). In the engineered strain, 172 protein abundances were significantly changed (ug/g_protein, $|\log_2\text{FC}| > 1.5$, BH-adjusted $p < 0.05$; **Figure S2A**). Functional group analysis (Proteomap (Liebermeister et al., 2014), gene set analysis (Väremo et al., 2013)) revealed clear trade-offs consistent with the shift from aerobic fermentation to anaerobic cultivation. Ribosomal proteins decreased from 25.5% to 18.5% of the proteome, while glycolysis expanded from 28.7% to 43.5%, becoming the largest functional group (**Figure S3**). At the same time, biosynthesis of cofactors, amino acids, and lipids was significantly upregulated ($\text{apval} < 0.05$), whereas cell wall and glycogen metabolism were strongly downregulated (**Table S1**). These global proteome reallocations show how *S. cerevisiae* balances protein investment between growth machinery and metabolic functions under redox-constrained, anaerobic conditions.

To evaluate predictive capacity, we revisited the coarse total-protein ecGEM simulations constrained by exchange fluxes and total protein abundance. While those runs were first analyzed for predicted exchange fluxes, here we focus on the predicted protein allocation. Both model and experiment agreed on strong glycolytic upregulation: upper glycolysis increased by 30–70% *in silico* and up to 43.5% *in vivo* (**Figure 2**), consistent with higher pyruvate demand for the 23BD pathway. The shift toward higher glycolytic investment, predicted by the ecGEM and confirmed experimentally, underscores the remarkable flexibility of *S. cerevisiae* metabolism in supporting redox and precursor supply under altered fermentation pathways – an insight that conventional GEMs cannot provide, as they do not account for proteome allocation. However, other predictions diverged. The ecGEM suggested a ~5% downregulation of lower glycolysis, while measured proteomics showed stable or slightly increased allocation. Even more strikingly, the model predicted downregulation of ~50% of the growth rate of biosynthetic enzymes in the engineered strain, whereas the proteome increased from 10.7% to 13.8%. These differences show that the ecGEM tends to minimize biosynthetic investment for efficiency, whereas cells appear to maintain or even increase biosynthetic capacity, likely to buffer against translational and metabolic stress.

Discrepancies were also apparent at the enzyme-family level. The ecGEM predicted high allocation to glycerol 3-phosphate dehydrogenases (Gpd) and phosphatases (Gpp) as the main NADH sink, but measured abundances were far lower. Conversely, despite complete PDC deletion, proteomics revealed strong upregulation of acetaldehyde and alcohol dehydrogenases (Ald), which the model predicted to be absent. This points to an ethanol/acetaldehyde cycle acting as a redox overflow valve *in vivo* – activated when NADH reoxidation shifts from alcohol fermentation to glycerol/2,3BD dehydrogenases. Such overflow strategies remain invisible to ecGEMs unless explicit cofactor-balancing

constraints are imposed. A secondary explanation could be regulatory: downregulation of Swi1, a transcriptional repressor of ADH1/ADH2, may have indirectly led to Ald upregulation (Peterson & Herskowitz, 1992; Taguchi & Young, 1987). Together, this shows that cells maintain redundant or “unneeded” proteins for robustness, even when this reduces efficiency – a signature of evolutionary incompleteness that may be improved via ALE or further engineering.

A further limitation arises from isoenzyme representation. The ecGEM systematically allocates flux to the isoenzyme with the highest k_{cat}/MW ratio, even if this is biologically inactive. For instance, the model routed all flux through Err1, annotated as an enolase, since it carries the same k_{cat} as Eno1/Eno2 but at lower molecular weight. *In vivo*, however, proteomics detected only Eno1 and Eno2 at high abundance, not Err1. This mismatch illustrates a fundamental modeling limitation: ecGEMs optimize catalytic efficiency, while real cells use isoenzymes tuned by regulation, compartmentation, and robustness. Isoenzyme choice can therefore be misrepresented, emphasizing the need for organism-specific curation of k_{cat} values and careful validation of annotation sources.

Overall, these results show that ecGEMs are powerful in predicting global proteome allocation trends (e.g., glycolysis upregulation), but can misrepresent specific pathways, especially in redox balancing, biosynthesis, and isoenzyme usage. This reflects the model’s bias toward stoichiometric efficiency rather than the robustness and flexibility observed *in vivo*. A limitation of current generation ecGEMs is also their reliance on a static biomass composition, whereas *in vivo* biomass allocation varies substantially with growth rate and condition. For example, ribosomal protein fraction scales directly with growth rate in many organisms (Björkeroth et al., 2020; Scott & Hwa, 2011), yet this dependency is not reflected in most ecGEM frameworks. Incorporating such variability could improve predictive accuracy, especially for proteome allocation and growth-rate trade-offs. A possible improvement would therefore be to implement automatic adjustment of biomass composition, even in the absence of experimental data, using empirical growth laws as constraints (Lange & Heijnen, 2001). However, this refinement must be balanced against the feasibility of maintaining efficient linear programming formulations, as dynamic biomass functions increase model complexity. The benefits of introducing variable biomass composition have been recognized before (Elsemman et al., 2022; O’Brien et al., 2013; Pramanik & Keasling, 1997), and doing so in ecGEMs could help reconcile discrepancies between predicted and observed proteome allocations.

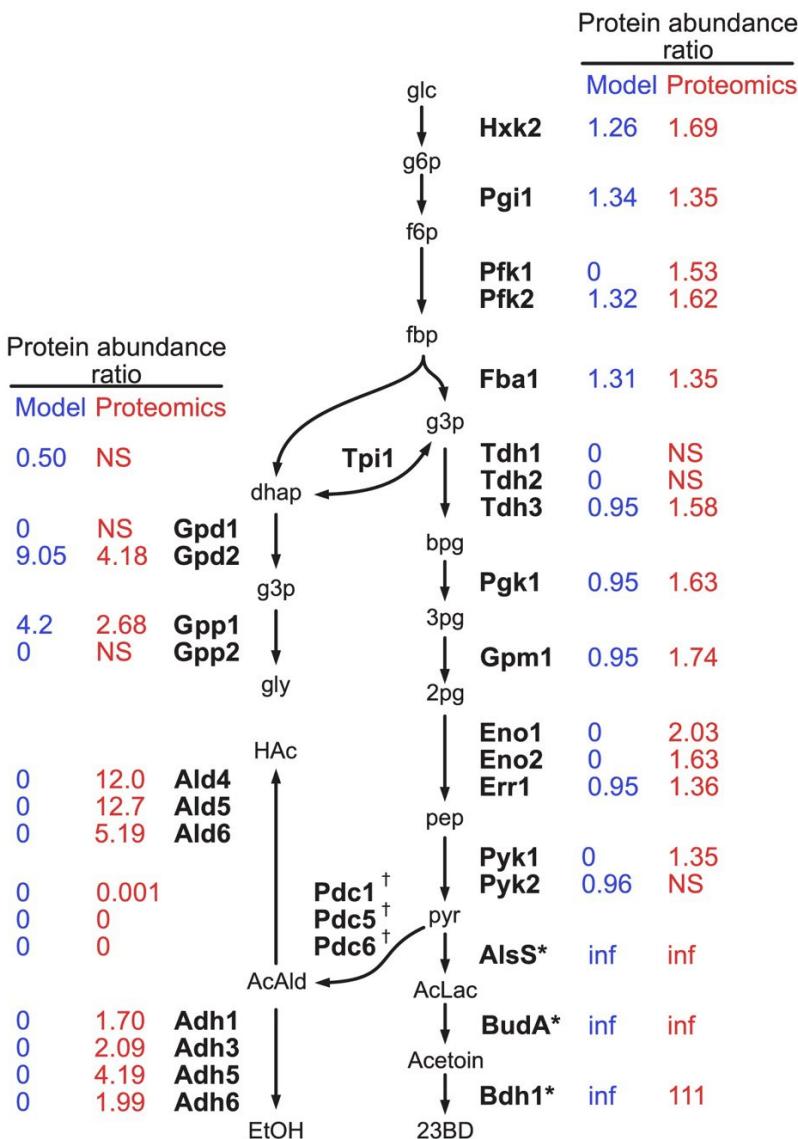


Figure 2. Comparison of ratios of predicted (blue) and measured (red) protein abundance of the 23BD-glycerol co-producing strain GSY014 and the reference strain GSY013. Predicted ratios were calculated from ecGEM mean fluxes of respective enzyme pseudoreactions. Measured ratios were calculated from absolute mean protein abundances ($\mu\text{g g}_{\text{total protein}}^{-1}$) of three replicates distinguishing between concrete values of differentially expressed proteins above 0.05 significance level and no significant change (NS) in expression of the respective enzymes in the two strains. * overexpressed genes; † deletion; Inf is due to 0 flux in the reference strain. Abbreviations: glc, glucose; g6p, glucose-6-phosphate; f6p, fructose-6-phosphate; fbp, fructose-1,6-bisphosphate; g3p, glyceraldehyde-3-phosphate; bpg, 2,3-bisphosphoglycerate; 3 pg, 3-phosphoglycerate; 2 pg, 2-phosphoglycerate; pep, phosphoenolpyruvate; pyr, pyruvate; AcLac, acetolactate; 23BD, 2,3-butanediol; dhap, dihydroxyacetone phosphate; g3p, glycerol-3-phosphate; gly, glycerol; Hac, acetate; AcAld, acetaldehyde; EtOH, ethanol. Reproduced from Sjöberg et al., Metabolic Engineering 82, 49-59, 2024, under a CC BY licence.

3.1.4 Proteomics-constrained modeling exposes limitations of *in vitro* enzyme data

We next refined the ecGEM by integrating 750 individual protein constraints, covering 72% of the total measured protein mass and 92% of the protein mass represented in the model. Two strain-specific ecGEMs were fitted: one for the engineered 23BD-producing strain GSY014 and one for the reference strain GSY013. In both models, additional flexibilization of individual protein constraints by adjusting their protein abundance was required to achieve feasible flux states. For the engineered strain, 46 proteins – including several glycolytic enzymes and the heterologous pathway enzymes BudA and AlsS – were flexibilized (Table S2), whereas the reference strain required flexibilization of 72 proteins, primarily biosynthetic enzymes. Despite these differences, the total mass of flexibilized protein was comparable between strains (50 vs. 54 mg gDCW⁻¹). This contrast suggests that the engineered strain faced specific bottlenecks in glycolysis and the heterologous pathway, while the reference strain was constrained more broadly at the level of biosynthetic capacity.

Flexibilization in this context indicates that the measured enzyme abundances were insufficient to support feasible flux states under the assumed k_{cat} values, exposing limitations of enzyme data – either experimental underestimation, or incorrect kinetic annotation in the model. Indeed, if the higher literature values of key glycolytic enzymes were chosen by the Bayesian procedure (F. Li et al., 2022), the flexibilization would have not been required. Similarly, the manually added k_{cat} values for AlsS and BudA had to be increased, suggesting that the heterologous enzymes are likely more efficient *in vivo* than current *in vitro* data indicates. These observations reinforce the need for improved correlation between *in vitro* enzymology and *in vivo* performance (García-Contreras et al., 2012; F. Li et al., 2022) to strengthen future ecGEM predictions.

After integrating individual protein constraints, the total protein constraint itself had to be adjusted upward to 382 mg gDCW⁻¹ to achieve feasible flux states. By comparison, in the coarse total-protein model (without enzyme-specific limits), the protein pool pseudoreaction was capped at only 123 mg gDCW⁻¹. Both values fall short of the experimentally measured proteome content of 450 mg gDCW⁻¹, but the comparison is instructive: the coarse model underestimated protein demand, while the individually constrained model narrowed the gap, albeit still underpredicting the total proteome. This threefold increase (123 → 382 mg gDCW⁻¹) illustrates how enzyme-specific constraints drastically reduce proteome allocation flexibility, forcing the solver to distribute protein mass more realistically across pathways. As a result, the solver required a higher effective proteome ceiling to achieve the same extracellular fluxes. In practice, this means that either (i) some enzymes operate at higher *in vivo* catalytic efficiencies than annotated k_{cat} values suggest, (ii) isoenzymes or complexes not captured in the model contribute additional capacity, or (iii) experimental proteome measurements underestimate the available protein pool. From a modeling perspective, the need for such a large upward adjustment underscores the sensitivity of ecGEMs to protein annotation and highlights a key trade-off: resolution increases, but so does dependence on accurate kinetic parameters.

The integration of individual protein constraints had major consequences for intracellular flux predictions. Flux variability was markedly reduced in both strains (Figure S4), meaning that fewer alternative flux states were compatible with the measured proteome. In other words, the proteome data served as an additional filter, forcing the model to abandon many of the equally optimal but biologically implausible flux values

that were possible under total protein constraints alone. The distribution of flux values obtained through sampling also became tighter after protein integration (**Figure 3A**), reflecting a reduction in the uncertainty of flux predictions. By contrast, extracellular fluxes predicted by the refined ecGEMs were only slightly lower than before, indicating that global phenotypes are robust, while intracellular pathway choices become more constrained when protein limits are considered.

When comparing mean fluxes between the coarse total-protein and proteomics-constrained models (**Figure 3A**), we observed a widespread across both pseudoreactions and metabolic reactions. This spread indicates that pathways were reallocated once enzyme capacities were enforced, decreased fluxes in the most optimal (as indicated by k_{cat}/MW) enzymes and increased flux through less efficient isoenzymes and alternate pathways. In addition, the number of non-zero fluxes decreased by 11.2%, meaning increased diversification in enzyme usage. This implies that the proteome constraints suppress minor or inefficient side pathways, focusing the solution on a leaner, more defined metabolic strategy.

An example of such reallocation is the switch between Pyk1 and Pyk2 isoenzymes (**Figure 3B**). Under total protein constraints, one isoform was favored, but once individual protein limits were introduced, the predicted flux shifted to the other isoenzyme. Such predictions highlight the potential of ecGEMs to resolve not only pathway-level shifts but also fine-scale differences in enzyme choice – provided kinetic annotations and proteomic data are accurate.

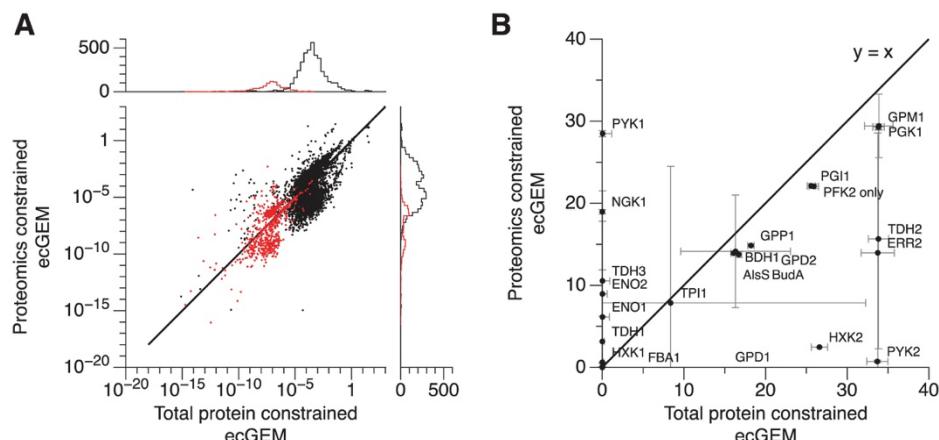


Figure 3. Comparison of mean sampled fluxes between proteome-constrained and total protein ecGEMs. **A.** Pairwise correlation of mean reaction fluxes obtained from sampling ($n = 10'000$) in both models. Each point represents a single reaction. Marginal histograms along the top and right axes show the distribution of mean flux values, expressed as counts of reactions. **B.** Detailed comparison of selected reactions shown in **Figure 2**, including isoenzymes. Red points denote protein pseudoreactions, black points denote metabolic reactions. Reproduced from Sjöberg et al., *Metabolic Engineering* 82, 49-59, 2024, under a CC BY license.

3.2 Study 2: Predicting and validating central carbon metabolism in the oleaginous yeast *R. toruloides* (Publications II and III)

The non-conventional Basidiomycota yeast *Rhodotorula toruloides* can accumulate lipids exceeding 50% of its dry biomass, making it a promising host for sustainable lipid production. As explained in the introductory section, lipid biosynthesis relies on a steady cytosolic supply of acetyl-CoA and NADPH, produced through interconnected central carbon pathways. Unlike *Saccharomyces cerevisiae*, *R. toruloides* possesses ATP-citrate lyase (ACL), which cleaves citrate into cytosolic acetyl-CoA and oxaloacetate, and phosphoketolase (PK), which generates acetyl-phosphate from sugar phosphates (**Figure 4**) (Z. Zhu et al., 2012). NADPH regeneration can occur either via the oxidative pentose phosphate pathway (oxPPP) or, if ACL catalyzes the conversion of citrate into acetyl-CoA and oxaloacetate (OAA), latter can be transformed into malate and further converted into pyruvate through cytosolic malic enzyme (cMAE) (**Figure 4**). Alternatively, NADPH can be regenerated in mitochondria via NADP-dependent IDH (Chawla et al., 2022; Sreeharsha & Mohan, 2020; Yang et al., 2012). Previous studies have shown that these strategies are organism-specific, and ACL and oxPPP dominate in oleaginous yeasts, such as *Yarrowia lipolytica* and *Rhodotorula glutinis* (Blazeck et al., 2014; Dulermo et al., 2015; Qiao et al., 2017; Wasyleenko et al., 2015; Yoon et al., 1984; H. Zhang et al., 2013; J. Zhu et al., 2023), whereas TCA- and malate-linked reactions contribute more in filamentous fungi (Wynn et al., 1997, 2001; Wynn & Ratledge, 1997; Y. Zhang et al., 2007). However, the relative importance and regulation of these pathways in *R. toruloides* remain unclear, particularly regarding how precursor supply strategies shift between fermentable carbon sources, such as glucose and xylose, and non-fermentable carbon sources, such as acetate.

Recent genome-scale models have generated hypotheses about substrate-specific metabolism in *R. toruloides* (Lopes et al., 2020, p. 202; Pinheiro et al., 2020), but these lacked (i) enzyme constraints, (ii) substrate-resolved biomass composition, and (iii) experimental validation. Given large shifts in cellular resource allocation between proteome and lipidome reported in previous studies (Pinheiro et al., 2020; Tiukova, Brandenburg, et al., 2019; Z. Zhu et al., 2012), this is a relevant challenge for an enzyme-constrained model. In **Publications II and III** we integrated absolute proteomics, enzyme-constrained genome-scale modeling (ecGEM), and knockout phenotyping to dissect substrate-dependent flux distributions through key metabolic nodes including pyruvate dehydrogenase (PDH), citrate synthase (CIT), ACL, malate dehydrogenase (MDH1), pyruvate carboxylase (PC), and others (**Figure 4**), providing a mechanistic framework for understanding lipid precursor supply in *R. toruloides*.

To obtain precise, high-quality experimental data, we performed controlled growth characterization on glucose, xylose, and – for the first time – acetate, analyzed global proteome allocation and growth constraints, constructed and curated the first ecGEM of *R. toruloides*, and validated predictions through gene knockout phenotyping. This integrative framework revealed substrate-specific essentiality of ACL, PK, and cMAE, and highlighted major discrepancies between model predictions and experimental outcomes. Together, these results demonstrate both the power and the current limitations of ecGEMs in oleaginous yeasts, providing a mechanistic understanding of how *R. toruloides* flexibly reorganizes its metabolism to support lipogenesis and informing future metabolic engineering strategies.

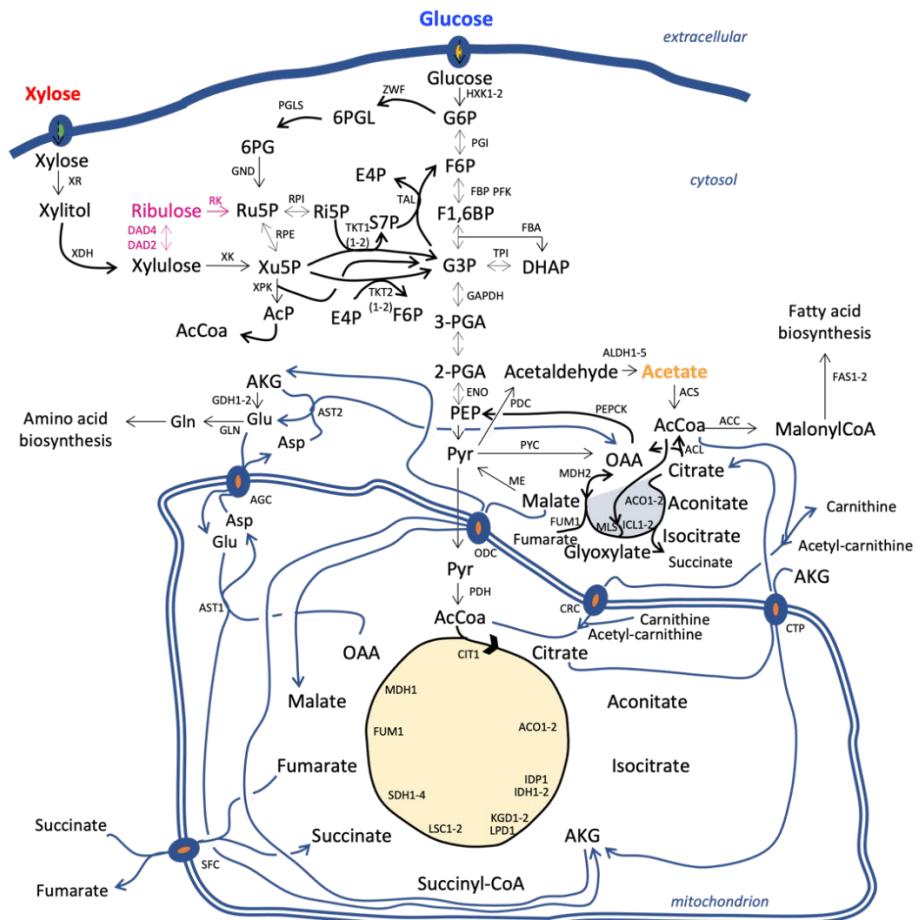


Figure 4. Metabolism of *Rhodotorula toruloides*. Blue arrows denote mitochondrial carrier proteins and enzymes in shuttling pathways. Pink arrows denote alternative xylose assimilation pathways. Reproduced from Rekēna et al., PLOS Comput. Biol. 19(4):e1011009, 2023, under a CC BY licence.

3.2.1 Global proteome allocation analysis reveals metabolic strategies and growth constraints in *R. toruloides*

We quantified 3,160 proteins across six conditions (glucose, xylose, acetate; exponential vs. nitrogen limitation) using stable isotope labeling (SILAC) and the total protein quantification approach (Sánchez et al., 2021). Replicate consistency was satisfactory (Figure 5A). Up to 4% g/gDCW of the proteome was uncharacterized, which limited the use of automated tools such as Proteomap and gene set analysis, but *manual* Gene Ontology grouping allowed us to track functional allocation throughout exponential growth and lipid accumulation.

Global proteome allocation showed substrate- and phase-specific patterns (Figure 5B). On glucose, the glycolytic fraction increased during nitrogen limitation, whereas on xylose it remained unchanged. The TCA cycle fraction did not shift significantly between phases, while the ETC fraction increased during lipid accumulation on all substrates, consistent with higher respiratory demand. Acetate-grown cells displayed a distinct pattern: proteome allocation changed little upon nitrogen starvation, in contrast to glucose

and xylose, where exponential and nitrogen-limited phases were clearly separated. On acetate, more proteome was invested into the TCA cycle, pyruvate metabolism, ETC, and amino acid biosynthesis, while on glucose and xylose a larger share was allocated to ribosome synthesis. This difference reflects the fact that acetate enters metabolism only as acetyl-CoA, providing no ATP through glycolysis. Thus, the TCA cycle must supply both energy and precursors, necessitating stronger investment in TCA and ETC enzymes.

Differential expression analysis revealed the largest number of significantly altered proteins between the exponential growth and nitrogen-limited phases on glucose (186 proteins), while none were detected between phases on acetate ($|\log_2\text{FC}| > 1$, BH-adjusted $p < 0.05$; **Figure S5**). This suggests that sugar-grown cells undergo stronger proteomic reprogramming during lipid accumulation, while acetate-grown cells rely more on constitutive metabolism.

Key enzymes for acetyl-CoA supply reflected clear substrate-specific strategies. On glucose, ACL was abundant ($1010 \pm 75 \mu\text{g g}^{-1}$ _protein) and upregulated 2.6-fold during lipid accumulation ($\text{apval} = 0.039$), at levels comparable to glycolytic enzymes (**Figure S6, Table S3**). Together with high PDH ($2304 \pm 11 \mu\text{g g}^{-1}$) and CIT ($2573 \pm 6 \mu\text{g g}^{-1}$), this points to citrate export and cleavage by ACL as a major cytosolic acetyl-CoA route. MDH1 was extremely abundant ($4195 \pm 3 \mu\text{g g}^{-1}$), though it declined during lipid accumulation ($3168 \pm 24 \mu\text{g g}^{-1}$, 1.3-fold \downarrow , $\text{apval} = 0.015$). In parallel, PC was also strongly expressed ($569 \pm 24 \rightarrow 712 \pm 5 \mu\text{g g}^{-1}$, 1.25-fold \uparrow , $\text{apval} = 0.048$), indicating that glucose-grown cells reinforce oxaloacetate regeneration through both pyruvate carboxylation and malate–oxaloacetate cycling. This dual system likely sustains the citrate–malate shuttle while buffering NADH/NAD⁺ redox balance, ensuring efficient acetyl-CoA and precursor supply under high glycolytic flux.

On xylose, ACL levels were similar, but induction was stronger (4.8-fold, $\text{apval} = 0.025$), supporting its role in lipogenesis. MDH1 was moderately abundant ($3237 \pm 8 \mu\text{g g}^{-1}$) and unchanged during lipid accumulation ($3089 \pm 305 \mu\text{g g}^{-1}$, $\text{apval} = 0.69$). By contrast, PC started lower ($249 \pm 9 \mu\text{g g}^{-1}$) but increased 2.4-fold to $604 \pm 6 \mu\text{g g}^{-1}$ ($\text{apval} = 0.025$). This pattern suggests that on xylose, ACL remains the main cytosolic acetyl-CoA supplier, while PC provides strong reinforcement of OAA regeneration under pentose metabolism, helping balance flux through glycolysis and the PPP.

On acetate, ACL abundance was lower ($148 \pm 10 \mu\text{g g}^{-1}$) and its 2.8-fold increase was not significant ($\text{apval} = 0.14$), suggesting a limited role. Instead, ACS was highly abundant ($3105 \pm 53 \mu\text{g g}^{-1}$), confirming its function as the main entry point for acetate assimilation. High abundances of CIT, ACO1, and MDH1 relative to other TCA enzymes (12.6, 4.9, and 10.8 mg g^{-1} _protein) with mild downregulation during nitrogen limitation (1.2–1.4-fold, $\text{apval} = 0.15$ –0.35) suggest an internal citrate–isocitrate–malate cycle that could replenish oxaloacetate and balance redox cofactors. This apparent downregulation during lipid accumulation likely reflects reduced acetate uptake demand rather than diminished lipogenic capacity, since lipid accumulation is primarily driven by flux redistribution under nitrogen limitation. By contrast, PDC was consistently low across substrates (47 – $59 \mu\text{g g}^{-1}$), indicating it plays only a minor role. PK showed moderate levels on glucose ($490 \pm 45 \mu\text{g g}^{-1}$, 1.6-fold \uparrow , $\text{apval} = 0.042$) and xylose ($632 \pm 50 \mu\text{g g}^{-1}$, 2.8-fold \uparrow , $\text{apval} = 0.064$). Although 4–6-fold lower than ACL or glycolytic enzymes, its consistent upregulation under nitrogen limitation suggests a conditional role in cytosolic acetyl-CoA supply.

NADPH regeneration enzymes showed less dramatic changes. cMAE abundance was modest (142 – $189 \mu\text{g g}^{-1}$) and unchanged across conditions (**Figure S6**), consistent with a

nonessential role. On acetate, IDP abundance was elevated ($1472 \pm 47 \mu\text{g g}^{-1}$, 2.4-fold ↑, $\text{apval} = 0.14$), suggesting possible contribution to NADPH supply, though evidence remains inconclusive. ZWF was abundant on glucose ($1309 \mu\text{g g}^{-1}$, 1.16-fold ↑, $\text{apval} = 0.015$), but lower on xylose ($735 \mu\text{g g}^{-1}$) and acetate ($637 \mu\text{g g}^{-1}$). Taken together, these data support the PPP as the dominant NADPH source on fermentative substrates, while the main NADPH pathway during acetate metabolism remains unresolved, as none of ZWF, cMAE, or IDP showed consistent evidence of a dominant role.

Ribosomal content and translation efficiency provided further insights into growth control, and in the future, it could be used to predict how yeast reallocates proteome under different substrates. Translation rates ranged from 0.8 to 6.6 amino acids s^{-1} (Figure 2C), which is in a similar range but below those reported for *S. cerevisiae* (2.8–10 aa s^{-1}) (Boehlke & Friesen, 1975). *R. toruloides* devoted 46% of its proteome to ribosomes, compared to 37% in *S. cerevisiae* (Metzl-Raz et al., 2017a), representing a 25% higher ribosome fraction. This suggests that *R. toruloides* offsets lower ribosome efficiency by investing more into ribosome abundance, perhaps a distinct allocation strategy in oleaginous yeasts. Growth rates correlated tightly with translation rate ($R^2 = 0.99$, $p < 0.01$) (Figure 2C), showing that translational capacity is the main determinant of growth. By contrast, correlation with ribosome fraction was weaker ($R^2 = 0.68$, $p < 0.043$) (Figure 2C). Ribosome fractions were lowest during nitrogen limitation on acetate and xylose. This indicates that growth is constrained more by ribosome activity than by ribosome abundance, consistent with reports of inactive ribosome pools maintained under nutrient limitation (Metzl-Raz et al., 2017b; Warner, 1999).

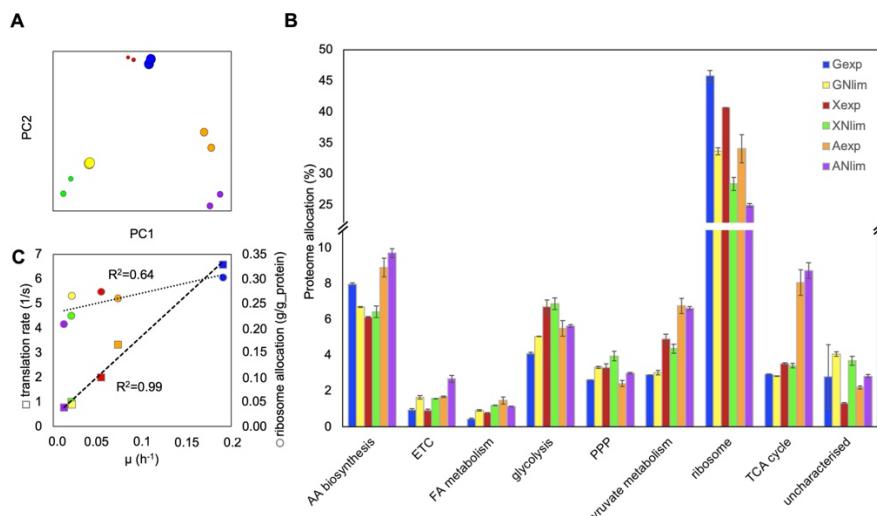


Figure 5. Absolute proteome quantification in *R. toruloides* wild type strain CCT 7815 cultivated in batch bioreactors with low-nitrogen chemically defined media containing glucose (G), xylose (X), or acetate (A) as the carbon source during exponential (exp) and nitrogen-limited (Nlim) phases. **A.** Principal component analysis. **B.** Allocation of absolute mean protein abundances (% of $\mu\text{g g}_{\text{total_protein}}^{-1}$) of two replicates to manually assigned major functional groups (annotation file available online: <https://doi.org/10.1371/journal.pcbi.1011009.s002>). **C.** Translation rate (s^{-1}) and ribosome allocation ($\text{g g}_{\text{protein}}^{-1}$). Error bars represent standard deviation of two experiments. Abbreviations: AA, amino acid; ETC, electron transport chain; FA, fatty acid; PPP, pentose phosphate pathway; TCA, tricarboxylic acid. Reproduced from Rekēna et al., PLOS Comput. Biol. 19(4):e1011009, 2023, under a CC BY licence.

3.2.2 Construction and curation of the first ecGEM of *R. toruloides* reveals limits of enzyme-constrained modeling in lipogenesis

To construct the first ecGEM of *R. toruloides*, we measured growth and exchange rates of the hydrolysate-tolerant strain CCT 7815 in actively controlled aerobic batch cultures. This strain was previously developed by short-term adaptation of CCT 0783 in sugarcane bagasse hydrolysate and is characterized by enhanced tolerance and lipid accumulation (Bonturi et al., 2017). Characterization for growth was performed in low-nitrogen media to induce lipid accumulation (Lopes et al., 2020). Online biomass and CO₂ measurements revealed clear transitions separating exponential growth from nitrogen limitation, allowing these phases to be analyzed independently (Figure S7).

During exponential growth, cells divided actively until nitrogen was depleted. Upon nitrogen exhaustion, growth ceased because nitrogen is required for amino acid, nucleotide, and cell wall biosynthesis. As carbon, minerals, and cofactors remained available, cells redirected metabolism toward maintenance and storage lipid accumulation. These distinct physiological states – growth-coupled metabolism and nitrogen-limited lipogenesis – represent fundamentally different proteome allocation and flux regimes and were therefore modeled separately in the ecGEM to capture the metabolic shift between biomass formation and storage metabolism.

Growth characterization and physiological parameters

All cultivations were performed in chemically defined media without amino acid supplementation. Although this resulted in a longer lag phase (10–16 h), it eliminated by-products from complex organic sources, improving reproducibility and interpretation of physiological data. Biomass yields on substrate differed markedly between phases and substrates (Table 8). During exponential growth, formation reached $515 \pm 27 \text{ mg g}^{-1}$ on glucose, $193 \pm 6 \text{ mg g}^{-1}$ on xylose, and $185 \pm 24 \text{ mg g}^{-1}$ on acetate. Under nitrogen limitation, yields fell on glucose ($236 \pm 29 \text{ mg g}^{-1}$) but rose on xylose ($316 \pm 34 \text{ mg g}^{-1}$), while acetate remained largely unchanged ($109 \pm 20 \text{ mg g}^{-1}$). Carbon balances closed well on xylose (89–109%), supported by significant xylitol and D-arabinitol secretion during exponential growth (0.223 ± 0.03 and $0.367 \pm 0.04 \text{ mmol (gDCW h)}^{-1}$). In contrast, balances did not close on glucose (106% in exponential, 68% in nitrogen limitation) or acetate (85% and 67%), despite detection of citrate secretion during exponential growth ($0.122 \pm 0.016 \text{ mmol (gDCW h)}^{-1}$). Missing byproducts remain to be identified.

Glucose cultures formed aggregates, complicating biomass estimation. Aggregation was reduced by switching the nitrogen source from ammonium sulfate to urea, though not eliminated. Given a recently reported EPS production by *R. toruloides* under ammonium sulfate at acidic pH (< 2.3; (Sepúlveda Del Rio Hamacek et al., 2025)), EPS secretion is not a possible explanation at pH 6, though not confirmed. All glucose data reported here (rates, yields, proteome) are from urea-grown cultures (a comparison of the growth curves when using ammonium sulfate or urea shown on Figure S7A).

Growth rates also reflected substrate differences (Table 8). Maximum specific growth was threefold higher on glucose ($0.19 \pm 0.025 \text{ h}^{-1}$) than on xylose (0.054 h^{-1}) or acetate (0.073 h^{-1}), consistent with published values (Tiukova, Brandenburg, et al., 2019). During nitrogen limitation, apparent rates were very low (0.012 – 0.021 h^{-1}) and were thus denoted as non-quantifiable. Since protein synthesis ceases when nitrogen is depleted, these values likely reflect lipid accumulation and measurement noise rather than true biomass growth. Supporting this, secreted byproducts decreased in concentration during nitrogen limitation, indicating re-assimilation as secondary carbon sources to support

maintenance metabolism and lipid storage. Examples include re-consumption of xylitol and D-arabinitol on xylose and citrate on acetate (-0.043 ± 0.001 mmol (gDCWh) $^{-1}$). Using stereoselective HPLC, we confirmed that the arabinitol produced was the D-isomer (**Figure S8**), validating a pathway proposed by Jagtap and Rao (Jagtap & Rao, 2018).

The specific CO_2 production, O_2 uptake, and substrate uptake rates further confirmed the metabolic shift (**Table 8**). As expected, metabolic activity was considerably higher during exponential growth: CO_2 production reached 2.7–5.2 mmol (gCDW h) $^{-1}$ compared to 1.3–2.2 mmol (gCDW h) $^{-1}$ under nitrogen limitation. A similar pattern was observed for oxygen uptake, which was -1.4 to -7.1 mmol (gCDW h) $^{-1}$ in the exponential phase but only -0.7 to -2.3 mmol (gCDW h) $^{-1}$ during nitrogen limitation. Substrate uptake also slowed down substantially, from -1.9 to -6.6 mmol (gCDW h) $^{-1}$ in exponential growth to -0.4 to -2.0 mmol (gCDW h) $^{-1}$ in nitrogen limitation. These measurements confirm that cells shift from a high-flux, energy-demanding metabolism during active growth to a reduced metabolic state under nitrogen limitation.

Table 8: Biomass yields, specific rates and carbon balances of *R. toruloides* wild type strain CCT 7815 cultivated in batch bioreactors with low-nitrogen chemically defined media containing glucose (G), xylose (X), or acetate (A) as the carbon source during exponential growth (exp) and nitrogen-limited (Nlim) phases. Average and standard deviation of duplicate experiments are presented.

Condition	μ (h $^{-1}$)	Specific rate of consumption or production (mmol (gCDW h) $^{-1}$)			Yield (mg (g substrate) $^{-1}$)	Yields on biomass (mg gCDW $^{-1}$)		Recovery of consumed substrate based on carbon
		CO_2	O_2	substrate		Total protein	Total lipid	
Gexp	0.191 ± 0.025	2.685 ± 0.584	-1.423 ± 0.779	-2.096 ± 0.557	515 ± 27	636 ± 24	57 ± 2	106 ± 16
GNlim	0.021 ± 0.007^a	1.547 ± 0.519	-0.704 ± 0.338	-0.410 ± 0.007	236 ± 29	227 ± 31	483 ± 41	69 ± 6
Xexp	0.054 ± 0.001	2.845 ± 0.325	-2.678 ± 0.012	-1.859 ± 0.022	193 ± 6	438 ± 13	202 ± 25	89 ± 6
XNlim	0.021 ± 0.001^a	1.277 ± 0.139	-1.028 ± 0.012	-0.434 ± 0.026	316 ± 34	197 ± 2	290 ± 6	109 ± 18
Aexp	0.073 ± 0.003	5.224 ± 0.724	-7.144	-6.627 ± 0.557	185 ± 24	386 ± 12	175 ± 19	85 ± 5
ANlim	0.012 ± 0.0^a	2.229 ± 0.198	-2.265	-1.971 ± 0.429	109 ± 20	217 ± 12	341 ± 13	67 ± 3

a – Below reliable measurement threshold

Model reconstruction and condition-specific parametrization of ecGEM

We based our enzyme-constrained genome-scale metabolic model (ecGEM) on rhtoGEM, one of the first GEMs of *R. toruloides*, originally derived from *S. cerevisiae* and subsequently parameterized with *R. toruloides*-specific data (Tiukova, Prigent, et al., 2019). Because one allele of the polyploid strain CCT 7815 shares 70–90% sequence identity with the *R. toruloides* NP11 genome (Bonturi et al., 2022), NP11 was selected as the reference for annotation.

Motivated by the absence of xylulokinase in the measured proteome, we incorporated a ribulose-mediated xylose assimilation pathway (**Figure 4**, pink pathway), supported by the detection of D-arabinitol dehydrogenase (DAD-4) on xylose (1913 ± 272 $\mu\text{g g}^{-1}$,

3.5-fold increase compared with the glucose condition, $\text{apval} = 0.06$; **Figure S6**) and the presence of D-arabinitol among identified extracellular metabolites (**Figure S7B**).

To fit the model to experimental data, we expanded the enzyme annotation coverage to 773 enzymes corresponding to 1515 reactions ($\approx 30\%$ of total reactions) by manually assigning NP11 enzyme identifiers (protein list available online: <https://doi.org/10.1371/journal.pcbi.1011009.s004>). Using the GECKO 2.0 framework, we reconstructed six condition-specific proteomics-constrained ecRhtoGEMs, representing growth under carbon- or nitrogen-limited regimes (**Table 8**).

The protein content during exponential growth ($386\text{--}636\text{ mg gCDW}^{-1}$) was comparable to *S. cerevisiae* (Xia et al., 2022) but decreased nearly twofold under nitrogen limitation ($197\text{--}227\text{ mg gCDW}^{-1}$). These values are higher, yet consistent with those previously reported for another *R. toruloides* strain under steady-state conditions (**Table 8**) (Shen et al., 2017). In contrast, the lipid content increased up to eightfold under nitrogen limitation (e.g., $483 \pm 41\text{ mg gCDW}^{-1}$ on glucose vs. $57 \pm 2\text{ mg gCDW}^{-1}$ in exponential growth), confirming the onset of lipid accumulation consistent with previous report by Tiukova et al. (**Table 8**) (Tiukova, Brandenburg, et al., 2019). Oleate (C18:1) was the dominant fatty acid species (**Table S4**), and overall fatty acid composition was consistent with the previous report (Tiukova, Brandenburg, et al., 2019).

To ensure feasible flux states consistent with the experimental data, measured fatty acid compositions, total lipid and protein content were introduced into the biomass stoichiometry of all condition-specific ecGEMs. These modifications required additional technical updates, including (i) revised growth- (GAM) and non-growth-associated maintenance (NGAM), (ii) recalibration of average enzyme saturation, and (iii) updated f-factor and protein pool constraint calculations, as described in **Publication II**. The GAM parameters were inherited from the *S. cerevisiae* GEM and re-parameterized according to the GECKO formalism, originally based on a fitted value of $59.276\text{ mmol gCDW}^{-1}$ for yeast (Förster et al., 2003). The NGAM requirements were also inherited from the *S. cerevisiae* GEM and re-fitted using the GECKO algorithm (from the starting value of $0.7\text{ mmol (gCDW h)}^{-1}$ for aerobic conditions). Although the rhtoGEM reflects ATP synthase stoichiometry typical of Complex I-deficient cells, both GEMs and annotated genomes of *R. toruloides* indicate the presence of Complex I (Dinh et al., 2019; Zhu et al., 2012). We retained the same ATP synthase stoichiometry from the yeastGEM for the ecRhtoGEM, as the functional presence of Complex I in *R. toruloides* remains to be experimentally confirmed.

As these functionalities were not natively implemented in the GECKO toolbox, we developed custom code modules that integrate these updates directly into the pipeline. All model edits, scripts, and data are available at:

 <https://github.com/alinarekena/ecRhtoGEM>.

Model evaluation

To achieve feasible flux states, in total 59 protein concentrations and k_{cat} values required flexibilization (**Table S5**). Without these adjustments, the proteomics-constrained ecGEMs were infeasible, while the conventional rhtoGEM could run but overpredicted exchange fluxes. Enzymes frequently flexibilized included ACL (M7WHC9), FAS (M7WSW5), and cytochrome c (M7WUI0) oxidase, reflecting lipid metabolism and respiratory demand. k_{cat} values for enzymes of the ribulose pathway were also increased. Detailed information about the curation of individual k_{cat} and protein concentration values is saved on the Github repository.

The ecGEM predicted growth rates were generally in good agreement with the experimental data. With measured boundary conditions – including extracellular fluxes (such as substrate uptake) and total protein abundance – applied, the model optimized protein allocation and predicted growth rates that closely matched the experimental measurements when above the reliable detection threshold (**Table 9**). We then examined predicted exchange fluxes, including substrate uptake and gas exchange.

On glucose, substrate uptake was overestimated during exponential growth (-2.46 vs. -2.01 mmol (gCDW h) $^{-1}$) but matched measurements under nitrogen limitation (-0.41 mmol (gCDW h) $^{-1}$). CO₂ production was strongly overestimated during exponential growth (2.6-fold higher, 7.10 vs. 2.69 mmol (gCDW h) $^{-1}$) but aligned with measurements in nitrogen limitation (1.45 vs. 1.55 mmol (gCDW h) $^{-1}$). Oxygen uptake was also overpredicted, by 4.2-fold during exponential growth and 1.7-fold during nitrogen limitation. In addition, the model predicted low pyruvate secretion (0.18 mmol (gCDW h) $^{-1}$), which was not detected experimentally. Pyruvate secretion was predicted only during growth on glucose, consistent with its direct formation via glycolysis and the high glycolytic flux under these conditions. Other substrates enter central metabolism downstream of pyruvate or generate less cytosolic NADH, reducing both carbon and redox pressure at the pyruvate node. Combined with constrained CO₂ efflux during exponential growth, this likely explains why minor pyruvate excretion was predicted exclusively on glucose.

On xylose, predicted uptake rates matched measurements within experimental uncertainty under both conditions. CO₂ production was close to measured during exponential growth (2.95 vs. 2.84 mmol (gCDW h) $^{-1}$) but underestimated in nitrogen limitation (1.11 vs. 2.84 mmol (gCDW h) $^{-1}$), consistent with the underpredicted growth rate. Oxygen uptake was slightly lower than measured in exponential growth (-2.29 vs. -2.68 mmol (gCDW h) $^{-1}$) but closely matched under nitrogen limitation (-0.97 vs. -1.03 mmol (gCDW h) $^{-1}$). A small amount of L-arabinitol secretion was predicted but not confirmed experimentally.

On acetate, predicted uptake rates were slightly below measured in both growth phases. CO₂ production was overestimated during exponential growth (6.73 vs. 5.22 mmol (gCDW h) $^{-1}$) and underestimated in nitrogen limitation (1.47 vs. 2.23 mmol (gCDW h) $^{-1}$). Oxygen uptake was likewise underestimated (exponential growth: -6.67 vs. -7.14 mmol (gCDW h) $^{-1}$; nitrogen limitation: -2.02 vs. -2.27 mmol (gCDW h) $^{-1}$). The model also predicted secretion of succinate in both phases (0.33 and 0.033 mmol (gCDW h) $^{-1}$), which remains to be validated experimentally.

Table 9: Biomass specific rates simulated using condition-specific, proteomics-constrained ecGEMs of *R. toruloides*. Median and standard deviation of 2000 flux samples are presented. Abbreviations: G, glucose; X, xylose; A, acetate; exp, exponential growth phase; Nlim, nitrogen-limited phase.

Condition	μ (h ⁻¹)	Specific rate of consumption or production (mmol (gCDW h) ⁻¹)		
		substrate	CO ₂	O ₂
Gexp	0.178 ± 0.00	-2.464 ± 0.011	7.101 ± 0.20	-5.944 ± 0.14
GNlim	0.0208 ± 0.00 ^a	-0.406 ± 0.0019	1.446 ± 0.029	-1.164 ± 0.027
Xexp	0.0535 ± 0.00013	-1.841 ± 0.0078	2.948 ± 0.055	-2.292 ± 0.048
XNlim	0.0151 ± 0.00 ^a	-0.433 ± 0.0015	1.110 ± 0.0082	-0.966 ± 0.0085
Aexp	0.0723 ± 0.00	-6.039 ± 0.0010	6.731 ± 0.070	-6.668 ± 0.0075
ANlim	0.0119 ± 0.00 ^a	-1.951 ± 0.00057	1.472 ± 0.044	-2.016 ± 0.046

a – Below reliable measurement threshold

Overall, comparison of predicted and measured rates for the wild-type phenotype in *R. toruloides* showed that the proteome-constrained ecGEM reproduced growth rates and many exchange fluxes reasonably well across conditions, though notable discrepancies remained. Some discrepancies may partly reflect experimental uncertainties (e.g., aggregate formation on glucose cultures), but also illustrate the model's prediction capacity at more extreme cases. For example, the least precise predictions were received on the glucose exponential phase, in which protein content (64%) was higher than normal. It is an inherent limitation of flux balance analysis that it provides one feasible solution that satisfies the constraints rather than representing the exact physiological state of the cell (Mahadevan & Schilling, 2003). Although random sampling (2000 iterations) broadened the solution space, it seemingly did not eliminate off-target effects within the permitted interval (simulations were allowed 10% deviation from measured rates). Particularly during lipid-accumulating conditions, where protein content was lower than in *S. cerevisiae*, it highlighted again the need to enable ecGEMs to automatically adjust biomass composition and current limitations in directly transferring ecGEM frameworks across species. Finally, achieving this agreement required flexibilization of measured protein concentration constraints. The need to flexibilize many enzyme k_{cat} and protein abundance parameters can be attributed to the lack of *R. toruloides*-specific kinetic data. All k_{cat} values in the model were drawn from BRENDA entries of distantly related organisms, which likely reduced predictive precision. This underscores the importance of generating organism-specific kinetic parameters to improve future ecGEM reconstructions in non-model organisms.

3.2.3 Predicted intracellular fluxes reveal substrate-dependent strategies for acetyl-CoA and NADPH supply during lipogenesis

Motivated by the ecGEM's ability to reproduce exchange fluxes, we compared predicted intracellular fluxes between exponential growth and nitrogen limitation to probe lipogenesis mechanisms in *R. toruloides* and their substrate dependence. From these fluxes and absolute proteomics, we calculated apparent catalytic activities (k_{app}) as flux/protein. To our knowledge, this is the first report of *in silico* k_{app} values in *R. toruloides*, with most values in the range of an “average enzyme” (~ 10 s⁻¹) as reported by Bar-Even (Bar-Even et al., 2011). Because the ecGEM constrains fluxes by literature

k_{cat} , k_{app} values are upper-bounded by k_{cat} , i.e., any *in vivo* activity enhancements beyond cataloged k_{cat} cannot be captured. Notably, the number of enzymes with low k_{app} ($0.1\text{--}1\text{ s}^{-1}$) increased during Nlim (Figure S9), indicating that fluxes declined more strongly than protein abundances and leaving a larger fraction of the proteome underutilized. This suggests that *R. toruloides* maintains enzyme pools while throttling catalytic throughput, a trade-off that preserves metabolic flexibility while conserving nitrogen under lipid-accumulating conditions. A full dataset of k_{app} values is available online: <https://doi.org/10.1371/journal.pcbi.1011009.s009>.

Fermentative substrates (glucose, xylose): acetyl-CoA and NADPH supply

On glucose (exponential phase), predicted fluxes through PDH ($68 \pm 9\%$ of carbon) and CIT ($70 \pm 14\%$) exceeded average TCA flux (Figure S10), initially suggesting a mitochondrial PDH \rightarrow citrate export \rightarrow cytosolic ACL route. The MDH1 flux was also exceptionally high ($218 \pm 6\%$ of carbon, only slightly reduced to $209 \pm 381\%$ under nitrogen limitation), suggesting substantial malate–oxaloacetate cycling for redox balance, though the very high SD indicates weak constraint. Despite strong PDH and CIT activity, ACL consistently carried 0% flux (Figure S10). Exported citrate was instead routed back into the TCA via auxiliary shunts (succinate, malate).

PK emerged as the dominant cytosolic acetyl-CoA route, consistent with previously reported GEM predictions under similar conditions (Lopes et al., 2020; Pinheiro et al., 2020). We compared alternative PK pathway variants and found that the PTA-based (t_0082) and ACK-based (t_0886) routes generated nearly identical flux distributions (full dataset available online: <https://doi.org/10.1371/journal.pcbi.1011009.s011>). As no functional difference emerged, we report results for the PTA variant as representative. On glucose, PK fluxes rose 4-fold under nitrogen limitation ($14 \pm 55\% \rightarrow 60 \pm 26\%$), paralleling the lipid accumulation ($5.7\% \rightarrow 483\text{ mg gDCW}^{-1}$). In contrast, PDC declined ($22 \pm 39\% \rightarrow 12 \pm 103\%$) and PC dropped sharply ($26 \pm 47\% \rightarrow 7 \pm 142\%$), indicating minor roles during lipogenesis. On xylose, PK was consistently higher than PDC ($13 \pm 46\% \rightarrow 22 \pm 29\%$ vs. $9 \pm 151\% \rightarrow 7 \pm 71\%$), again marking PK the dominant cytosolic acetyl-CoA source. PC flux remained low ($7 \pm 87\% \rightarrow 4 \pm 55\%$), suggesting limited anaplerotic contribution. Importantly, some fluxes – especially PDC on xylose – had very large SDs ($\pm 53\text{--}151\%$), indicating weak constraints and representing alternative feasible states rather than robust predictions. By contrast, PDC on glucose (Nlim) had lower SDs ($\pm 16\text{--}26\%$), supporting higher confidence. Thus, PK is robustly predicted as the main cytosolic acetyl-CoA route on fermentative substrates, while ACL acts as a conditional backup only when PK is blocked.

Simulations with PK deletion confirmed this (full dataset available online: <https://doi.org/10.1371/journal.pcbi.1011009.s012>). PDC partially compensated: glucose 1.2-fold increase ($22\% \rightarrow 26\%$), xylose 1.9-fold increase ($9\% \rightarrow 17\%$), with further increases on glucose under Nlim (3.4-fold, $41 \pm 16\%$). In these scenarios, ACL became conditionally active (glucose 0% $\rightarrow 24 \pm 18\%$, xylose $2 \pm 26\% \rightarrow 18 \pm 19\%$), highlighting ACL's potential role as a backup acetyl-CoA route, especially under Nlim.

Acetate: ACS entry, internal cycling, and uncertain cytosolic acetyl-CoA routes

On acetate, all carbon was obligatorily channeled through ACS ($100 \pm 0.1\% / 100 \pm 1\%$), consistent with previously reported modeling under similar conditions (Lopes et al., 2020). From there, the model distributed carbon primarily through the glyoxylate shunt (ICL + MLS; $\sim 51\%$), a substantial fraction via the carnitine carrier (CRC; $\sim 29\%$), and the remainder to lipids via ACC ($\sim 18\%$) (Figure S10). At the malate branch, carbon flowed

into the TCA via PDH (~18%), through anaplerotic shunts (malate–OAA– α KG), or toward gluconeogenesis via cytosolic MDH. The MLS–MAE–PDH route was less preferred than the CRC route.

Predicted CIT flux ($67 \pm 2\%$) was ~3-fold higher than average TCA flux (~12%), which would be compatible with PDH → citrate export → ACL; yet ACL and PK remained at 0%, and *extra* exported citrate again re-entered the TCA cycle. Importantly, experimental data confirmed that a fraction of citrate was secreted extracellularly during exponential growth, indicating overflow metabolism. This suggests that citrate export on acetate serves dual purposes: (i) internal redox/OAA balancing via cycling and (ii) relieving metabolic pressure by secreting excess carbon when ACS-TCA-glyoxylate fluxes saturate. Proteomics supported this view: CIT, ACO1, and MDH1 were highly abundant and moderately regulated, consistent with reinforced internal TCA/glyoxylate cycling for OAA and redox balance during acetate assimilation. However, the absence of ACL flux despite active citrate export indicates that cytosolic acetyl-CoA is supplied exclusively by ACS, while citrate overflow represents a redox/energy balancing strategy rather than a lipogenic route.

Where NADPH is spent (turnover analysis)

During lipid accumulation on glucose, ~75% of NADPH turnover was directed into FAS1-2 and ~12% into NADP-GDH (Figure 6). In exponential growth, the split shifted to ~13% (FAS1-2) and ~46% (GDH). Thus, NADPH allocation switches from biomass synthesis (via GDH-mediated amino acid production) to storage compound formation (via FAS1-2) as cells transition from growth to lipid accumulation, reflecting a general growth-versus-storage trade-off. The magnitude of this reallocation (up to 75% to FAS) appears characteristic of oleaginous yeasts. While not resolving whether NADPH limits lipogenesis in *R. toruloides*, these results pinpoint where NADPH is predominantly spent and identify FAS as the dominant sink under lipid-accumulating conditions.

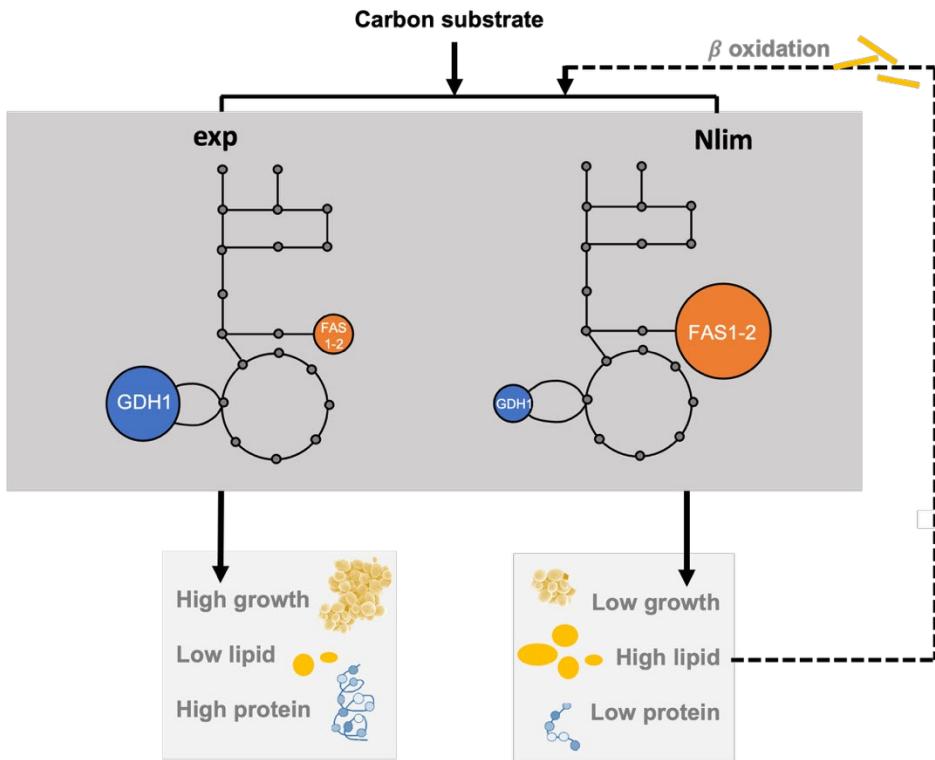


Figure 6. Comparison of NADPH expenditure on glucose across metabolism during exponential (exp) and nitrogen-limited (Nlim) phases in *R. toruloides*. Blue: nitrogen assimilation pathways. Orange: lipid biosynthetic pathways. Large orange circle represents 75 % of NADPH turnover, small orange circle is 13 %. Large blue circle represents 46 % of NADPH turnover, small blue circle is 12 %. GDH1: glutamate dehydrogenase (NADP-dependent), FAS1-2: fatty acid synthase. Reproduced from Rekēna et al., *PLOS Comput. Biol.* 19(4):e1011009, 2023, under a CC BY licence.

NADPH regeneration and the LXR cofactor switch

On glucose, oxPPP carried $76 \pm 43\%$ of carbon; on xylose $42 \pm 32\%$. These relative contributions changed little under Nlim. Alternative NADPH routes (cMAE, IDP) were minor on fermentative substrates ($\leq 9\%$ on glucose; $\leq 3\%$ on xylose), suggesting the oxPPP as the dominant source of NADPH regeneration, which is consistent with previous studies in *Y. lipolytica* (Wasylewko et al., 2015; J. Zhu et al., 2023). On acetate, oxPPP was essentially inactive, and cMAE carried the largest predicted NADPH share ($18 \pm 35\%$, decreasing to $15 \pm 30\%$ in Nlim), although proteome data did not conclusively support this, leaving the acetate NADPH source unresolved.

Cofactor specificity of L-xylulose reductase (LXR) strongly influenced xylose predictions. With NADP⁺/NADPH-dependence (Verho et al., 2004), oxPPP demand was lower ($42 \pm 32\%$), consistent with NADPH drainage by LXR. Switching LXR to NAD⁺/NADH-dependence increased oxPPP flux to $83 \pm 18\%$ on xylose, while other fluxes remained unchanged. The ribulose-mediated xylose pathway was preferred even without proteome constraints on XK, underscoring that cofactor choice is a critical, testable lever for redox design. A full dataset of flux values is available online: <https://doi.org/10.1371/journal.pcbi.1011009.s013>.

Together, the simulations outline substrate-dependent strategies in *R. toruloides*: (i) on glucose/xylose, PK + oxPPP dominate (with ACL acting as a conditional backup when PK is unavailable); (ii) on acetate, ACS is obligatory, with reinforced internal TCA/glyoxylate cycling and cMAE predicted for NADPH (pending validation); and (iii) citrate export – both cytosol and extracellularly – recurs across conditions without ACL activity *in silico*, plausibly due to ATP cost, compartmentation gaps, or missing regulation in the model. Given the large SDs for some routes, these are hypotheses rather than final calls and should be resolved by knockout phenotyping, or direct measurements (e.g., cofactor ratios, isotope tracing, enzyme assays)

3.2.4 Knockout engineering demonstrates ACK, PK, and cMAE are dispensable for growth but differ in metabolic impact

We validated model predictions by engineering the first single gene knockouts of ACL, PK and cMAE in haploid *R. toruloides* strain NBRC 0880 using a one-step plasmid-based CRISPR/Cas9 strategy that relied on frameshift mutations. Since no episomal plasmids are available for this yeast, Cas9, gRNA, and the G418 resistance marker were randomly integrated into the genome via plasmid pPBO.202 (JBEI223791) (Otoupal et al., 2019). Given the low homologous recombination frequency in *R. toruloides* (Koh et al., 2014), knockouts were likely obtained through Cas9-induced double-stranded breaks followed by error-prone non-homologous end joining. Successful gene disruption was verified by sequencing genomic regions flanking the gRNA cut sites and aligning them against the wild-type reference genome (Figure 7A). This yielded multiple mutant variants per gene, from which one representative per target was selected for detailed physiological characterization (for full details see **Publication III**). Growth assays confirmed that integration of the CRISPR/Cas9 cassette itself did not reduce fitness, as wild-type and cassette-containing control strains exhibited comparable growth rates, consistent with earlier findings (Otoupal et al., 2019). This provides a baseline for attributing phenotypic changes specifically to the targeted knockouts. Indeed, no significant growth differences were observed among PK and cMAE knockout variants during *initial* screening, whereas the ACL knockout displayed a distinct phenotype (Figure 7B). These results demonstrated that central carbon metabolism genes in *R. toruloides* can be disrupted using a chromosomal integration strategy with a selectable drug marker. The recovery of viable knockout strains carrying frameshift mutations confirms that the targeted genes ACL, PK and cMAE are not essential for growth under the tested cultivation conditions. However, conditional essentiality cannot be excluded, as gene function may vary depending on environmental context.

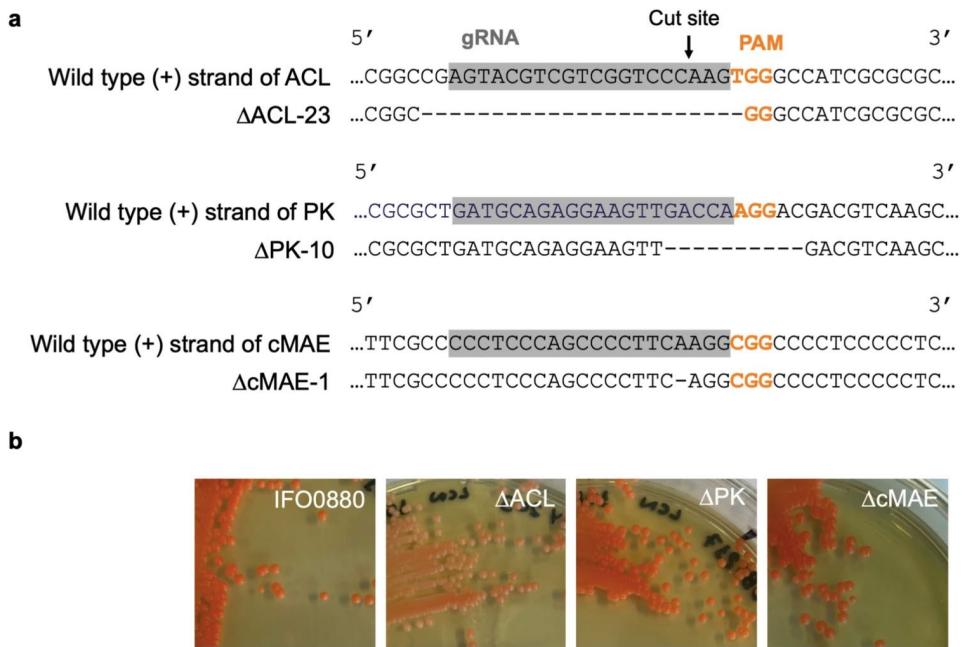


Figure 7. Confirmation of CRISPR/Cas9 based gene disruption by introducing frameshift mutations in *R. toruloides* NBRC 0880. **A:** partial gene sequencing of the targets ATP citrate lyase (ACL), phosphoketolase (PK) and cytosolic malic enzyme (cMAE) near the cut site. Example of one edited colony per gene target is shown. **B:** *R. toruloides* strains after 9 days growth on YPD agar – phenotype comparison. Reproduced from Rekēna et al., Appl Microbiol Biotechnol 109:77, 2025, under a CC BY NC ND licence.

3.2.5 Knockout phenotyping validates and challenges ecGEM predictions in *R. toruloides*

To validate model predictions, we compared lipid content, growth rates, and exchange fluxes of ACL, PK, and cMAE knockout strains with the wild type in sensor-monitored aerobic batch cultures. These cultivations were performed in low-nitrogen media to induce lipid accumulation. The multi-channel bioreactor system employed a reverse-spin mixing mechanism, ensuring efficient aeration and biomass formation – critical for the strictly aerobic *R. toruloides* – and enabling high-throughput, parallel strain characterization under reproducible, though less tightly controlled, conditions optimized for comparative screening. Only on acetate, the pH rose steadily during growth and reached ~9 by the end of cultivation (Figure S11). Experiments confirmed that integration of the CRISPR/Cas9 cassette did not impair growth, as several knockouts grew comparably to wild type. Distinct phenotypes can therefore be attributed to gene disruption.

ΔACL reveals an indispensable route for cytosolic acetyl-CoA and OAA supply

ACL knockout produced the strongest phenotypes among all mutants. Lipid accumulation was reduced across all substrates ($apval \leq 0.02$), most dramatically on glucose where lipid content dropped by -81% ($apval < 0.0001$) (Figure 8A). On xylose, the effect was milder (377 vs. 426 mg gCDW $^{-1}$ in wild type, $apval = 0.02$) (Figure 8B), suggesting pentose metabolism can partially bypass ACL by providing alternative cytosolic acetyl-CoA routes.

Lipid composition also altered: ACL knockout caused an increase in polyunsaturated fatty acids (PUFAs) (**Figure 8A-C**), mutants accumulated more linoleic acid (C18:3) ($p\text{val} < 0.0001$) and less oleic acid (C18:1) ($p\text{val} < 0.0001$) on glucose and acetate (**Table S6**), consistent with earlier findings in *Y. lipolytica* (Dulermo et al., 2015). Growth mirrored these trends: on glucose ACL mutants grew 41% slower ($p\text{val} < 0.002$) (**Figure 8D**) and arrested after ~ 18 h despite residual glucose and optimal pH (**Figure S12**, **Figure S11**). On xylose, they displayed a shorter lag phase compared to the wild type (**Figure S12**) but reached only 45% of the wild-type growth rate ($p\text{val} < 0.01$) (**Figure 8E**). Notably, they still consumed all substrate and accumulated more biomass, indicating that compensatory routes can support growth, albeit at reduced efficiency. These results, together with high ACL abundance and upregulation in the proteome for the wild type, confirm ACL as a central supplier of cytosolic acetyl-CoA during growth and lipogenesis on fermentative substrates, consistent with earlier works in other microorganisms (H. Chen et al., 2014; Dulermo et al., 2015; Hynes & Murray, 2010; Nowrouzian et al., 1999). The pervasive growth defects of ACL mutants also underscore its broader role in providing cytosolic OAA via citrate cleavage, as anticipated from PK deletion simulations. On glucose, the absence of ACL leaves cells without sufficient OAA compensation, leading to arrest. By contrast, on xylose, elevated PPP flux into glycolysis and buffering by PC – facilitated by high PC abundance and lower ATP demand at slower growth – can partially sustain OAA regeneration, allowing growth to continue, though at a reduced rate. If ACL were active, cytosolic OAA is produced stoichiometrically with acetyl-CoA, which would reduce the need for excessive MDH-mediated OAA regeneration and likely tighten those wide MDH1 distributions observed in predictions. Nonetheless, mitochondrial MDH1 remains essential to regenerate OAA and sustain the citrate-malate shuttle. Its high abundance on glucose as a carbon source, together with pyruvate carboxylation, indicates reinforced anaplerosis and redox buffering under high glycolytic flux, consistent with proteomic evidence.

On acetate, ACL mutants were most impaired, showing a ~ 60 h lag, 40% slower growth ($p\text{val} = 0.009$) (**Figure 8F**), and premature arrest before carbon substrate depletion (**Figure S12**). This severe phenotype contrasted with the wild type proteome (low ACL abundance: $148 \pm 10 \mu\text{g g}^{-1}$, 2.8-fold increase, $p\text{val} = 0.14$) and the ecGEM, which predicted zero ACL flux in the wild type. Proteomics for the wild type provided additional context: CIT, ACO1, and MDH1 were highly abundant, suggesting reinforced citrate-isocitrate-malate cycling to regenerate OAA and maintain redox balance. However, this internal cycling does not resolve the cytosolic acetyl-CoA demand for lipogenesis. The lack of ACL flux in the ecGEM, despite the severe knockout phenotype on acetate, reflects inherent model simplifications in energy accounting and compartmentation. First, ACL is an ATP-consuming reaction, and the model solution likely prioritizes energy-efficient pathways. With acetate assimilation modeled entirely through cytosolic ACS, which provides acetyl-CoA at lower energetic cost, the model solution bypasses ACL even when this contradicts biology. Second, many yeasts, including *S. cerevisiae*, *Y. lipolytica*, and *Candida* spp., encode both cytosolic and mitochondrial ACS isoforms (Krivoruchko et al., 2015), but in *R. toruloides* these are not well annotated. *In vivo*, mitochondrial ACS would convert acetate to acetyl-CoA for the TCA cycle, while cytosolic acetyl-CoA would be regenerated via citrate export and ACL cleavage. In the ecGEM, however, acetate is assumed to freely enter the cytosol, and cytosolic ACS therefore becomes the cheapest assimilation route, bypassing the need for citrate shuttling and ACL activity. Moreover, transport penalties and carrier coupling

(e.g., citrate-malate exchange) are not fully specified, so the energetic and regulatory advantages of the mitochondrial route are invisible to the solver. As both ACS isoforms are annotated to the same gene, the enzyme constraint cannot distinguish between cytosolic and mitochondrial pools, further biasing the solution. Together, these simplifications explain why the mitochondrial ACS is present in the network but not used in a biologically realistic way. Third, in this scenario, ACL becomes essential for exporting citrate to the cytosol and regenerating acetyl-CoA and OAA, thereby sustaining lipid synthesis and redox balance. Without ACL, this shuttle collapses: acetyl-CoA remains trapped in mitochondria, cytosolic OAA supply drops, and lipid synthesis stalls. PC cannot compensate under these conditions. Although PC abundance on acetate was comparable to glucose and xylose, glycolysis is inactive, so pyruvate is scarce. This substrate limitation caps PC flux and prevents it from meeting cytosolic OAA demand. ACL, by contrast, bypasses glycolysis and directly provides both cytosolic acetyl-CoA and OAA, uniquely positioning it to balance lipid synthesis and redox requirements. If ACL were active, exported citrate would be cleaved by cytosolic ACL to yield acetyl-CoA and OAA, directly supplying the cytosolic acetyl-CoA pool for lipogenesis while providing a parallel source of OAA that can re-enter the TCA cycle via malate shuttling. Although the ACL step itself is cataplerotic, the resulting citrate-malate cycle contributes to overall anaplerotic and redox balancing functions, reducing dependence on MDH- and PC-mediated OAA regeneration. This not only makes the proteome signal functionally meaningful rather than purely “internal cycling”, but also explains why ACL mutants arrest on acetate despite ACL’s low proteome abundance and the model’s zero predicted flux. Thus, the discrepancy between phenotype (ACL essential), proteomics (low ACL abundance), and modeling (zero flux) can be explained by (i) compartmentation errors, (ii) a small ACL pool providing essential flux despite low abundance, and (iii) acetate-specific where glycolysis-independent OAA supply is required. Together, these factors suggest that ACL plays a context-dependent but indispensable role under acetate metabolism, highlighting need for improved isoform annotation, compartmentation, and energy accounting in ecGEMs, as well as the limitations of using proteome abundance or model predictions alone to infer enzyme indispensability.

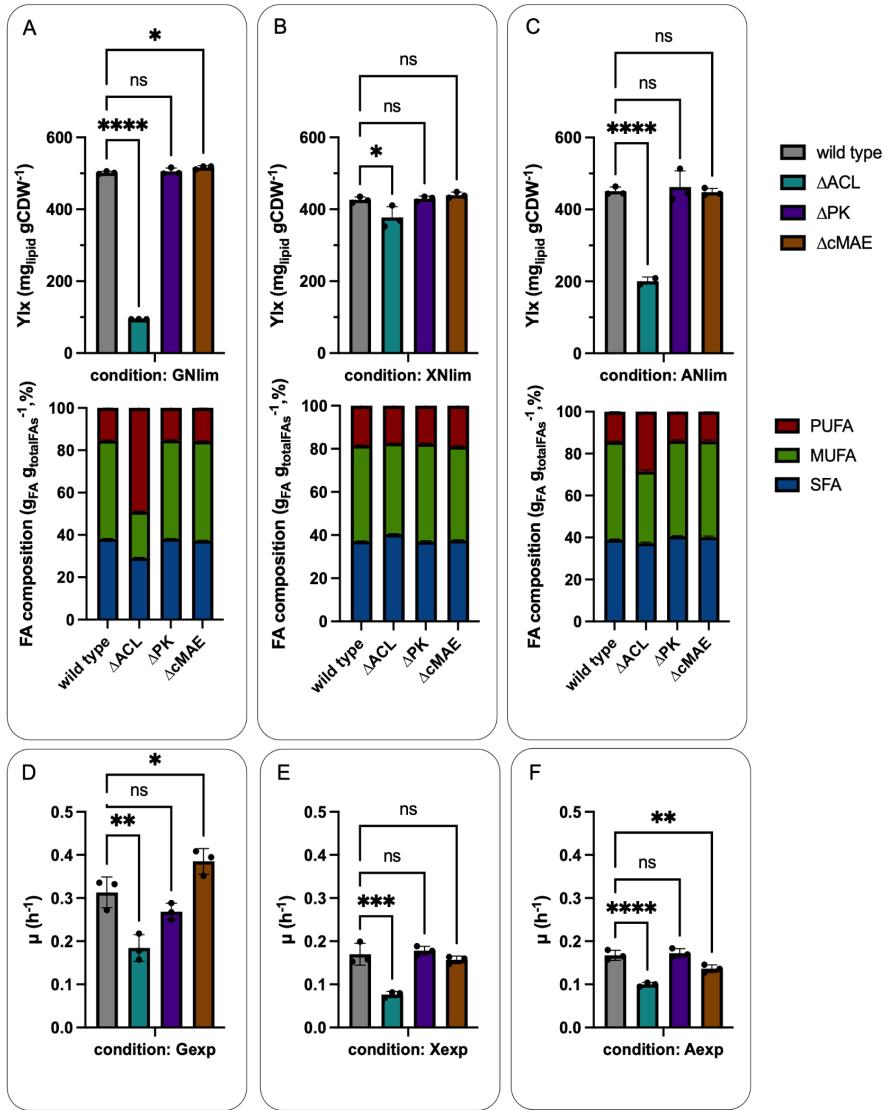


Figure 8. Growth and lipid accumulation data of *R. toruloides* NBRC 0880 (formerly known as IFO 0880) wild type and mutant strains cultivated in Falcon tube bioreactors in low-nitrogen chemically defined media with glucose (G), xylose (X), or acetate (A) as a carbon source during exponential growth (exp) and nitrogen-limited (Nlim) phases. **A:** Total lipid yield on biomass and FA composition of glucose-grown strains. **B:** Total lipid yield on biomass and FA composition of xylose-grown strains. **C:** Total lipid yield on biomass and FA composition of acetate-grown strains. **D:** Specific growth rate of glucose-grown strains. **E:** Specific growth rate of xylose-grown strains. **F:** Specific growth rate of acetate-grown strains. Strains include ATP citrate lyase knockout (ΔACL , green), phosphoketolase knockout (ΔPK , blue), and cytosolic malic enzyme knockout (ΔcMAE , orange). Error bars represent standard deviation of three biological replicates. Abbreviations: FA, fatty acid; SFAs, saturated fatty acids; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids. Asterisks denote statistical significance (ANOVA Dunnett's multiple comparison test against the wild type NBRC 0880 strain, adjusted p value * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). ns is used to denote changes that are statistically nonsignificant. Layouts and color schemes differ from the published versions for copyright compliance.

ΔPK indicates overestimation of phosphoketolase flux in silico

Although the ecGEM consistently predicted PK as the dominant cytosolic acetyl-CoA supplier on glucose and xylose, this was not borne out experimentally. PK knockout strains showed no significant changes in lipid content, growth rate, or fatty acid profile under any of tested conditions (Figure 8A–F), indicating that PK flux is dispensable *in vivo*. The model's preference for PK likely stems from its stoichiometric “cheapness” – a short, ATP-efficient route to acetyl-CoA – and possibly high k_{cat} values inherited from non-*R. toruloides* enzymes in BRENDA. This interpretation is consistent with ATP balance analysis, where ACL cleavage would appear as a costly ATP sink, while PK routes minimize energetic demand (Figure S13). Under nitrogen limitation on glucose/xylose, ATP consumption shifted toward ACC/FAS, mirroring the observed increase in NADPH flux to FAS and the rise in PK flux *in silico* (Figure 9). On acetate, as discussed above, ATP spending is dominated by ACS-driven assimilation and respiratory ATP production, leaving little energetic slack (Figure S14). In respiratory metabolism, mitochondrial NADH (mNADH) generated by PDH and the TCA cycle serves as the principal electron donor for oxidative phosphorylation. Although the ACL pathway is experimentally established as the main source of cytosolic acetyl-CoA during lipid accumulation in fully respiratory oleaginous yeasts, the model did not predict significant ACL flux (Figure 9). The model occasionally predicted citrate export to the cytosol followed by re-import into the mitochondria through auxiliary pathways, effectively creating a futile citrate cycle. This behavior likely reflects the rigid ATP–NADH coupling imposed by the steady-state energy balance. *In vivo*, mitochondria can produce *more ATP than strictly needed* for growth under high respiratory capacity, providing surplus energy to drive lipid biosynthesis. In contrast, GEMs solve ATP balance algebraically – every mole of NADH oxidation must immediately meet ATP demand to satisfy GAM and NGAM under steady-state assumptions. Consequently, cytosolic ATP required for ACL activity is never “left over.” When citrate is exported from the mitochondria, it carries carbon atoms but no reducing equivalents. The NADH produced by PDH remains in the matrix to sustain respiration, while the cytosol must generate its own NADPH for fatty acid synthesis. This separation of carbon and redox fluxes renders the ACL route energetically costly and redox-imbalanced in stoichiometric models, even though real cells employ regulatory mechanisms to coordinate these compartments during lipid accumulation. The established AMP-mediated regulatory switch that triggers citrate overflow under nitrogen limitation (Papanikolaou & Aggelis, 2011a; Ratledge & Wynn, 2002b) is not represented in GEM. *In vivo*, a decline in AMP inhibits isocitrate dehydrogenase, promoting citrate accumulation that feeds ACL and lipid biosynthesis. In the model, however, AMP functions only as a metabolite within the ATP - ADP - AMP balance, lacking regulatory role. As a result, ATP/NADH turnover remains tightly balanced, and the dynamic energy-charge control that drives this metabolic transition is not captured.

Also, *in vivo* PK may be constrained by (i) instability of its product acetyl-phosphate, (ii) insufficient activity of downstream PTA/ACK enzymes, or (iii) lack of integration with OAA/redox balancing that ACL provides via citrate cleavage. Proteomic data showing PK upregulation under nitrogen limitation suggest regulatory priming, but the absence of a phenotype upon deletion implies that this induction does not translate into substantial flux. This discrepancy underscores a broader GEM limitation: pathway selection *in silico* is driven by stoichiometric and energetic efficiency, but real cells prioritize robustness and metabolic integration. Refining models with constraints on PK (PTA/ACK capacity, acetyl-P instability, or thermodynamic penalties) would help avoid overestimating its

contribution and better reconcile model predictions with experimental evidence. These results highlight how knockout phenotyping, when integrated with proteome data, provides a critical check on ecGEM predictions and helps identify where additional biological realism is required.

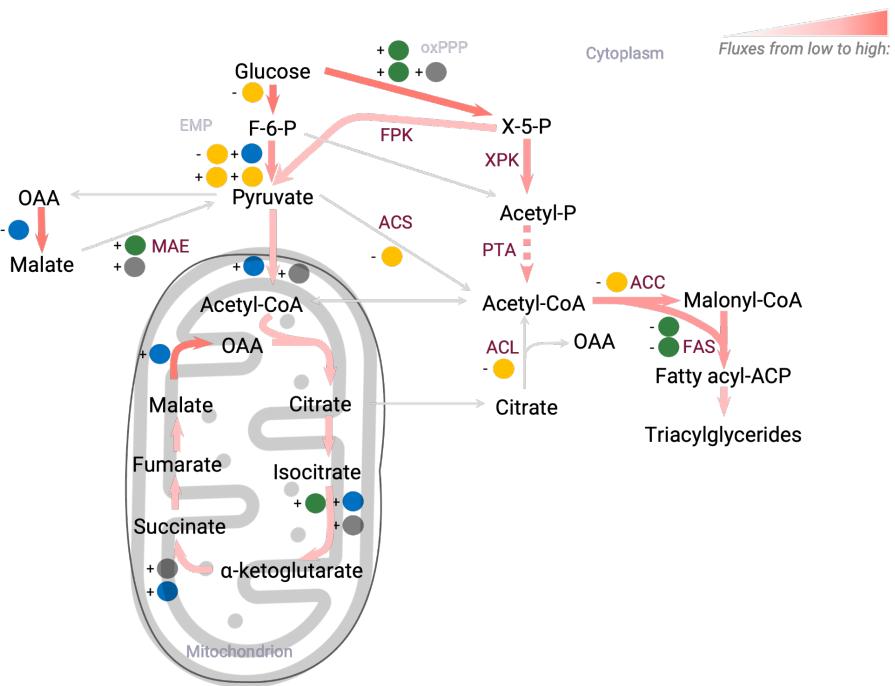


Figure 9. Central carbon and cofactor flux map in *R. toruloides* during respiratory growth on glucose. Nodes represent major intermediates (glucose, pyruvate, oxaloacetate (OAA), xylulose-5-phosphate (X-5-P), acetyl-CoA, and citrate) and compartments (cytoplasm, mitochondria). Reaction arrows are scaled by simulated flux and annotated with cofactor usage: ATP (amber badge), NADH (blue badge), NADPH (green badge), and CO₂ release (gray badge). In glycolysis, cofactor badges indicate per-reaction stoichiometries (-2 ATP at hexokinase and phosphofructokinase, +1 NADH at glyceraldehyde-3-phosphate dehydrogenase, +2 ATP at phosphoglycerate kinase and pyruvate kinase) rather than the overall glycolytic net yield (+2 ATP, +2 NADH per glucose). Color-coded cofactor badges summarize the energetic and redox logic underlying this predicted flux distribution.

***ΔcMAE* confirms substrate-dependent and nonessential role in NADPH supply**

For NADPH supply, all three approaches consistently pointed to the oxPPP as the major source on fermentative carbon sources. cMAE knockout displayed condition-specific effects. On acetate, growth slowed by 19% (apval = 0.009) (Figure 8F) without affecting lipid accumulation, consistent with model predictions that cMAE contributes to NADPH supply under this condition. The absence of a lipid phenotype on acetate (lipid yields of 448 vs. 451 mg gDCW⁻¹ for the *ΔcMAE* mutant and the wild type, respectively; difference not statistically significant, Figure 8C) suggests that remaining NADPH sources were preferentially directed into fatty acid synthesis at the expense of growth, reflecting a lipid-first allocation strategy typical of oleaginous yeasts. This observation is consistent with the early biochemical findings in *R. glutinis* (Yoon et al., 1984). Although recent genetic engineering studies in oleaginous yeasts have been performed only on glucose,

the reported effects of MAE deletion or overexpression on lipogenesis were relatively minor and varied depending on experimental design (Dulermo et al., 2015; S. Zhang et al., 2016). This contrasts with filamentous fungi, in which MAE has been identified as one of the rate-limiting steps in fatty acid biosynthesis (Wynn et al., 1997, 2001; Wynn & Ratledge, 1997; Y. Zhang et al., 2007). The proteomics-constrained ecGEM consistently favored cMAE as the dominant NADPH source on acetate ($\sim 18 \pm 35\%$ of carbon, \downarrow to $15 \pm 30\%$ at Nlim) because it provides a direct, ATP-neutral conversion of malate to pyruvate while generating NADPH, making it a stoichiometrically and energetically “cheap” solution compared to oxPPP (which requires costly gluconeogenic generation of G6P) or IDP (which consumes isocitrate and produces CO_2). The solver minimizes energetic cost and carbon loss, biasing predictions toward cMAE. However, this predicted dominance of cMAE is biologically unlikely. Malate availability in the cytosol is limited on acetate because flux through glycolysis is absent and malate must be generated via costly cycling through the glyoxylate shunt or TCA. Moreover, cMAE activity uncouples NADPH regeneration from OAA balance, whereas oxPPP and IDP are more integrated into central carbon and redox metabolism. Thus, *in vivo*, cMAE is more plausibly a secondary or backup source of NADPH rather than the main driver. Proteomics, however, revealed a substantial and even upregulated abundance of ZWF ($637 \pm 52 \rightarrow 951 \pm 16 \mu\text{g g}^{-1}$ protein), suggesting that cells maintain significant oxPPP capacity even when acetate is the sole carbon source. The most plausible explanation is that oxPPP on acetate is constrained by gluconeogenic supply of G6P – energetically costly, but still feasible – so ZWF may act as a redox reserve rather than a high-flux NADPH route. Meanwhile, IDP, which showed elevated abundance on acetate ($1472 \pm 47 \mu\text{g g}^{-1}$ protein, 2.4-fold increase, apval 0.14), represents an alternative, model-underestimated source of NADPH that could compensate when cMAE is disrupted. Together, these results highlight that NADPH supply during acetate metabolism in *R. toruloides* cannot be attributed to a single pathway. Instead, a distributed system of partially redundant routes – oxPPP, IDP, and possibly transhydrogenase-like cycling – appears to balance redox demands. This points to gaps in the ecGEM, including, e.g., gluconeogenic limits, thermodynamic penalties, enzyme saturation, inaccurate k_{cat} values or missing alternative NADPH sources, and underscores the need for isotope tracing and cofactor profiling to resolve acetate-specific NADPH metabolism.

By contrast, on glucose, the cMAE knockout unexpectedly improved performance: growth increased by 23% (apval = 0.04) (**Figure 8D**) and lipid accumulation modestly rose ($+2.9\% \text{ g gDCW}^{-1}$, apval = 0.03) (**Figure 8A**). This suggests that cMAE activity on glucose may impose a metabolic burden, either by drawing malate away from malate-OAA cycling or by generating redundant NADPH in a condition where the oxPPP already dominates redox supply ($76 \pm 43\%$ of carbon flux). Proteomics data supported this interpretation, showing cMAE at modest, unregulated levels across conditions. Together, these results indicate that cMAE is a supportive but nonessential enzyme: beneficial on acetate, but potentially maladaptive on glucose where excess NADPH or futile cycling can hinder growth efficiency. This contrasts with previous report in *R. toruloides*, where MAE overexpression, although producing only a relatively minor effect, still resulted in a significant increase in lipid production (S. Zhang et al., 2016). It is important to note, however, that knockout and overexpression phenotypes are perturbations in opposite directions and may not be directly comparable.

In summary, knockout phenotyping established ACL as essential for lipogenesis on glucose and acetate but dispensable on xylose, where alternative acetyl-CoA pathways compensate. PK appears redundant despite model predictions and proteome upregulation, while cMAE contributes conditionally in a substrate- and redox-dependent manner, even disfavouring lipid production. Integrating proteomics, modeling, and knockout phenotyping thus refines our understanding: ACL is indispensable *in vivo*, PK is overestimated *in silico*, and cMAE is largely dispensable but reflects metabolic flexibility.

4 Conclusions

Potential and limitations of enzyme-constrained genome-scale models (ecGEMs) in understanding and engineering yeast metabolism were examined focusing on two distinct studies: (i) anaerobic 2,3-butanediol (23BD)-glycerol co-production in *Saccharomyces cerevisiae*, and (ii) lipogenesis in the oleaginous yeast *Rhodotorula toruloides*. By integrating ecGEM simulations with quantitative proteomics, growth physiology, and CRISPR/Cas-based strain engineering, we demonstrated how modeling and experiments can be combined to reveal metabolic strategies, bottlenecks, and design opportunities.

Broader conclusions and future perspectives

The predictive value of enzyme-constrained genome-scale models (ecGEMs) ultimately depends on how accurately simulated quantities reproduce experimental observations. It is therefore necessary to distinguish calibration (parameterization with exchange fluxes, biomass composition, and proteomics) from model validation, in which independent phenotypes, fluxes, and proteome allocation patterns are quantitatively tested. Validation provides the critical link between theoretical modeling and biological experimentation, establishing when a model can be relied upon for design or hypothesis generation. Importantly, inconsistencies between predictions and measurements also highlight gaps in our current understanding of cellular metabolism, pointing to specific biological mechanisms that call for further investigation.

The studies described in the thesis demonstrate how enzyme-constrained genome-scale modeling, when integrated with quantitative proteomics and genetic perturbations, can bridge the gap between systems-level prediction and experimental validation in both model and non-model yeasts. The need for flexibilization of measured enzyme data in the ecGEM fitting process carries important implications for both biological interpretation and predictive credibility. The parametrization process emphasises that ecGEMs are not static predictive machines but iterative diagnostic tools – each round of adjustment refines both experimental and computational understanding. In this sense, flexibilization is not only an alarming signal, but also an informative point, pinpointing where future work should focus.

Nevertheless, by quantifying proteome allocation dynamics and demonstrating that enzyme constraints reshape feasible flux space, this thesis established ecGEMs as promising platforms for rational strain design and for interpreting adaptive evolution (**Aim I**). Moreover, across *S. cerevisiae* and *R. toruloides*, ecGEMs proved capable of reproducing major physiological trade-offs – between growth, redox balance, and resource allocation – while generating mechanistic hypotheses far faster and at greater resolution than traditional experimentation alone (**Aim II**). These insights enabled the identification of substrate-dependent strategies for cytosolic acetyl-CoA and NADPH supply, revealed context-specific enzyme essentialities, but also exposed key discrepancies between *in silico* efficiency and *in vivo* robustness. Importantly, this work also emphasized that not always enzyme abundance is a reliable proxy for flux, underscoring the necessity of coupling modeling with isotope tracing, cofactor profiling, and genetic perturbations to refine quantitative predictions. Future advances – including automatic biomass composition, ribosome allocation scaling with growth rate, improved enzyme annotation, thermodynamics-based constraints, organism-specific energy maintenance measurements and kinetic datasets – will further enhance their predictive power

(Aim III). Together, ecGEMs, as part of an integrative cycle with experiments, have the potential to shorten design-build-test-learn cycles, improve the efficacy of metabolic engineering experiments, and deepen our mechanistic understanding of cellular metabolism under both engineered and natural conditions.

Study 1: Anaerobic 23BD-glycerol co-production in *S. cerevisiae*

In this study, enzyme-constrained genome-scale modeling was integrated with quantitative proteomics to dissect metabolic trade-offs and proteome allocation in *Saccharomyces cerevisiae* engineered for anaerobic (R,R)-2,3-butanediol–glycerol co-production. The model successfully reproduced key physiological behaviors and provided mechanistic insight into the interplay between enzyme capacity, energy yield, and proteome organization.

The ecGEM captured the major trade-offs of 2,3-butanediol pathway operation – namely, a lower ATP yield and a corresponding reduction in biomass yield – while predicting compensatory increases in specific glucose uptake rate that offset the thermodynamic penalties on growth (Section 3.1.1). These predictions aligned closely with experimental observations (Section 3.1.2), confirming the model’s ability to describe key constraints in anaerobic co-production (Aim II).

Comparisons between predicted and measured proteome allocation further revealed both strengths and limitations of the ecGEM (Section 3.1.3). The predicted increase in glycolytic investment and its experimental confirmation underscored the remarkable metabolic flexibility of *S. cerevisiae* under redox stress – a phenomenon unattainable in conventional GEMs that lack proteome accounting. Under detailed enzyme constraints, fluxes through the most catalytically efficient enzymes (as defined by k_{cat}/Mw) often decreased, while flux through less efficient isoenzymes or alternative routes increased (Section 3.1.4), revealing a shift toward a leaner and more physiologically realistic metabolic strategy. Collectively, these results guided phenotype interpretations that would be difficult to derive through experiments alone. However, discrepancies at the enzyme-family level – particularly in redox balancing (Gpd/Gpp, Ald), biosynthetic regulation, and isoenzyme selection – reflected the model’s bias toward stoichiometric efficiency rather than the robustness and redundancy observed *in vivo* (Aim III). By incorporating the both coarse total-protein and proteomics- constraints (Section 3.1.1), only a fraction of total proteome mass contributed catalytically in the model due to incomplete annotation, missing or uncertain k_{cat} values, and the presence of structural, regulatory, or stress-related proteins that are not explicitly represented (Aim III).

Overall, this work demonstrates that enzyme-constrained modeling, when paired with absolute proteomics, can accurately predict macroscopic trade-offs and capture global proteome redistribution, while highlighting critical gaps in enzyme annotation and kinetic data. The ecGEM framework thus provides a mechanistically grounded context for evaluating adaptive limits in metabolic engineering: distinguishing “hard limits” imposed by thermodynamics and ATP yield from “soft limits” imposed by enzyme capacity and regulation.

Study 2: Lipogenesis in *R. toruloides*

In this study, we examined how central carbon metabolism reallocates fluxes to supply cytosolic acetyl-CoA and NADPH in *R. toruloides*, and how these strategies vary between substrates such as glucose, xylose, and acetate. Using growth experiments and detailed physiological characterisation, global absolute proteomics, ecGEMs with

substrate-resolved biomass composition, and targeted gene knockouts, we predicted intracellular fluxes across substrates, and validated predictions to dissect substrate-specific strategies between lipogenesis and exponential growth conditions.

Growth experiments on glucose, xylose, and acetate revealed nitrogen limitation consistently triggering lipid accumulation (Section 3.2.2) but through different metabolic routes. Absolute proteomics showed how resource allocation shifts between growth and storage (Section 3.2.1). Ribosomal investment was high but ribosome efficiency was low compared to *S. cerevisiae*, consistent with microbial growth laws but highlighting distinct proteome allocation in basidiomycetous oleaginous yeast. Key lipogenic enzymes such as ACL, PK, ACS, ZWF and cMAE displayed clear substrate-specific patterns, but abundance alone did not predict essentiality. The proteomics-constrained ecGEM reproduced growth rates and substrate uptake within ~30% of experimental values, and it systematically generated mechanistic hypotheses about acetyl-CoA and NADPH supply (Section 3.2.3). On glucose and xylose, the model consistently preferred the phosphoketolase (PK) pathway and the oxidative pentose phosphate pathway (oxPPP); on acetate, acetyl-CoA synthetase (ACS) and malic enzyme (cMAE) were predicted to dominate. Knockout phenotyping provided the decisive evidence to these predictions (Section 3.2.4 and 3.2.5). Results established ACL as a central node in lipogenesis, uncovered clear discrepancies between modeling and proteome data, and phenotypes, and highlighted the importance of integrating all three approaches (**Aim II**).

These findings refine ecGEM predictions and underscore the importance of isoform annotation, compartmentation, and energy accounting in non-model GEMs (**Aim III**). Nevertheless, the modeling insights proved especially valuable because they generated hypotheses that would have been difficult or slow to uncover using experimental approaches alone. For example, the model highlighted the PK pathway as a potential dominant source of cytosolic acetyl-CoA on glucose and xylose, narrowing the search space. Similarly, the simulations revealed that ATP-consuming ACL could act as a conditional backup, becoming active when PK was blocked *in silico*. This finding anticipated the strong ACL knockout phenotype on glucose/xylose. Accounting for cofactor NADPH regeneration prompted investigation into substrate-dependent NADPH strategies and led to the discovery that enzyme cofactor specificity (e.g., NADPH- vs NADH-dependent LXR) strongly shifts redox fluxes. Finally, the counterintuitive prediction of citrate export without ACL flux identified redox/energy balancing as a bottleneck and motivated the hypothesis that citrate shuttling plays roles in overflow and redox balancing. However, the model's energetic logic - favoring PK as the "cheaper" route - forced deeper reflection on why cells use ACL *in vivo*, pointing to integration with OAA and redox balance rather than energetic efficiency. Also, predicting cMAE dominance on acetate revealed its bias toward stoichiometric optimality, overestimating energy and carbon efficiency. These modeling discrepancies and their mechanistic interpretation were not examined in **Publication III** when it was published and are presented here for the first time, extending the original studies with new insights into enzyme essentiality and metabolic mechanisms in *R. toruloides*.

Taken together, the *R. toruloides* case study shows that ecGEMs can capture broad metabolic strategies and point to critical nodes, but that proteomics and phenotyping are essential to distinguish between "cheap but unused" routes and "low-abundance but indispensable" ones. Nevertheless, it elucidated how flexibly *R. toruloides* reorganizes metabolism to achieve lipogenesis, providing both mechanistic insight and guidance for metabolic engineering. In future work, ecGEMs could also be used for proteome

allocation modeling, linking expression constraints more directly to fluxes. Future improvements should include generation of *R. toruloides*-specific kinetic and maintenance parameters to fit GEMs, targeted testing of uncharacterized enzymes identified in the proteome. Such steps will bring us closer to a predictive design framework for engineering *R. toruloides* as a robust lipid cell factory.

References

Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198–207. <https://doi.org/10.1038/nature01511>

Alberts, B., Heald, R., Johnson, A., Morgan, D., Raff, M., Roberts, K., Walter, P., Wilson, J. H., & Hunt, T. (2022). *Molecular biology of the cell* (Seventh edition, international student edition). W. W. Norton & Company.

Bairoch, A. (1996). The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Research*, 24(1), 21–25. <https://doi.org/10.1093/nar/24.1.21>

Bar-Even, A., Noor, E., Savir, Y., Lieberman, W., Davidi, D., Tawfik, D. S., & Milo, R. (2011). The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry*, 50(21), 4402–4410. <https://doi.org/10.1021/bi2002289>

Benjamini, Y., & Hochberg, Y. (1995a). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Benjamini, Y., & Hochberg, Y. (1995b). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Bhosale, S. H., Rao, M. B., & Deshpande, V. V. (1996). Molecular and industrial aspects of glucose isomerase. *Microbiological Reviews*, 60(2), 280–300. <https://doi.org/10.1128/mr.60.2.280-300.1996>

Björkeroth, J., Campbell, K., Malina, C., Yu, R., Di Bartolomeo, F., & Nielsen, J. (2020). Proteome reallocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. *Proceedings of the National Academy of Sciences*, 117(35), 21804–21812. <https://doi.org/10.1073/pnas.1921890117>

Blazeck, J., Hill, A., Liu, L., Knight, R., Miller, J., Pan, A., Otoupal, P., & Alper, H. S. (2014). Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nature Communications*, 5(1), 3131. <https://doi.org/10.1038/ncomms4131>

Boehlke, K. W., & Friesen, J. D. (1975). Cellular content of ribonucleic acid and protein in *Saccharomyces cerevisiae* as a function of exponential growth rate: Calculation of the apparent peptide chain elongation rate. *Journal of Bacteriology*, 121(2), 429–433. <https://doi.org/10.1128/jb.121.2.429-433.1975>

Bonturi, N., Crucello, A., Viana, A. J. C., & Miranda, E. A. (2017). Microbial oil production in sugarcane bagasse hemicellulosic hydrolysate without nutrient supplementation by a *Rhodotoruloides toruloides* adapted strain. *Process Biochemistry*, 57, 16–25. <https://doi.org/10.1016/j.procbio.2017.03.007>

Bonturi, N., Pinheiro, M. J., De Oliveira, P. M., Rusadze, E., Eichinger, T., Liudžiūtė, G., De Biaggi, J. S., Brauer, A., Remm, M., Miranda, E. A., Ledesma-Amaro, R., & Lahtvee, P.-J. (2022). Development of a dedicated Golden Gate Assembly Platform (RtGGA) for *Rhodotorula toruloides*. *Metabolic Engineering Communications*, 15, e00200. <https://doi.org/10.1016/j.mec.2022.e00200>

Bordel, S., Agren, R., & Nielsen, J. (2010). Sampling the Solution Space in Genome-Scale Metabolic Networks Reveals Transcriptional Regulation in Key Enzymes. *PLoS Computational Biology*, 6(7), e1000859. <https://doi.org/10.1371/journal.pcbi.1000859>

Boyer, P. D. (1997). The ATP synthase—A splendid molecular machine. *Annual Review of Biochemistry*, 66(1), 717–749. <https://doi.org/10.1146/annurev.biochem.66.1.717>

Branduardi, P., Valli, M., Brambilla, L., Sauer, M., Alberghina, L., & Porro, D. (2004). The yeast *Zygosaccharomyces bailii*: A new host for heterologous protein production, secretion and for metabolic engineering applications. *FEMS Yeast Research*, 4(4–5), 493–504. [https://doi.org/10.1016/S1567-1356\(03\)00200-9](https://doi.org/10.1016/S1567-1356(03)00200-9)

Cammack, R. (Ed.). (2006). *Oxford dictionary of biochemistry and molecular biology* (Rev. ed.). Oxford University Press.

Celińska, E., & Grajek, W. (2009). Biotechnological production of 2,3-butanediol—Current state and prospects. *Biotechnology Advances*, 27(6), 715–725. <https://doi.org/10.1016/j.biotechadv.2009.05.002>

Chambers, P. J., & Pretorius, I. S. (2010). Fermenting knowledge: The history of winemaking, science and yeast research. *EMBO Reports*, 11(12), 914–920. <https://doi.org/10.1038/embor.2010.179>

Chawla, K., Kaur, S., Kaur, R., & Bhunia, R. K. (2022). Metabolic engineering of oleaginous yeasts to enhance single cell oil production. *Journal of Food Process Engineering*, 45(7), e13634. <https://doi.org/10.1111/jfpe.13634>

Chen, H., He, X., Geng, H., & Liu, H. (2014). Physiological characterization of ATP-citrate lyase in *Aspergillus niger*. *Journal of Industrial Microbiology and Biotechnology*, 41(4), 721–731. <https://doi.org/10.1007/s10295-014-1418-3>

Chen, Y., Zhang, Y., Siewers, V., & Nielsen, J. (2015). Ach1 is involved in shuttling mitochondrial acetyl units for cytosolic C2 provision in *Saccharomyces cerevisiae* lacking pyruvate decarboxylase. *FEMS Yeast Research*, 15(3). <https://doi.org/10.1093/femsyr/fov015>

Choi, M.-H., Kim, S.-J., Kim, J.-W., Park, Y.-C., & Seo, J.-H. (2016). Molecular cloning and expression of *Enterobacter aerogenes* α -acetolactate decarboxylase in pyruvate decarboxylase-deficient *Saccharomyces cerevisiae* for efficient 2,3-butanediol production. *Process Biochemistry*, 51(2), 170–176. <https://doi.org/10.1016/j.procbio.2015.11.023>

Crabtree, H. G. (1929). Observations on the carbohydrate metabolism of tumours. *Biochemical Journal*, 23(3), 536–545. <https://doi.org/10.1042/bj0230536>

Cregg, J. M., Cereghino, J. L., Shi, J., & Higgins, D. R. (2000). Recombinant protein expression in *Pichia pastoris*. *Molecular Biotechnology*, 16(1), 23–52. <https://doi.org/10.1385/MB:16:1:23>

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, 227(5258), 561–563. <https://doi.org/10.1038/227561a0>

Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, 12, 138–163.

Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology*, 38(3), 367–379. [https://doi.org/10.1016/0022-2836\(68\)90392-6](https://doi.org/10.1016/0022-2836(68)90392-6)

Cueto-Rojas, H. F., Van Maris, A. J. A., Wahl, S. A., & Heijnen, J. J. (2015). Thermodynamics-based design of microbial cell factories for anaerobic product formation. *Trends in Biotechnology*, 33(9), 534–546. <https://doi.org/10.1016/j.tibtech.2015.06.010>

Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., & Ralser, M. (2020). DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods*, 17(1), 41–44. <https://doi.org/10.1038/s41592-019-0638-x>

Dijken, J. P., & Scheffers, W. A. (1986). Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiology Letters*, 32(3–4), 199–224. <https://doi.org/10.1111/j.1574-6968.1986.tb01194.x>

Dixon, M., & Webb, E. C. (1958). *Enzymes* (1st Edition). Longmans Green, London, and Academic Press, New York.

Domenzain, I., Li, F., Kerkhoven, E. J., & Siewers, V. (2021). Evaluating accessibility, usability and interoperability of genome-scale metabolic models for diverse yeasts species. *FEMS Yeast Research*, 21(1), foab002. <https://doi.org/10.1093/femsyr/foab002>

Domenzain, I., Sánchez, B., Anton, M., Kerkhoven, E. J., Millán-Oropeza, A., Henry, C., Siewers, V., Morrissey, J. P., Sonnenschein, N., & Nielsen, J. (2022). Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nature Communications*, 13(1), 3766. <https://doi.org/10.1038/s41467-022-31421-1>

Dulermo, T., Lazar, Z., Dulermo, R., Rakicka, M., Haddouche, R., & Nicaud, J.-M. (2015). Analysis of ATP-citrate lyase and malic enzyme mutants of *Yarrowia lipolytica* points out the importance of mannitol metabolism in fatty acid synthesis. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1851(9), 1107–1117. <https://doi.org/10.1016/j.bbalip.2015.04.007>

Edwards, J. S., Covert, M., & Palsson, B. (2002). Metabolic modelling of microbes: The flux-balance approach. *Environmental Microbiology*, 4(3), 133–140. <https://doi.org/10.1046/j.1462-2920.2002.00282.x>

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767), 335–338. <https://doi.org/10.1038/35002125>

Elsemann, I. E., Rodriguez Prado, A., Grigaitis, P., Garcia Albornoz, M., Harman, V., Holman, S. W., Van Heerden, J., Bruggeman, F. J., Bisschops, M. M. M., Sonnenschein, N., Hubbard, S., Beynon, R., Daran-Lapujade, P., Nielsen, J., & Teusink, B. (2022). Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *Nature Communications*, 13(1), 801. <https://doi.org/10.1038/s41467-022-28467-6>

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1), 238. <https://doi.org/10.1186/s13059-019-1832-y>

Feist, A. M., & Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*, 13(3), 344–349. <https://doi.org/10.1016/j.mib.2010.03.003>

Ferguson, S. J. (2010). ATP synthase: From sequence to ring size to the P/O ratio. *Proceedings of the National Academy of Sciences*, 107(39), 16755–16756. <https://doi.org/10.1073/pnas.1012260107>

Gancedo, J. M. (1998). Yeast Carbon Catabolite Repression. *Microbiology and Molecular Biology Reviews*, 62(2), 334–361. <https://doi.org/10.1128/MMBR.62.2.334-361.1998>

García-Contreras, R., Vos, P., Westerhoff, H. V., & Boogerd, F. C. (2012). Why in vivo may not equal in vitro – new effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions. *The FEBS Journal*, 279(22), 4145–4159. <https://doi.org/10.1111/febs.12007>

Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767), 339–342. <https://doi.org/10.1038/35002131>

Garg, S. K., & Jain, A. (1995). Fermentative production of 2,3-butanediol: A review. *Bioresource Technology*, 51(2–3), 103–109. [https://doi.org/10.1016/0960-8524\(94\)00136-O](https://doi.org/10.1016/0960-8524(94)00136-O)

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343–345. <https://doi.org/10.1038/nmeth.1318>

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>

Grand View Research. (n.d.). *Bio-based Platform Chemicals Market (2025–2033)* (No. GVR-2-68038-793-3). <https://www.grandviewresearch.com/industry-analysis/bio-based-platform-chemicals-market>

Hagman, A., & Piškur, J. (2015). A Study on the Fundamental Mechanism and the Evolutionary Driving Forces behind Aerobic Fermentation in Yeast. *PLOS ONE*, 10(1), e0116942. <https://doi.org/10.1371/journal.pone.0116942>

Heckmann, D., Lloyd, C. J., Mih, N., Ha, Y., Zielinski, D. C., Haiman, Z. B., Desouki, A. A., Lercher, M. J., & Palsson, B. O. (2018). Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nature Communications*, 9(1), 5252. <https://doi.org/10.1038/s41467-018-07652-6>

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., ... Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, 14(3), 639–702. <https://doi.org/10.1038/s41596-018-0098-2>

Hellerstedt, S. T., Nash, R. S., Weng, S., Paskov, K. M., Wong, E. D., Karra, K., Engel, S. R., & Cherry, J. M. (2017). Curated protein information in the *Saccharomyces* genome database. *Database*, 2017. <https://doi.org/10.1093/database/bax011>

Hohmann, S. (2016). Nobel Yeast Research. *FEMS Yeast Research*, 16(8), fow094. <https://doi.org/10.1093/femsyr/fow094>

Huang, P.-S., Boyken, S. E., & Baker, D. (2016). The coming of age of de novo protein design. *Nature*, 537(7620), 320–327. <https://doi.org/10.1038/nature19946>

Hughes, C. S., Moggridge, S., Müller, T., Sorensen, P. H., Morin, G. B., & Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols*, 14(1), 68–85. <https://doi.org/10.1038/s41596-018-0082-x>

Hynes, M. J., & Murray, S. L. (2010). ATP-Citrate Lyase Is Required for Production of Cytosolic Acetyl Coenzyme A and Development in *Aspergillus nidulans*. *Eukaryotic Cell*, 9(7), 1039–1048. <https://doi.org/10.1128/EC.00080-10>

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

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Research*, 49(D1), D545–D551. <https://doi.org/10.1093/nar/gkaa970>

Karlebach, G., & Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10), 770–780. <https://doi.org/10.1038/nrm2503>

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., & Subhraveti, P. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, 20(4), 1085–1093. <https://doi.org/10.1093/bib/bbx085>

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3), 437–467. [https://doi.org/10.1016/0022-5193\(69\)90015-0](https://doi.org/10.1016/0022-5193(69)90015-0)

Kerkhoven, E. J., Lahtvee, P.-J., & Nielsen, J. (2014). Applications of computational modeling in metabolic engineering of yeast. *FEMS Yeast Research*, n/a-n/a. <https://doi.org/10.1111/1567-1364.12199>

Kim, S.-J., Kim, J.-W., Lee, Y.-G., Park, Y.-C., & Seo, J.-H. (2017). Metabolic engineering of *Saccharomyces cerevisiae* for 2,3-butanediol production. *Applied Microbiology and Biotechnology*, 101(6), 2241–2250. <https://doi.org/10.1007/s00253-017-8172-1>

Klipp, E. (2016). *Systems Biology: A Textbook* (2., vollst. überarb. und erw. Aufl). Wiley.

Koh, C. M. J., Liu, Y., Moehninsi, Du, M., & Ji, L. (2014). Molecular characterization of KU70 and KU80 homologues and exploitation of a KU70-deficient mutant for improving gene deletion frequency in *Rhodosporidium toruloides*. *BMC Microbiology*, 14(1), 50. <https://doi.org/10.1186/1471-2180-14-50>

Korosuo, A., Borzacchiello, M. T., Giuntoli, J., Lasarte Lopez, J., M'barek, R., Mubareka, S. B., Camia, A., & European Commission (Eds.). (2024). *Trends in the EU bioeconomy: Update 2024*. Publications Office. <https://doi.org/10.2760/0141556>

Koutinas, A. A., Vlysidis, A., Pleissner, D., Kopsahelis, N., Lopez Garcia, I., Kookos, I. K., Papanikolaou, S., Kwan, T. H., & Lin, C. S. K. (2014). Valorization of industrial waste and by-product streams via fermentation for the production of chemicals and biopolymers. *Chemical Society Reviews*, 43(8), 2587. <https://doi.org/10.1039/c3cs60293a>

Krivoruchko, A., Zhang, Y., Siewers, V., Chen, Y., & Nielsen, J. (2015). Microbial acetyl-CoA metabolism and metabolic engineering. *Metabolic Engineering*, 28, 28–42. <https://doi.org/10.1016/j.ymben.2014.11.009>

Kruger, N. J., & Von Schaewen, A. (2003). The oxidative pentose phosphate pathway: Structure and organisation. *Current Opinion in Plant Biology*, 6(3), 236–246. [https://doi.org/10.1016/S1369-5266\(03\)00039-6](https://doi.org/10.1016/S1369-5266(03)00039-6)

Kurtzman, C. (with Fell, J. W., & Boekhout, T.). (2011). *The Yeasts: A Taxonomic Study* (5th ed). Elsevier Science & Technology.

Kwak, S., Jo, J. H., Yun, E. J., Jin, Y.-S., & Seo, J.-H. (2019). Production of biofuels and chemicals from xylose using native and engineered yeast strains. *Biotechnology Advances*, 37(2), 271–283. <https://doi.org/10.1016/j.biotechadv.2018.12.003>

Lange, H. C., & Heijnen, J. J. (2001). Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, 75(3), 334–344. <https://doi.org/10.1002/bit.10054>

Ledesma-Amaro, R., & Nicaud, J.-M. (2016). Metabolic Engineering for Expanding the Substrate Range of *Yarrowia lipolytica*. *Trends in Biotechnology*, 34(10), 798–809. <https://doi.org/10.1016/j.tibtech.2016.04.010>

Levy, S. E., & Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, 17(1), 95–115. <https://doi.org/10.1146/annurev-genom-083115-022413>

Lewis, N. E., Nagarajan, H., & Palsson, B. O. (2012). Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4), 291–305. <https://doi.org/10.1038/nrmicro2737>

Li, F., Yuan, L., Lu, H., Li, G., Chen, Y., Engqvist, M. K. M., Kerkhoven, E. J., & Nielsen, J. (2022). Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nature Catalysis*, 5(8), 662–672. <https://doi.org/10.1038/s41929-022-00798-z>

Li, Y., Wang, S., Umarov, R., Xie, B., Fan, M., Li, L., & Gao, X. (2018). DEEPre: Sequence-based enzyme EC number prediction by deep learning. *Bioinformatics*, 34(5), 760–769. <https://doi.org/10.1093/bioinformatics/btx680>

Liebermeister, W., Noor, E., Flamholz, A., Davidi, D., Bernhardt, J., & Milo, R. (2014). Visual account of protein investment in cellular functions. *Proceedings of the National Academy of Sciences*, 111(23), 8488–8493. <https://doi.org/10.1073/pnas.1314810111>

Lopes, H. J. S., Bonturi, N., Kerkhoven, E. J., Miranda, E. A., & Lahtvee, P.-J. (2020). C/N ratio and carbon source-dependent lipid production profiling in *Rhodotorula toruloides*. *Applied Microbiology and Biotechnology*, 104(6), 2639–2649. <https://doi.org/10.1007/s00253-020-10386-5>

Lu, H., Xiao, L., Liao, W., Yan, X., & Nielsen, J. (2024). Cell factory design with advanced metabolic modelling empowered by artificial intelligence. *Metabolic Engineering*, 85, 61–72. <https://doi.org/10.1016/j.ymben.2024.07.003>

Mahadevan, R., & Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, 5(4), 264–276. <https://doi.org/10.1016/j.ymben.2003.09.002>

Malina, C., Yu, R., Björkeroth, J., Kerkhoven, E. J., & Nielsen, J. (2021). Adaptations in metabolism and protein translation give rise to the Crabtree effect in yeast. *Proceedings of the National Academy of Sciences*, 118(51), e2112836118. <https://doi.org/10.1073/pnas.2112836118>

Mans, R., Wijsman, M., Daran-Lapujade, P., & Daran, J.-M. (2018). A protocol for introduction of multiple genetic modifications in *Saccharomyces cerevisiae* using CRISPR/Cas9. *FEMS Yeast Research*, 18(7). <https://doi.org/10.1093/femsyr/foy063>

McAdams, H. H., & Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3), 814–819. <https://doi.org/10.1073/pnas.94.3.814>

Merico, A., Sulo, P., Piškur, J., & Compagno, C. (2007). Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. *The FEBS Journal*, 274(4), 976–989. <https://doi.org/10.1111/j.1742-4658.2007.05645.x>

Metsalu, T., & Vilo, J. (2015). ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research*, 43(W1), W566–W570. <https://doi.org/10.1093/nar/gkv468>

Metzl-Raz, E., Kafri, M., Yaakov, G., Soifer, I., Gurvich, Y., & Barkai, N. (2017a). Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife*, 6, e28034. <https://doi.org/10.7554/elife.28034>

Metzl-Raz, E., Kafri, M., Yaakov, G., Soifer, I., Gurvich, Y., & Barkai, N. (2017b). Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife*, 6, e28034. <https://doi.org/10.7554/elife.28034>

Millán-Oropeza, A., Blein-Nicolas, M., Monnet, V., Zivy, M., & Henry, C. (2022). Comparison of Different Label-Free Techniques for the Semi-Absolute Quantification of Protein Abundance. *Proteomes*, 10(1), 2. <https://doi.org/10.3390/proteomes10010002>

Neidhardt, F. C., Ingraham, J. L., & Schaechter, M. (1990). *Physiology of the bacterial cell: A molecular approach* (2. [print.]). Sinauer.

Nelson, D. L., Cox, M. M., Hoskins, A. A., & Lehninger, A. L. (2021). *Lehninger principles of biochemistry* (Eighth edition). Macmillan International Higher Education.

Ng, C., Jung, M., Lee, J., & Oh, M.-K. (2012). Production of 2,3-butanediol in *Saccharomyces cerevisiae* by in silico aided metabolic engineering. *Microbial Cell Factories*, 11(1), 68. <https://doi.org/10.1186/1475-2859-11-68>

Nielsen, J. (2019). Yeast Systems Biology: Model Organism and Cell Factory. *Biotechnology Journal*, 14(9), 1800421. <https://doi.org/10.1002/biot.201800421>

Nielsen, J., & Keasling, J. D. (2016). Engineering Cellular Metabolism. *Cell*, 164(6), 1185–1197. <https://doi.org/10.1016/j.cell.2016.02.004>

Nonklang, S., Abdel-Banat, B. M. A., Cha-aim, K., Moonjai, N., Hoshida, H., Limtong, S., Yamada, M., & Akada, R. (2008). High-Temperature Ethanol Fermentation and Transformation with Linear DNA in the Thermotolerant Yeast *Kluyveromyces marxianus* DMKU3-1042. *Applied and Environmental Microbiology*, 74(24), 7514–7521. <https://doi.org/10.1128/AEM.01854-08>

Nowrouzian, M., Masloff, S., Pöggeler, S., & Kück, U. (1999). Cell Differentiation during Sexual Development of the Fungus *Sordaria macrospora* Requires ATP Citrate Lyase Activity. *Molecular and Cellular Biology*, 19(1), 450–460. <https://doi.org/10.1128/MCB.19.1.450>

O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., & Palsson, B. Ø. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9(1), 693. <https://doi.org/10.1038/msb.2013.52>

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

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, 28(3), 245–248. <https://doi.org/10.1038/nbt.1614>

Otoupal, P. B., Ito, M., Arkin, A. P., Magnuson, J. K., Gladden, J. M., & Skerker, J. M. (2019). Multiplexed CRISPR-Cas9-Based Genome Editing of *Rhodosporidium toruloides*. *mSphere*, 4(2), e00099-19. <https://doi.org/10.1128/mSphere.00099-19>

Oud, B., Flores, C.-L., Gancedo, C., Zhang, X., Trueheart, J., Daran, J.-M., Pronk, J. T., & Van Maris, A. J. (2012). An internal deletion in MTH1 enables growth on glucose of pyruvate-decarboxylase negative, non-fermentative *Saccharomyces cerevisiae*. *Microbial Cell Factories*, 11(1), 131. <https://doi.org/10.1186/1475-2859-11-131>

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

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

Papapetridis, I., Goudriaan, M., Vázquez Vitali, M., De Keijzer, N. A., Van Den Broek, M., Van Maris, A. J. A., & Pronk, J. T. (2018). Optimizing anaerobic growth rate and fermentation kinetics in *Saccharomyces cerevisiae* strains expressing Calvin-cycle enzymes for improved ethanol yield. *Biotechnology for Biofuels*, 11(1), 17. <https://doi.org/10.1186/s13068-017-1001-z>

Park, Y.-K., Nicaud, J.-M., & Ledesma-Amaro, R. (2018). The Engineering Potential of *Rhodosporidium toruloides* as a Workhorse for Biotechnological Applications. *Trends in Biotechnology*, 36(3), 304–317. <https://doi.org/10.1016/j.tibtech.2017.10.013>

Peterson, C. L., & Herskowitz, I. (1992). Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell*, 68(3), 573–583. [https://doi.org/10.1016/0092-8674\(92\)90192-F](https://doi.org/10.1016/0092-8674(92)90192-F)

Pfeiffer, T., & Morley, A. (2014). An evolutionary perspective on the Crabtree effect. *Frontiers in Molecular Biosciences*, 1. <https://doi.org/10.3389/fmolb.2014.00017>

Pinheiro, M. J., Bonturi, N., Belouah, I., Miranda, E. A., & Lahtvee, P.-J. (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, 1008. <https://doi.org/10.3389/fbioe.2020.01008>

Pramanik, J., & Keasling, J. D. (1997). Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnology and Bioengineering*, 56(4), 398–421. [https://doi.org/10.1002/\(SICI\)1097-0290\(19971120\)56:4%253C398::AID-BIT6%253E3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0290(19971120)56:4%253C398::AID-BIT6%253E3.0.CO;2-J)

Qiao, K., Wasylenko, T. M., Zhou, K., Xu, P., & Stephanopoulos, G. (2017). Lipid production in *Yarrowia lipolytica* is maximized by engineering cytosolic redox metabolism. *Nature Biotechnology*, 35(2), 173–177. <https://doi.org/10.1038/nbt.3763>

Ratledge, C., & Wynn, J. P. (2002a). The Biochemistry and Molecular Biology of Lipid Accumulation in Oleaginous Microorganisms. In *Advances in Applied Microbiology* (Vol. 51, pp. 1–52). Elsevier. [https://doi.org/10.1016/S0065-2164\(02\)51000-5](https://doi.org/10.1016/S0065-2164(02)51000-5)

Ratledge, C., & Wynn, J. P. (2002b). The Biochemistry and Molecular Biology of Lipid Accumulation in Oleaginous Microorganisms. In *Advances in Applied Microbiology* (Vol. 51, pp. 1–52). Elsevier. [https://doi.org/10.1016/S0065-2164\(02\)51000-5](https://doi.org/10.1016/S0065-2164(02)51000-5)

Ryu, J. Y., Kim, H. U., & Lee, S. Y. (2019). Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. *Proceedings of the National Academy of Sciences*, 116(28), 13996–14001. <https://doi.org/10.1073/pnas.1821905116>

Sánchez, B. J., Lahtvee, P., Campbell, K., Kasvandik, S., Yu, R., Domenzain, I., Zelezniak, A., & Nielsen, J. (2021). Benchmarking accuracy and precision of intensity-based absolute quantification of protein abundances in *Saccharomyces cerevisiae*. *PROTEOMICS*, 21(6), 2000093. <https://doi.org/10.1002/pmic.202000093>

Sánchez, B. J., & Nielsen, J. (2015). Genome scale models of yeast: Towards standardized evaluation and consistent omic integration. *Integrative Biology*, 7(8), 846–858. <https://doi.org/10.1039/C5IB00083A>

Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E. J., & Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, 13(8), 935. <https://doi.org/10.15252/msb.20167411>

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

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>

Schomburg, I. (2002). BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, 30(1), 47–49. <https://doi.org/10.1093/nar/30.1.47>

Scott, M., & Hwa, T. (2011). Bacterial growth laws and their applications. *Current Opinion in Biotechnology*, 22(4), 559–565. <https://doi.org/10.1016/j.copbio.2011.04.014>

Sepúlveda Del Rio Hamacek, H., Tingajeva, O., Ostertag, K., Rekēna, A., Illarionov, A., Jõul, P., Monteiro De Oliveira, P., De La Caridad Martín-Hernández, G., Müller, B., Bonturi, N., Passoth, V., Lahtvee, P.-J., & Kumar, R. (2025). Acidification by nitrogen metabolism triggers extracellular biopolymer production in an oleaginous yeast. *Applied and Environmental Microbiology*, e00947-25. <https://doi.org/10.1128/aem.00947-25>

Shapiro, E., Biezuner, T., & Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics*, 14(9), 618–630. <https://doi.org/10.1038/nrg3542>

Sharp, P. M., & Li, W.-H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15(3), 1281–1295. <https://doi.org/10.1093/nar/15.3.1281>

Shen, H., Zhang, X., Gong, Z., Wang, Y., Yu, X., Yang, X., & Zhao, Z. K. (2017). Compositional profiles of *Rhodosporidium toruloides* cells under nutrient limitation. *Applied Microbiology and Biotechnology*, 101(9), 3801–3809. <https://doi.org/10.1007/s00253-017-8157-0>

Sreeharsha, R. V., & Mohan, S. V. (2020). Obscure yet Promising Oleaginous Yeasts for Fuel and Chemical Production. *Trends in Biotechnology*, 38(8), 873–887. <https://doi.org/10.1016/j.tibtech.2020.02.004>

Stephanopoulos, G., Aristidou, A. A., & Nielsen, J. H. (1998). *Metabolic engineering: Principles and methodologies*. Academic Press.

Sukhija, P. S., & Palmquist, D. L. (1988). Rapid method for determination of total fatty acid content and composition of feedstuffs and feces. *Journal of Agricultural and Food Chemistry*, 36(6), 1202–1206. <https://doi.org/10.1021/jf00084a019>

Taguchi, A. K. W., & Young, E. T. (1987). The Cloning and Mapping of *ADR6*, a Gene Required for Sporulation and for Expression of the Alcohol Dehydrogenase II Isozyme From *Saccharomyces cerevisiae*. *Genetics*, 116(4), 531–540. <https://doi.org/10.1093/genetics/116.4.531>

Tammekivi, E., Vahur, S., Vilbaste, M., & Leito, I. (2021). Quantitative GC–MS Analysis of Artificially Aged Paints with Variable Pigment and Linseed Oil Ratios. *Molecules*, 26(8), 2218. <https://doi.org/10.3390/molecules26082218>

Tehlivets, O., Scheuringer, K., & Kohlwein, S. D. (2007). Fatty acid synthesis and elongation in yeast. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1771(3), 255–270. <https://doi.org/10.1016/j.bbalip.2006.07.004>

Teusink, B., Passarge, J., Reijenga, C. A., Esgalhado, E., Van Der Weijden, C. C., Schepper, M., Walsh, M. C., Bakker, B. M., Van Dam, K., Westerhoff, H. V., & Snoep, J. L. (2000). Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry*, 267(17), 5313–5329. <https://doi.org/10.1046/j.1432-1327.2000.01527.x>

Teusink, B., Walsh, M. C., Van Dam, K., & Westerhoff, H. V. (1998). The danger of metabolic pathways with turbo design. *Trends in Biochemical Sciences*, 23(5), 162–169. [https://doi.org/10.1016/S0968-0004\(98\)01205-5](https://doi.org/10.1016/S0968-0004(98)01205-5)

The Gene Ontology Consortium, Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L.-P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., ... Elser, J. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, 49(D1), D325–D334. <https://doi.org/10.1093/nar/gkaa1113>

The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., Da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., ... Zhang, J. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>

Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1), 93–121. <https://doi.org/10.1038/nprot.2009.203>

Thomas, R. (1973). Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42(3), 563–585. [https://doi.org/10.1016/0022-5193\(73\)90247-6](https://doi.org/10.1016/0022-5193(73)90247-6)

Tiukova, I. A., Brandenburg, J., Blomqvist, J., Sampels, S., Mikkelsen, N., Skaugen, M., Arntzen, M. Ø., Nielsen, J., Sandgren, M., & Kerkhoven, E. J. (2019). Proteome analysis of xylose metabolism in *Rhodotorula toruloides* during lipid production. *Biotechnology for Biofuels*, 12(1), 137. <https://doi.org/10.1186/s13068-019-1478-8>

Tiukova, I. A., Prigent, S., Nielsen, J., Sandgren, M., & Kerkhoven, E. J. (2019). Genome-scale model of *Rhodotorula toruloides* metabolism. *Biotechnology and Bioengineering*, 116(12), 3396–3408. <https://doi.org/10.1002/bit.27162>

Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, 11(12), 2301–2319. <https://doi.org/10.1038/nprot.2016.136>

van Dijken, J. P., & Scheffers, W. A. (1986). Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiology Letters*, 32(3–4), 199–224. <https://doi.org/10.1111/j.1574-6968.1986.tb01194.x>

Väremo, L., Nielsen, J., & Nookaew, I. (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*, 41(8), 4378–4391. <https://doi.org/10.1093/nar/gkt111>

Varma, A., & Palsson, B. O. (1994). Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology*, 12(10), 994–998. <https://doi.org/10.1038/nbt1094-994>

Verduyn, C. (1991). Physiology of yeasts in relation to biomass yields. *Antonie van Leeuwenhoek*, 60(3–4), 325–353. <https://doi.org/10.1007/BF00430373>

Verduyn, C., Postma, E., Scheffers, W. A., & Van Dijken, J. P. (1990). Physiology of *Saccharomyces Cerevisiae* in Anaerobic Glucose-Limited Chemostat Culturesx. *Journal of General Microbiology*, 136(3), 395–403. <https://doi.org/10.1099/00221287-136-3-395>

Verduyn, C., Postma, E., Scheffers, W. A., & Van Dijken, J. P. (1992). Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, 8(7), 501–517. <https://doi.org/10.1002/yea.320080703>

Verho, R., Putkonen, M., Londesborough, J., Penttilä, M., & Richard, P. (2004). A Novel NADH-linked L-Xylulose Reductase in the L-Arabinose Catabolic Pathway of Yeast. *Journal of Biological Chemistry*, 279(15), 14746–14751. <https://doi.org/10.1074/jbc.M312533200>

Villadsen, J., Nielsen, J., & Lidén, G. (2011). *Bioreaction Engineering Principles*. Springer US. <https://doi.org/10.1007/978-1-4419-9688-6>

Walker, G. (2000). *Yeast physiology and biotechnology* (repr). Wiley.

Walker, G., & Stewart, G. (2016). *Saccharomyces cerevisiae* in the Production of Fermented Beverages. *Beverages*, 2(4), 30. <https://doi.org/10.3390/beverages2040030>

Wang, H., Marcišauskas, S., Sánchez, B. J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., & Kerkhoven, E. J. (2018). RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLOS Computational Biology*, 14(10), e1006541. <https://doi.org/10.1371/journal.pcbi.1006541>

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences*, 24(11), 437–440. [https://doi.org/10.1016/S0968-0004\(99\)01460-7](https://doi.org/10.1016/S0968-0004(99)01460-7)

Wasylewko, T. M., Ahn, W. S., & Stephanopoulos, G. (2015). The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in *Yarrowia lipolytica*. *Metabolic Engineering*, 30, 27–39. <https://doi.org/10.1016/j.ymben.2015.02.007>

Wynn, J. P., Kendrick, A., & Ratledge, C. (1997). Sesamol as an inhibitor of growth and lipid metabolism in *Mucor circinelloides* via its action on malic enzyme. *Lipids*, 32(6), 605–610. <https://doi.org/10.1007/s11745-997-0077-1>

Wynn, J. P., & Ratledge, C. (1997). Malic Enzyme is a Major Source of NADPH for Lipid Accumulation by *Aspergillus nidulans*. *Microbiology*, 143(1), 253–257. <https://doi.org/10.1099/00221287-143-1-253>

Wynn, J. P., Ratledge, C., Hamid, A. A., & Li, Y. (2001). Biochemical events leading to the diversion of carbon into storage lipids in the oleaginous fungi *Mucor circinelloides* and *Mortierella alpina*. *Microbiology*, 147(10), 2857–2864. <https://doi.org/10.1099/00221287-147-10-2857>

Xia, J., Sánchez, B. J., Chen, Y., Campbell, K., Kasvandik, S., & Nielsen, J. (2022). Proteome allocations change linearly with the specific growth rate of *Saccharomyces cerevisiae* under glucose limitation. *Nature Communications*, 13(1), 2819. <https://doi.org/10.1038/s41467-022-30513-2>

Yang, F., Zhang, S., Zhou, Y. J., Zhu, Z., Lin, X., & Zhao, Z. K. (2012). Characterization of the mitochondrial NAD+-dependent isocitrate dehydrogenase of the oleaginous yeast *Rhodosporidium toruloides*. *Applied Microbiology and Biotechnology*, 94(4), 1095–1105. <https://doi.org/10.1007/s00253-011-3820-3>

Yoon, S. H., Park, J. S., & Rhee, J. S. (1984). Production of NADPH for Lipogenesis in Oleaginous Yeast *Rhodotorula glutinis*. 12(3), 247–251.

Zhang, H., Zhang, L., Chen, H., Chen, Y. Q., Ratledge, C., Song, Y., & Chen, W. (2013). Regulatory properties of malic enzyme in the oleaginous yeast, *Yarrowia lipolytica*, and its non-involvement in lipid accumulation. *Biotechnology Letters*, 35(12), 2091–2098. <https://doi.org/10.1007/s10529-013-1302-7>

Zhang, S., Ito, M., Skerker, J. M., Arkin, A. P., & Rao, C. V. (2016). Metabolic engineering of the oleaginous yeast *Rhodosporidium toruloides* IFO0880 for lipid overproduction during high-density fermentation. *Applied Microbiology and Biotechnology*, 100(21), 9393–9405. <https://doi.org/10.1007/s00253-016-7815-y>

Zhang, Y., Adams, I. P., & Ratledge, C. (2007). Malic enzyme: The controlling activity for lipid production? Overexpression of malic enzyme in *Mucor circinelloides* leads to a 2.5-fold increase in lipid accumulation. *Microbiology*, 153(7), 2013–2025. <https://doi.org/10.1099/mic.0.2006/002683-0>

Zhu, J., Gu, Y., Yan, Y., Ma, J., Sun, X., & Xu, P. (2023). Knocking out central metabolism genes to identify new targets and alternating substrates to improve lipid synthesis in *Y. lipolytica*. *Frontiers in Bioengineering and Biotechnology*, 11, 1098116. <https://doi.org/10.3389/fbioe.2023.1098116>

Zhu, Z., Zhang, S., Liu, H., Shen, H., Lin, X., Yang, F., Zhou, Y. J., Jin, G., Ye, M., Zou, H., & Zhao, Z. K. (2012). A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. *Nature Communications*, 3(1), 1112. <https://doi.org/10.1038/ncomms2112>

Disclosure of Copyrighted Materials

All reproduced figures from open-access publications are used in accordance with their respective Creative Commons licenses. Figures originally published under the CC BY 4.0 license were reused and, where applicable, adapted without requiring permission for their reuse or adaptation. For materials published under the CC BY-NC-ND license, no modifications were made to the original content; instead, data were reformatted and presented in new figure layouts created by the author. Specific sources and licensing details are cited in the corresponding figure legends.

Acknowledgements

I would like to express my sincere gratitude to my advisor for *all* the opportunities and guidance provided throughout my doctoral studies. You taught me the importance of learning from actions rather than words, a lesson that will stay with me beyond science. I also wish to express my gratitude to our collaborators who developed the conceptual framework of **Study 1** and invited us to be part of the study.

I am also thankful to my former advisors for the moments we shared and learned science together. These experiences have contributed greatly to my growth as a researcher.

To my mentors, I am deeply grateful for your time, encouragement, and belief in me.

I extend my thanks to my lab mates and colleagues for sharing your perspectives and for the time we spent working side by side.

Finally, I wish to thank my parents. None of this would have been possible without your support. I am proud to be your daughter.

I would like to acknowledge Estonian Research Council, Estonian Doctoral School, Yeast4Bio COST Action for the funding support. I also acknowledge the use of the AI language model ChatGPT (OpenAI, 2025) to improve phrasing and clarity of writing of this thesis. All scientific interpretations, data analyses, and conclusions remain my own.

Abstract

Computational design of yeast-based cell factories

The design of efficient yeast-based cell factories requires predictive computational models that accurately capture quantitative links between genotype and phenotype. Among the modeling approaches available, genome-scale metabolic models (GEMs) enable the prediction of cellular growth and metabolic fluxes under defined environmental and genetic conditions. However, conventional GEMs rely solely on stoichiometric and steady-state assumptions and therefore fail to capture enzyme capacity limitations or kinetic and regulatory control. This limitation restricts their predictive power, particularly in non-model yeasts that display distinct metabolic regulation and for which experimental data remain scarce. To address these shortcomings, enzyme-constrained genome-scale metabolic models (ecGEMs) extend the GEM framework by coupling metabolic fluxes to absolute proteomics data and reaction-specific catalytic constants (k_{cat}).

The overarching aim of this thesis was to develop, apply, and critically evaluate enzyme-constrained models for predicting metabolism in both the model yeast *Saccharomyces cerevisiae* and the lipid-producing yeast *Rhodotorula toruloides*. The work was motivated by the exploration of thermodynamically feasible, redox neutral and ATP yielding pathways for anaerobic product formation, as well as by the goal of extending advanced metabolic modeling toward nonconventional, industrially relevant yeast species.

Study 1 integrated absolute quantitative proteomics with physiological data from controlled bioreactor cultivations to parametrize ecGEMs and assess the predictive power for yields, metabolic reaction rates (fluxes), and proteome allocation in *Saccharomyces cerevisiae* strains engineered for redox-neutral, ATP-yielding co-production of (R,R)-2,3-butanediol and glycerol. The model was further constrained by genetic perturbations, and the predicted fluxes, validated experimentally with the engineered strain, reproduced several observed physiological trends, including growth reduction and proteome reallocation.

Study 2 focused on *R. toruloides*, an oleaginous yeast naturally capable of synthesizing high amounts of lipids and carotenoids. Absolute quantitative proteomics and physiological data from bioreactor cultivations were integrated to construct new ecGEMs and to simulate carbon flow through intracellular metabolic pathways during exponential growth and lipid-accumulation phases on three carbon substrates – glucose, xylose and acetate. Model-based hypotheses were evaluated through targeted gene deletions and subsequent growth experiments assessing the effects on growth and lipid accumulation. Although the ecGEM accurately reproduced growth and global proteome limitations, it systematically underestimated the activity of key lipid-precursor biosynthetic pathways. These results emphasized the need for accurate catalytic constants, compartmentalization, and maintenance parameters for reliable model predictions and provided new insight into the roles of specific genes in lipid-precursor production on fermentative and non-fermentative substrates.

The thesis demonstrates the integration of genetic, biochemical and computational approaches to develop and validate predictive models of yeast metabolism, highlighting both strengths and current critical gaps in modeling frameworks, and provides perspectives on how future improvements can advance *in silico* biomanufacturing design.

Lühikokkuvõte

Pärmpõhiste rakuvalbrikute *in silico* disain

Töhusate pärmpõhiste rakuvalbrikute disainimine vajab ennustusvõimelisi arvutuslikke mudeliteid, mis suudavad täpselt kirjeldada kvantitatiivseid seoseid genotübi ja fenotübi vahel. Üheks selliseks lähenemiseks on ülegenoomsed metabolismimudelid (GEM-id), mis võimaldavad prognoosida rakkude kasvu ja metaboolsete reaktsioonide kiirusti määratletud keskkonna- ja geneetilistes tingimustes. Traditsioonilised GEM-id tuginevad siiski üksnes stöhhiomeetrilistele ja püsiseisundi eeldustele ega arvesta seetõttu ensüümide mahulisi piiranguid ega geneetilist ja regulatoorset kontrolli. See piirang vähendab nende ennustusvõimet, eriti mitte-mudel tüüpi pärmpides, millel on spetsiifiline metaboolne regulatsioon ning mille eksperimentaalsed andmed on puudulikud. Nende puuduste kõrvaldamiseks laiendavad ensüümipiiranguga ülegenoomsed metabolismimudelid (ecGEM-id) GEM-raamistikku, sidudes metaboolsed vood absoluutsete proteoomika andmete ja reaktsioonispetsiifiliste katalüütiliste konstantidega.

Käesoleva väitekirja üldeesmärk oli välja töötada, rakendada ja kriitiliselt hinnata ensüümipiiranguga mudelite pärmpide metabolismi ennustamiseks nii mudelpärmas *Saccharomyces cerevisiae* kui ka lipiide tootvas pärmis *Rhodotorula toruloides*. Töö oli motiveeritud termodünaamiliselt võimalike, redoksneutraalse ja ATP-d tootvate sünteesiradade uurimisest anaeroobsete produktide moodustumisel ning soovist laiendada täiustatud metabolismimudelite kasutust mitte-konventionsaalsetele, kuid tööstuslikult olulistele pärmpõhikidele.

Uuring 1 seob absoluutse kvantitatiivse proteoomika ja füsioloogilised andmed kontrollitud bioreaktori kasvatusest, et ecGEM-e parameetrite kaudu väljendada ja hinnata nende ennustusvõimet saagikuse, metaboolsete reaktsioonide kiiruste ning proteoomi jaotuse osas *S. cerevisiae* tüvedes, mis olid geneetiliselt muundatud redoksneutraalse, ATP-d tootva (R,R)-2,3-butanidiooli ja glütserooli koosproduktiooni jaoks. Järgnevalt täiendati mudelit geneetiliste muudatustega ning prognoositud vood, mida kontrolliti eksperimentaalselt muundatud tüves, taasesitasid mitmeid tähendatud füsioloogilisi trende, sealhulgas kasvu vähinemist ja proteoomi ümberjaotumist.

Uuring 2 keskendus *R. toruloides*'ele, looduslikult lipiidsele pärnile, mis on võimeline sünteesima suuri koguseid lipiide ja karotenoide. Absoluutne kvantitatiivne proteoomika ja füsioloogilised andmed bioreaktori kultiveerimistest seoti uute ecGEM-ide koostamiseks ning süsinikuvoogude modelleerimiseks rakusisestes metabolismiradades eksponentsialse kasvu ja lipiidide akumulatsiooni faasis kolmel erineval süsinikusubstraadil – glükoosil, ksüloosil ja atsetaadil. Mudelil põhinevaid hüponeese kontrolliti määratud geenide eemaldamiste ja sellele järgnenud kasvukatsetega, hinnates möjusid kasvule ja lipiidide akumulatsioonile. Kuigi ecGEM ennustas täpselt kasvu ja proteoomi üldisi piiranguid, alahindas see süstemaatiliselt peamiste lipiid-prekursorite biosünteesiliste radade aktiivsust. Need tulemused rõhutasid täpsete katalüütiliste konstantide, raku kompartmentide ja alalhoiuer energia tähtsust usaldusväärsete mudelprognooside saavutamisel ning andsid uusi teadmisi konkreetsete geenide rollist lipiid-prekursorite tootmises fermenteeruvatel ja mitte-fermenteeruvatel substraatidel.

Väitekirjas näidati, kuidas siduda geneetilised ja biokeemilised andmed mudelitega, et pärnide metabolismi ennustavate mudelite arendamiseks ja valideerimiseks. Töö toob esile nii ensüümipiiranguga modelleerimise tugevused kui ka praegused kitsaskohad ning pakub nägemuse, kuidas edasised täiustused võivad edendada *in silico* biotööstuslike tootmislahenduste kavandamist.

Appendix 1

Supplementary materials

Supplementary Tables

Supplementary tables to Study 1

Table S1. Gene set analysis (Väremo et al., 2013) of measured proteome data. Results provide functional groups with the number of proteins included in each group and the adjusted p-value without considering directionality (Non-directional), upregulated, and downregulated proteins in the dataset. Reproduced from Sjöberg et al., Metabolic Engineering 82,49-59, 2024, under a CC BY licence.

Protein functional groups	Number of proteins	GSY014 vs GSY013			% of total protein	
		Upregulated	Downregulated	Non-directional	GSY013	GSY014
Central carbon metabolism, glycolysis	47	0,0012999	1	0,0012999	28,673181	43,4930885
Biosynthesis, amino acid metabolism	116	0,0012999	1	0,0012999	6,85647662	8,7849401
Translation, ribosome	110	1	0,0038996	0,0012999	25,4856646	18,5188884
Biosynthesis, cofactor biosynthesis	76	0,0012999	1	0,0019498	0,92952053	1,06945689
Central carbon metabolism, carbohydrate metabolism	31	1	0,0090991	0,0062394	0,61589211	0,4048838
Biosynthesis, lipid and steroid metabolism	62	0,76222	1	0,0090991	0,94723025	1,77506151
Biosynthesis, purine metabolism	28	0,52801	1	0,031754	1,49754419	1,74714157
Central carbon metabolism, other central metabolism enzymes	24	0,83247	1	0,052157	0,84297743	1,00011134
Biosynthesis, pyrimidine metabolism	18	0,85242	1	0,11699	0,33569691	0,33994558
Translation, tRNA loading	36	0,0048745	1	0,16027	0,823187	0,84344909
Central carbon metabolism, TCA cycle and anaplerotic enzymes	19	0,016248	1	0,31693	0,58057421	0,66133807
Folding sorting degradation, peptidases	32	0,56219	1	0,60894	0,53092279	0,61846825
Translation, translation factors	42	1	0,20101	0,60894	4,43020195	3,99360999
Central carbon metabolism, pentose phosphate metabolism	16	0,016248	1	0,9998	0,73470907	0,65379915
Other enzymes, other enzymes	116	0,18025	1	0,9998	1,15630208	1,18128839
Energy metabolism, oxidative phosphorylation	30	0,048745	1	0,9998	0,68681761	0,64869067
Translation, mRNA surveillance pathway	26	0,78887	1	0,9998	0,08479007	0,08025251
Folding sorting degradation, sulfur relay system	5	1	0,42236	0,9998	0,1406847	0,0941634
DNA maintenance, DNA repair and recombination proteins	8	1	0,74093	0,9998	0,0284869	0,02259921

Protein functional groups	Number of proteins	GSY014 vs GSY013			% of total protein	
		Upregulated	Downregulated	Non-directional	GSY013	GSY014
Biosynthesis, glycan metabolism	34	1	0,026517	0,9998	0,10913066	0,10280548
Folding sorting degradation, protein processing in endoplasmatic reticulum	42	1	0,10362	0,9998	0,30244859	0,2703966
Translation, RNA transport	35	0,78887	1	0,9998	0,67169539	0,68602054
Membrane transport, transport	12	0,78887	1	0,9998	0,50463491	0,56536671
Folding sorting degradation, chaperones and folding catalysts	61	0,91505	1	0,9998	3,78810893	3,4605689
DNA maintenance, chromosome-related	113	1	0,026517	0,9998	0,73873438	0,49726759
Transcription, basal transcription factors	28	1	0,76016	0,9998	0,06836305	0,06688099
Folding sorting degradation, ubiquitin mediated proteolysis	16	0,83247	1	0,9998	0,10112648	0,10017797
DNA maintenance, DNA replication complex	24	0,78887	1	0,9998	0,07823503	0,07682752
DNA maintenance, DNA replication control	11	0,83247	1	0,9998	0,01923367	0,01451861
Translation, ribosome biogenesis	11	1	0,13356	0,9998	0,08240959	0,06981363
Transcription, transcription factors	27	1	0,0038996	0,9998	0,07595343	0,06522586
Folding sorting degradation, proteasome	32	0,048745	1	0,9998	0,61754182	0,64410242
Transcription, RNA polymerase	29	1	0,80527	0,9998	0,23821556	0,23391551
Folding sorting degradation, ubiquitin labeling	22	0,56219	1	0,9998	0,06581475	0,07028825
Translation, ribosome biogenesis in eukaryotes	55	1	0,027297	0,9998	0,45656996	0,40651673
Folding sorting degradation, protein export	17	1	0,2278	0,9998	0,07331038	0,06628602
Folding sorting degradation, RNA degradation	36	0,83247	1	0,9998	0,2183356	0,19722009
Transcription, spliceosome	56	1	0,13779	0,9998	0,24065933	0,21380151
Folding sorting degradation, SNARE interactions in vesicular transport	15	1	0,16417	0,9998	0,03068322	0,02902006

Table S2. Enzymes under flexibilization of individual protein constraint by the ecGEM, their original and modified values. Constraints were adjusted by algorithmically increasing measured absolute protein concentration to achieve feasible flux states in proteome-constrained ecGEM simulations. Reproduced from Sjöberg et al., Metabolic Engineering 82,49-59, 2024, under a CC BY licence.

protein_IDs	previous values	modified values	flex mass	Fold change
ecGSY014				
YJR148W	7.34168641470523E-06	0.0000126212361363917	0.219756980729922	1.7191194806566
YDR226W	0.0000139550415479416	0.000168346889374346	3.7447125122903	12.0635176037278
YKL060C	0.000206226010683615	0.000292033602901766	3.39971739750939	1.41608520638938
YDL093W	4.29282746859551E-08	8.37144957433201E-07	0.0673295604365612	19.5010156722439
YBR084W	6.22617721052581E-07	8.67798100279856E-07	0.0260421821357636	1.39378959341661
YNL241C	2.58197414940101E-06	3.18689327441421E-06	0.0347954924979718	1.23428550791399
YDL171C	0.0000300929156952497	0.000037854908282196	1.84812608823608	1.25793421500103
YOL059W	0.0000219021897875799	0.000204832337515445	9.04061827038093	9.35213964914136
YFL018C	4.54830791768671E-06	6.71221577596403E-06	0.116871235246372	1.47576107366493
YMR189W	1.52847123109353E-06	1.64706294915667E-06	0.0135728612675926	1.07758845286103
YHR025W	1.88917786160494E-06	3.87905309650122E-06	0.0770323485745901	2.05330221962573
YMR267W	2.15051372298204E-07	5.98668176529226E-07	0.0136460668302905	2.78383797383573
YOL126C	2.61220332601348E-08	6.86660983272699E-08	0.00173284104221694	2.62866590986477
YMR205C	0.0000145403890848472	0.000165499675615655	15.7928669199861	11.3820665079812
YKL127W	1.07106731177744E-06	0.0000031734716608686	0.13268494599571	2.96290590327345
YGL062W	0.0000052095268499725	8.13483406971729E-06	0.380577408507306	1.56153030860379
YGR193C	1.02697872233855E-06	1.35473535384973E-06	0.0148674505011362	1.31914646757709
YER026C	8.56279415104335E-08	1.32074551074816E-07	0.00143075529500412	1.5424235214007
YPL134C	1.61860509441388E-08	2.18248443334158E-06	0.0740999908715941	134.837363410863
YMR241W	1.34354661858203E-06	8.75370421444899E-06	0.253309642374452	6.51537065657431
YLR348C	1.19982250688092E-07	1.22199353909357E-06	0.0363567188984942	10.1847859336319
YKL120W	1.06277750770121E-06	4.28034467750603E-06	0.113105948220296	4.0275077701762
YLR092W	2.99026459251971E-08	9.82215988507091E-07	0.0948977103248827	32.8471263367179
YPR058W	2.71071659041941E-07	4.68727934167305E-06	0.147392432918425	17.2916613940368
YBR166C	5.36288471365389E-07	7.67601444340613E-07	0.0117790857551849	1.43132191968681
YBR196C	0.000033415299772487	0.000161232669566215	7.83499003517426	4.82511516173703
YDL141W	1.38582492389754E-07	1.31961345476461E-06	0.0901859217797978	9.5222234209303
YDR127W	2.93766613000817E-06	3.74161803359305E-06	0.140493047204765	1.27367027701771
YDR234W	2.12981217246513E-06	3.65222820028056E-06	0.114409397024567	1.71481234237352
YDR408C	9.40434807258418E-07	1.71705829642162E-06	0.0182816703374924	1.82581321232381
YER003C	2.42485462627788E-06	3.18602976437456E-06	0.0366797054601386	1.31390547286749
YFL045C	4.61330423427245E-06	0.0000096809357682781	0.147277838751778	2.09848197228312

protein_IDs	previous values	modified values	flex mass	Fold change
YFR025C	3.77922203229925E-07	1.91881040330703E-06	0.0594495931846907	5.07726295758188
YGL148W	5.51201738995808E-06	0.00000876853510456	0.132989442472673	1.59080323667605
YGR185C	2.05557549493691E-06	2.84420186039857E-06	0.0347148752043302	1.38365234816436
YGR260W	2.44909772125445E-08	5.63256300935921E-07	0.0323982486681124	22.9985229273504
YGR264C	1.96088696356967E-06	2.49745273903439E-06	0.0459713566758069	1.27363422034687
YIL020C	2.44021595010777E-07	2.49266082016012E-06	0.0664612981255358	10.2149189707986
YIL078W	3.45783586439813E-06	4.40299273563897E-06	0.0798841294694281	1.27333769106052
YKL024C	2.00877843670793E-06	3.41266756710619E-06	0.0321953332718581	1.69887704126246
YKL067W	2.34893066259546E-06	5.45897202205949E-06	0.0533884316959986	2.32402433540868
YNL316C	2.63438769577246E-07	0.0000019489768964799	0.0644294252147536	7.39821591031391
YPR021C	2.07088128455421E-08	3.69413229671712E-06	0.383149052904022	178.384551749539
YPR167C	9.22944062003563E-07	0.0000027830796212845	0.0565107322773989	3.01543694342957
alsS	0.0000233136832958054	0.0000332553612977999	0.617378203923858	1.42643103090378
budA	0.0000098616428532387	0.000148926183203201	4.0050587620789	15.1015591843596
ecGSY013				
YIL020C	2.0707823658965E-07	5.98178772978745E-06	0.170678641964138	28.8866074402646
YEL038W	0.0000008048548373412	2.70581093510809E-06	0.0478796473683084	3.36186205210198
YMR009W	1.16501402921321E-07	2.70581093510809E-06	0.0540624526534808	23.2255652486473
YOL066C	1.20412014020544E-07	1.56625388876581E-07	0.00242756539414323	1.30074552901223
YDR035W	2.24006592433249E-06	3.78476718515613E-06	0.0634401393254218	1.68957848251004
YHR208W	0.0000084052125018761	0.0000174378786847	0.393783869553284	2.07465054343454
YJR148W	7.88024792371187E-07	0.00003029096672734	1.22803606066011	38.4391037193052
YDR234W	2.24870718594819E-06	8.76447114590527E-06	0.489658944856739	3.89755998498739
YDR127W	2.26896505942829E-06	8.97898529229527E-06	1.17259650125522	3.95730434674816
YLR017W	1.77082396295614E-06	4.23907237116664E-06	0.0934384560259301	2.39384177074841
YPR118W	5.6314822574007E-07	2.56025647086442E-06	0.0899094736870967	4.54632786510127
YGR248W	1.96770511848961E-07	0.0000182597126530684	0.513843198379862	92.7969972812002
YLR355C	0.0000216812254346756	0.0000274693317437741	0.256806237673576	1.2669639834952
YDL141W	1.53643649761605E-07	3.16675558420593E-06	0.230087343931395	20.6110411274368
YBL015W	4.47104149183052E-07	3.53972271132916E-06	0.181571872671028	7.91699812626865
YER069W	2.61195211998824E-06	0.0000105065708895909	0.748948509196984	4.02249750643906
YOL140W	2.30889510107621E-06	4.98957191315747E-06	0.12513619174294	2.16102148202045
YDR441C	4.37016690637517E-07	1.26418375799461E-06	0.0165315128580786	2.8927585263401
YOL052C	2.23231350623218E-07	0.0000020663008633722	0.0852082920846427	9.25632021489587
YDL166C	7.67561525177403E-07	0.000129536970463576	2.92601440236634	168.764283011237
YDR354W	7.80393173535528E-07	0.0000019824638829337	0.0497344975720542	2.54033985709057

protein_IDs	previous values	modified values	flex mass	Fold change
YGL148W	4.59344419904105E-06	0.0000210423798025189	0.671740480749337	4.58095905615046
YKL060C	0.000166494456565975	0.000258458741108414	3.64364702499974	1.55235643539878
YLR420W	1.67789634570431E-06	2.79014627766009E-06	0.0448379646694435	1.66288357728612
YJR016C	0.000013358200510892	0.000021812245399003	0.531425404148656	1.6328730341499
YPR183W	1.62832445226523E-06	2.18319255776276E-06	0.0168468665783485	1.34076016283219
YDL093W	6.14640877647385E-08	2.00914789783968E-06	0.165114505462341	32.68815939366
YBR084W	8.00966090229429E-07	9.23964328829121E-07	0.0130644325703651	1.15356235438689
YNL241C	2.23925876425137E-06	0.0000192960109867273	0.981119738913814	8.61714210736978
YBR196C	0.0000271400984376489	0.000140533326241554	6.95081436375666	5.17806987931206
YGL245W	0.0000101282043288754	0.0000175666676915239	0.601339288162993	1.73443061781855
YOL059W	5.19035224622731E-06	0.0000142030682311429	0.445418788597911	2.73643628743437
YDR408C	6.83011607822419E-07	4.12052781741949E-06	0.0809189253229425	6.03288109635
YFL018C	4.09784741765133E-06	0.0000124560036680142	0.451418502698974	3.03964555009062
YMR189W	1.45170183601458E-06	2.58447794063851E-06	0.129646598990324	1.78030906658752
YFR015C	7.67624547943735E-07	9.75926187524028E-06	0.723910516990889	12.713587523201
YBR121C	5.24054570962615E-06	0.0000109947699542727	0.433511400889731	2.09802004666744
YFR025C	3.25041369955677E-07	4.60468445344007E-06	0.165114536068283	14.1664565777211
YDL182W	0.0000143545481066437	0.0000531492758699868	1.82716999403632	3.70260878121184
YHR025W	1.41234840466105E-06	9.30972743160293E-06	0.305724521497828	6.59166491842866
YNL277W	8.3127351740627E-07	9.30692166821144E-07	0.00533470133220675	1.119597999134
YNL037C	3.56102272441385E-06	0.0000131511879411508	0.377119533211921	3.69309295641058
YOR136W	3.20336686269192E-06	0.0000131511879411508	0.39531676027151	4.10542672908195
YLR089C	7.41153578714527E-07	6.13399176049652E-06	0.358197111659942	8.27627624916208
YER023W	2.94336588028675E-06	5.42235641133499E-06	0.0746955792367534	1.84222982526614
YKL085W	1.73771857515319E-06	0.0000038919557242198	0.0767976711269934	2.23969276721157
YER003C	3.16805095543162E-06	7.64570678735549E-06	0.215770443419264	2.41337872872497
YGR264C	1.93102229968995E-06	5.99328718502518E-06	0.348042749826164	3.10368615939209
YKL067W	1.56308873879918E-06	0.0000133163833123186	0.201762578697484	8.51927531801471
YKL184W	2.8292572133238E-07	2.67418669331654E-06	0.125026290804071	9.45190377432993
YPR167C	9.82223656438124E-07	6.67872315197366E-06	0.17305908502442	6.79959509038193
YMR205C	9.08386092379404E-06	0.000146472314789958	14.3731307832339	16.1244558914692
YFL045C	7.02273880199969E-06	0.0000232319224192831	0.471078750509953	3.30810002682542
YNL316C	2.41017136003625E-07	0.0000046770767970966	0.169567670775733	19.4055778549549
YBR166C	4.22893381168624E-07	1.84205924207083E-06	0.072267785768282	4.35584789002959
YGL062W	4.30507251069082E-06	0.0000287024253754188	3.17405340028065	6.66711775565727
YGR193C	7.04940711725255E-07	3.08756134986342E-06	0.108078650421745	4.37988797995082

protein_IDs	previous values	modified values	flex mass	Fold change
YOR095C	1.15115947187808E-06	2.94757501561345E-06	0.050762697259299	2.56052709257091
YLR146C	1.22672087100474E-07	2.13793707576971E-07	0.00310636793461093	1.74280647399326
YIL078W	3.00598648838671E-06	0.000010566125847277	0.638978744935365	3.51502772487436
YGR185C	2.37817115319451E-06	6.82540185651007E-06	0.195764516166143	2.8700212965504
YKL024C	0.0000012947721571925	8.18958312083876E-06	0.158118424036861	6.3251152531705
YGR094W	2.00384927360481E-06	2.95486716964651E-06	0.119607808952952	1.47459552400909
YKR067W	2.22543137443432E-07	2.10622462934014E-06	0.157558786565913	9.46434319896977
YJR073C	3.00543928315243E-07	3.74102409118646E-06	0.0796472189848735	12.4475117902314
YMR241W	0.0000013769362330796	0.0000179989783945226	0.568209717671406	13.0717588528169
YPR021C	1.41172822036582E-08	0.0000228110029068117	2.37778333333464	1615.82113169776
YLR348C	1.44155342130625E-07	3.22997631528754E-07	0.00590022889168312	2.24062200369982
YGR260W	3.00305509144484E-08	1.35167994073388E-06	0.079476394819749	45.0101613048849
YKL120W	9.79540093332928E-07	0.0000102728272260145	0.326683484058915	10.4873984188445
YGL080W	1.28834409133766E-08	0.0000716013801329679	1.07348024617186	5557.62863464727
YPR058W	2.68406914610939E-07	7.89455089521339E-06	0.25452514824156	29.4126211564136

Supplementary tables to Study 2

Table S3. Gene-protein names of *R. toruloides* analyzed in **Study 2**. Asterisks (*) denote isoenzymes. Subsystem as defined for **Figure S6** and **Figure S10** in the Supplementary figures. Abbreviations: PPP, pentose phosphate pathway; TCA, tricarboxylic acid. Modified from Rekēna et al., PLOS Comput. Biol. 19(4):e1011009, 2023, under a CC BY licence.

Gene name	Subsystem	Reaction ID	Locus tag	Protein accession (UniProt)	Protein name
ACC	FA metabolism	r_0109No1	RHTO_02004	M7XLR4	acetyl-coa carboxylase
ACL	FA metabolism	y200003No1	RHTO_03915	M7WHC9	atp citrate synthase
ACO1	TCA cycle	r_0280No1	RHTO_00539	M7X6X3	aconitate hydratase, mitochondrial (aconitase) (ec 4.2.1.-)
ACO2	TCA cycle	r_0280No2	RHTO_08030	M7WQ73	aconitate hydratase, mitochondrial (aconitase) (ec 4.2.1.-)
ACS	Pyruvate metabolism	r_0112No1	RHTO_08027	M7XFRO	acetyl-coenzyme a synthetase (ec 6.2.1.1)
AGC	other	r_1118No1	RHTO_04513	M7WMP3	mitochondrial carrier protein

Gene name	Subsystem	Reaction ID	Locus tag	Protein accession (UniProt)	Protein name
ALDH1	Pyruvate metabolism	r_2116No1	RHTO_02062	M7WSZ7	Succinate-semialdehyde dehydrogenase (ec 1.2.1.16)
ALDH2	Pyruvate metabolism	r_2116No2	RHTO_04310	M7WN91	aldehyde dehydrogenase
ALDH3	Pyruvate metabolism	r_2116No3	RHTO_04425	M7X6W7	aldehyde dehydrogenase, mitochondrial
ALDH4	Pyruvate metabolism	r_2116No4	RHTO_05680	M7WLP6	aldehyde dehydrogenase
ALDH5	Pyruvate metabolism	r_2116No5	RHTO_05838	M7WKL7	aldehyde dehydrogenase (nad)
CIT1	TCA cycle	r_0300No1	RHTO_06406	M7XE29	citrate synthase
CRC	other	r_1120No1	RHTO_01354	M7WMH0	carnitine acyl carnitine carrier, mitochondrial
CTP1	other	r_1112No1	RHTO_05609	M7WEL7	mitochondrial carrier protein, tricarboxylate carrier
CTP2	other	r_1112No2	RHTO_08140	M7WQJ3	mitochondrial carrier protein
DAD-4	xylose metabolism	t_0883No1	RHTO_07844	M7XGH5	D-arabinitol dehydrogenase
DAD02/LXR	xylose metabolism	t_0884_REVNo 1	RHTO_00373	M7X791	I-xylulose reductase
ENO	glycolysis	G,X: r_0366No1, A: r_0366_REVNo 1	RHTO_00323	M7X749	enolase
FAS1*	FA metabolism	r_2141No1	RHTO_02139	M7XM89	fatty acid synthase subunit alpha, fungi type
FAS2*	FA metabolism	r_2141No1	RHTO_02032	M7WSW5	fatty acid synthase subunit beta, fungi type
FBA	PPP	G: r_0450No1; RHTO_03043 A: r_0450_REVNo 1; X: r_0990No1	M7X5F4		Fructose-bisphosphate aldolase (fbp aldolase) (ec 4.1.2.13)

Gene name	Subsystem	Reaction ID	Locus tag	Protein accession (UniProt)	Protein name
FBP	PPP	r_0449No1	RHTO_03046	M7XYF6	fructose-1,6-bisphosphatase i
FUM	TCA cycle	r_0451No1	RHTO_05746	M7XEU6	fumarate hydratase, class ii
GAPDH1	glycolysis	r_0486No1	RHTO_01292	M7WNF8	glyceraldehyde 4-phosphate dehydrogenase
GAPDH2	glycolysis	G,X: r_0486No2, A: r_0486_REVNo 2	RHTO_03746	M7WI96	Glyceraldehyde-3-phosphate dehydrogenase (ec 1.2.1.12)
GND1	PPP	r_0091No1	RHTO_02788	M7X3Z4	6-phosphogluconate dehydrogenase, decarboxylating (ec 1.1.1.44)
GND2	PPP		RHTO_00011	M7XZ45	Phosphogluconate dehydrogenase (decarboxylating)
GPI	PPP	G,X:r_0467No, A:r_0467_REV No1	RHTO_04058	M7WNZ9	Glucose-6-phosphate isomerase (ec 5.3.1.9)
ICL1	Glyoxylate shunt	r_0662No1	RHTO_03507	M7WQF1	isocitrate lyase
ICL2	Glyoxylate shunt	r_0662No2	RHTO_05768	M7WLD5	isocitrate lyase
IDH1	TCA cycle	r_0658No1	RHTO_01289	M7XE28	isocitrate dehydrogenase [nad] subunit, mitochondrial
IDH2	TCA cycle	r_0658No2	RHTO_01290	M7WW42	isocitrate dehydrogenase [nad] subunit, mitochondrial
IDP	TCA cycle	r_2131No1	RHTO_04315	M7WN97	isocitrate dehydrogenase (nadp+), mitochondrial
KGD1*	TCA cycle	r_0831No1,r_0 832No1	RHTO_02312	M7WKF8	2-oxoglutarate dehydrogenase e2 component (dihydrolipoamide succinyltransferase)
KGD2*	TCA cycle	r_0831No1,r_0 832No1	RHTO_07893	M7WR40	Dihydrolipoyl dehydrogenase (ec 1.8.1.4)

Gene name	Subsystem	Reaction ID	Locus tag	Protein accession (UniProt)	Protein name
LPD1*	TCA cycle	r_0831No1,r_0 832No1	RHTO_07860	M7XGI7	2-oxoglutarate dehydrogenase e1 component (succinyl-transferring)
LSC1*	TCA cycle	r_1022No1	RHTO_05264	M7WM30	succinate--coa ligase [adp-forming] subunit beta, mitochondrial (ec 6.2.1.5) (succinyl-coa synthetase beta chain) (scs-beta)
LSC2*	TCA cycle	r_1022No1	RHTO_01205	M7WVW2	succinate--coa ligase [adp-forming] subunit alpha, mitochondrial (ec 6.2.1.5) (succinyl-coa synthetase subunit alpha) (scs-alpha)
MDH1	TCA cycle	r_0713No1	RHTO_04363	M7XHF8	malate dehydrogenase (ec 1.1.1.37)
MDH2	Pyruvate metabolism	G,X: r_0714_REVNo1, A:r_0714No1	RHTO_03745	M7WQ86	l-malate dehydrogenase
ME	Pyruvate metabolism	t_0027No1	RHTO_03795	M7WHN9	malic enzyme
MS	Glyoxylate shunt	r_0716No1	RHTO_01459	M7WMR9	malate synthase
ODC	other	r_2132No1	RHTO_03731	M7WI80	Mitochondrial 2-oxodicarboxylate carrier
PDC	Pyruvate metabolism	r_0959No1	RHTO_00098	M7X6G2	pyruvate decarboxylase
PDH	Pyruvate metabolism	r_0961No1	RHTO_03543	M7WIG9	acetyltransferase component of pyruvate dehydrogenase complex (ec 2.3.1.12)
PDH	Pyruvate metabolism	r_0961No1	RHTO_01852	M7WTI0	pyruvate dehydrogenase e1 component subunit alpha (ec 1.2.4.1)

Gene name	Subsystem	Reaction ID	Locus tag	Protein accession (UniProt)	Protein name
PDH	Pyruvate metabolism	r_0961No1	RHTO_07250	M7WZJ7	pyruvate dehydrogenase e1 component subunit beta (ec 1.2.4.1)
PDH	Pyruvate metabolism	r_0961No1	RHTO_01754	M7XMA7	pyruvate dehydrogenase x component
PEPCK	Pyruvate metabolism	r_0884No1	RHTO_07480	M7XSC4	Phosphoenolpyruvate carboxykinase (atp)
PFK	PPP	G:r_0886No1, X:r_0887No1	RHTO_00494	M7X6S3	Atp-dependent 6-phosphofructokinase (atp-pfk) (phosphofructokinase) (ec 2.7.1.11) (phosphohexokinase)
PGK	glycolysis	G,X: r_0892No1, A: r_0892_REVNo1	RHTO_00033	M7X689	phosphoglycerate kinase (ec 2.7.2.3)
PGIS	PPP	r_0091No1	RHTO_07939	M7WWW0	6-phosphogluconolactonase (6pgl) (ec 3.1.1.31)
PGM	glycolysis	G,X: r_0893No3, A: r_0893_REVNo3	RHTO_07773	M7XRQ3	protein of phosphoglycerate mutase 1 family
PKT	PPP	t_0081No1	RHTO_04463	M7WGA7	phosphoketolase
PYC	Pyruvate metabolism	r_0958No2	RHTO_02628	M7WS17	pyruvate carboxylase (ec 6.4.1.1)
PYC	Pyruvate metabolism	r_0958No1	RHTO_01350	M7XNE0	urea carboxylase / allophanate hydrolase
PYK	glycolysis	r_0962No1	RHTO_01610	M7WUI5	pyruvate kinase (ec 2.7.1.40)
RK	xylose metabolism	t_0885No1	RHTO_00950	M7WVT6	protein of carbohydrate kinase, fgg type family
RPE	PPP	r_0984No1, except Anlim	RHTO_05984	M7XEA2	Ribulose-phosphate 3-epimerase (ec 5.1.3.1)
RPI	PPP	r_0982No1	RHTO_06311	M7WUP8	ribose 5-phosphate isomerase a

Gene name	Subsystem	Reaction ID	Locus tag	Protein accession (UniProt)	Protein name
SDH1*	TCA cycle		RHTO_07066	M7X332	succinate dehydrogenase assembly factor 2, mitochondrial (sdh assembly factor 2) (sdhaf2)
SDH2*	TCA cycle	r_1021No1	RHTO_05714	M7X560	succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (ec 1.3.5.1)
SDH4*	TCA cycle	r_1021No1	RHTO_06068	M7XJJ4	succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial (ec 1.3.5.1)
SFC	other	r_1265No1	RHTO_05007	M7WW62	mitochondrial carrier protein, succinate:fumarate antiporter
SHD3*	TCA cycle	r_1021No1	RHTO_00534	M7X6W8	succinate dehydrogenase (ubiquinone) cytochrome b subunit
TAL	PPP	r_1048No1, except Anlim	RHTO_06955	M7X0R7	transaldolase (ec 2.2.1.2)
TKT1	PPP	r_1049No1	RHTO_03248	M7XNL9	transketolase (ec 2.2.1.1)
TKT2	PPP	r_1049No2	RHTO_03251	M7WY13	transketolase
TPI	PPP	G,X: r_1054No1; A: r_1054_REVNo 1	RHTO_01329	M7WME7	triosephosphate isomerase (ec 5.3.1.1)
XDH	xylose metabolism	r_1092No1	RHTO_01970	M7WT79	xylitol dehydrogenase
XK	xylose metabolism	r_1094No1	RHTO_04556	M7X6R2	xylulokinase
XR	Xylose metabolism	r_1093No1	RHTO_03963	M7X8C7	xylose reductase
ZWF	PPP	r_0466No1	RHTO_07853	M7WR01	Glucose-6-phosphate 1-dehydrogenase (ec 1.1.1.49)

Table S4. Fatty acid profiles in *R. toruloides* wild-type strain CCT 7815 batch bioreactor cultivations in low-nitrogen chemically defined media with glucose (G), xylose (X), or acetate (A) as a carbon source during exponential growth (exp) and nitrogen-limited (Nlim) phases. The mean and standard deviation of duplicate experiments is presented. Abbreviations: C14:0, myristic acid; C16:0, palmitic acid; C16:1, palmitoleic acid; C18:0, stearic acid; C18:1, oleic acid; C18:2, linoleic acid; C18:3, α -Linolenic acid.

Condition	Yields on biomass (mg gCDW ⁻¹)		Major fatty acid residues (mg _{FA} gCDW ⁻¹)						
	Total lipid	Total fatty acid	C14:0	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3
Gexp	57 ± 2	51 ± 2	ND	7.5 ± 0.34	ND	2.0 ± 0.24	16.2 ± 1.2	19.8 ± 0.37	5.7 ± 0.16
GNlim	483 ± 41	435 ± 36	3.0 ± 0.11	101.9 ± 6.3	ND	32.3 ± 3.6	233.5 ± 21.2	53.6 ± 4.4	10.8 ± 1.1
Xexp	202 ± 25	182 ± 23	1.7 ± 0.24	47.8 ± 6.2	2.0 ± 0.31	10.4 ± 1.4	78.1 ± 10	36.5 ± 3.9	5.7 ± 0.71
XNlim	290 ± 6	261 ± 6	0.34 ± 0.011	7.44 ± 0.22	0.32 ± 0.0043	1.61 ± 0.077	11.35 ± 0.25	4.35 ± 0.074	0.66 ± 0.017
Aexp	175 ± 19	158 ± 17	ND	25.2 ± 2.6	ND	18.0 ± 2.4	67.9 ± 12.3	37.3 ± 0.19	9.6 ± 0.16
ANlim	341 ± 13	307 ± 12	ND	57.3 ± 0.21	ND	37.6 ± 2.5	160.1 ± 7.8	42.1 ± 1.3	9.4 ± 0.30

Table S5. Enzymes under flexibilization of individual protein constraint by the ecGEM of *R. toruloides* in low-nitrogen chemically defined media containing glucose (G), xylose (X), or acetate (A) as the carbon source during exponential growth (exp) and nitrogen-limited (Nlim) phases. Constraints were adjusted by algorithmically increasing measured absolute protein concentration to achieve feasible flux states in proteome-constrained ecGEM simulations. Detailed report on previous values and modified values available at Github repository github.com/alinarekena/results/generate_protModels. Modified from Rekēna et al., PLOS Comput. Biol. 19(4):e1011009, 2023, under a CC BY licence.

Protein names	Locus tag	Protein accession (UniProt)
mitochondrial carrier protein, tricarboxylate carrier	RHTO_05609	M7WEL7
1,3-beta-glucan synthase, glycosyltransferase family 48 protein	RHTO_05280	M7WFB4
arginine biosynthesis bifunctional protein argJ, mitochondrial [cleaved into: arginine biosynthesis bifunctional protein argJ]	RHTO_05134	M7WFJ0
2,5-diamino-6-(ribosylamino)-4(3h)-pyrimidinone 5-phosphate reductase	RHTO_04544	M7WG08
atp citrate synthase	RHTO_03915	M7WHC9
mitochondrial 2-oxodicarboxylate carrier	RHTO_03731	M7WI80
carnitine o-acetyltransferase	RHTO_01903	M7WLQ0
carnitine acyl carnitine carrier, mitochondrial	RHTO_01354	M7WMH0
mitochondrial carrier protein	RHTO_04513	M7WMP3
sterol 24-c-methyltransferase	RHTO_00856	M7WPW0

Protein names	Locus tag	Protein accession (UniProt)
fatty acid synthase subunit beta, fungi type	RHTO_02032	M7WSW5
dol-p-man:protein alpha-mannosyltransferase, glycosyltransferase family 39 protein	RHTO_02306	M7WT33
phospholipid:diacylglycerol acyltransferase	RHTO_01945	M7WT53
cytochrome c oxidase, subunit via	RHTO_01605	M7WUI0
protein of carbohydrate kinase, fgg type family	RHTO_00950	M7WVT6
mitochondrial carrier protein, succinate:fumarate antiporter	RHTO_05007	M7WW62
5-methyltetrahydropteroylglutamate-homocysteine s-methyltransferase	RHTO_07691	M7WXM9
glutamate dehydrogenase	RHTO_04650	M7X2B5
fructose-bisphosphate aldolase (fbp aldolase) (ec 4.1.2.13)	RHTO_03043	M7X5F4
long-chain acyl-coa synthetase	RHTO_00058	M7X6B9
aconitase hydratase, mitochondrial (aconitase) (ec 4.2.1.-)	RHTO_00539	M7X6X3
l-xylulose reductase	RHTO_00373	M7X791
d-arabinitol dehydrogenase	RHTO_07844	M7XGH5
3-hydroxy-3-methylglutaryl coenzyme a reductase (hmg-coa reductase) (ec 1.1.1.34)	RHTO_04045	M7XI04
delta-9 fatty acid desaturase	RHTO_03911	M7XI95
phosphoribosylformylglycinamidine synthase	RHTO_02389	M7XKV7
fatty acid synthase subunit alpha, fungi type	RHTO_02139	M7XM89
dolichol-phosphate mannosyltransferase subunit 1 (ec 2.4.1.83)	RHTO_06635	M7WJE8
farnesyl-diphosphate farnesyltransferase	RHTO_02590	M7WJI1
chorismate synthase (ec 4.2.3.5)	RHTO_05784	M7WKX0
phosphomevalonate kinase (ec 2.7.4.2)	RHTO_02073	M7WL39
homoaconitase, mitochondrial (ec 4.2.1.36) (homoaconitate hydratase)	RHTO_05318	M7WMT2
c-22 sterol desaturase	RHTO_01252	M7WNB5
glyceraldehyde 4-phosphate dehydrogenase	RHTO_01292	M7WNF8
aconitase hydratase, mitochondrial (aconitase) (ec 4.2.1.-)	RHTO_08030	M7WQ73
multifunctional tryptophan biosynthesis protein [includes: anthranilate synthase component 2 (as) (ec 4.1.3.27) (anthranilate)]	RHTO_02564	M7WRD1
short-chain dehydrogenase/reductase sdr family protein	RHTO_07171	M7WS89
adenylosuccinate synthetase (ampsase) (adss) (ec 6.3.4.4) (imp-- aspartate ligase)	RHTO_02257	M7WSH3
succinate-semialdehyde dehydrogenase (ec 1.2.1.16)	RHTO_02062	M7WSZ7
asparaginyl-tRNA synthetase	RHTO_04575	M7WUY3
phosphoribosylamine--glycine ligase / phosphoribosylformylglycinamidine cyclo-ligase	RHTO_01050	M7WW40

Protein names	Locus tag	Protein accession (UniProt)
transketolase	RHTO_03251	M7WY13
atp phosphoribosyltransferase	RHTO_07750	M7WYE8
nadh-ubiquinone oxidoreductase 64 kda subunit	RHTO_07075	M7WZ39
diphosphomevalonate decarboxylase (ec 4.1.1.33)	RHTO_06005	M7X109
dihydroorotate	RHTO_04690	M7X2F3
c-1-tetrahydrofolate synthase	RHTO_02988	M7X4L7
succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (ec 1.3.5.1)	RHTO_05714	M7X560
phosphoribosyl-atp pyrophosphohydrolase / phosphoribosyl-amp cyclohydrolase	RHTO_00514	M7X6U2
nadh dehydrogenase	RHTO_00388	M7X7A4
glycine hydroxymethyltransferase	RHTO_04131	M7X852
squalene monooxygenase	RHTO_01745	M7XBZ9
ribose-phosphate pyrophosphokinase	RHTO_04328	M7XHB5
glutamyl-trna synthetase	RHTO_03621	M7XIT2
acetylornithine aminotransferase	RHTO_07003	M7XJ70
ribose-phosphate pyrophosphokinase	RHTO_02591	M7XKF0
anthranilate phosphoribosyltransferase	RHTO_07763	M7XRP3
ketol-acid reductoisomerase (fragment)	RHTO_04566	M7XVW7
gmp synthase (glutamine-hydrolysing)	RHTO_00066	M7XZ93

Table S6. Fatty acid profiles in *R. toruloides* NBRC 0880 (formerly known as IFO 0880) wild type and mutant strains cultivated in Falcon tube bioreactors in low-nitrogen chemically defined media with glucose (G), xylose (X), or acetate (A) as a carbon source during nitrogen-limited (Nlim) phase. **A:** Growth and lipid metrics (titer, lipid yield on biomass, harvest time). **B:** Fatty acid composition. Strains include ATP citrate lyase knockout (Δ ACL), phosphoketolase knockout (Δ PK), and cytosolic malic enzyme knockout (Δ cMAE). The mean and standard deviation of triplicate experiments is presented. Abbreviations: C16:0, palmitic acid; C18:0, stearic acid; C18:1, oleic acid; C18:2, linoleic acid; C18:3, α -Linolenic acid. Layouts differ from the published versions for copyright compliance.

A.

Condition	Strain	Titer (g L ⁻¹)	Yield on biomass (mg gCDW ⁻¹)	Harvest time (h)
		CDW	Total lipid	
GNlim	wild type	10.23 \pm 0.15	501.8 \pm 3.7	54
	Δ ACL	2.34 \pm 0.065	94.3 \pm 0.39	54
	Δ PK	10.23 \pm 0.15	505.2 \pm 9.4	54
	Δ cMAE	9.62 \pm 0.26	516.4 \pm 5.0	48
XNlim	wild type	6.24 \pm 0.13	426.1 \pm 8.4	90
	Δ ACL	7.28 \pm 0.39	377.1 \pm 30.1	81

	ΔPK	5.76±0.075	429.4±7.3	91
	ΔcMAE	6.93±0.27	439.4±8.2	80
ANlim	wild type	2.53±0.03	450.7±11.5	66
	ΔACL	1.63±0.03	200.2±11.8	66
	ΔPK	2.70±0.09	462.3±44.5	66
	ΔcMAE	2.62±0.12	447.9±10.5	126

B.

Condition	Strain	Major fatty acid residues (g _{FA} gCDW ⁻¹)					Yield on biomass (mg gCDW ⁻¹)
		C16:0	C18:0	C18:1	C18:2	C18:3	
GNlim	wild type	102.4±0.83	57±0.63	206.0±1.2	51.4±0.75	16.2±0.13	433.0±3.5
	ΔACL	13.8±0.070	8.0±0.090	16.5±0.35	25.0±0.29	15.09±0.18	78.4±0.98
	ΔPK	103.2±1.7	57.5±1.6	208.0±3.9	51.3±0.77	15.9±0.25	435.9±8.1
XNlim	wild type	106.0±1.4	54.5±0.58	214.0±2.4	53.4±0.43	18.1±0.23	446.0±5.0
	ΔACL	83.6±1.8	43.8±0.79	166.9±3.1	56.2±1.4	12.6±0.31	363.1±7.1
	ΔPK	96.0±7.6	32.2±2.8	139.4±11.3	49.8±3.8	7.4±0.62	324.7±25.9
ANlim	wild type	87.8±1.3	44.9±0.64	168.6±3.7	58.7±1.3	13.9±0.18	373.9±6.9
	ΔACL	70.2±2.8	74.5±1.9	186.8±4.2	39.9±1.1	16.5±0.63	387.7±10.2
	ΔPK	31.1±1.8	28.3±2.3	59.6±5.1	35.4±0.58	14.6±0.16	169.0±10.0
	ΔcMAE	77.1±7.6	77.6±8.2	186.3±16.5	39.1±4.3	17.3±2.0	397.4±38.4
		73.6±0.31	74.5±1.5	181.8±6.8	38.3±0.82	16.9±0.11	385.0±9.1

Supplementary figures

Supplementary figures to Study 1

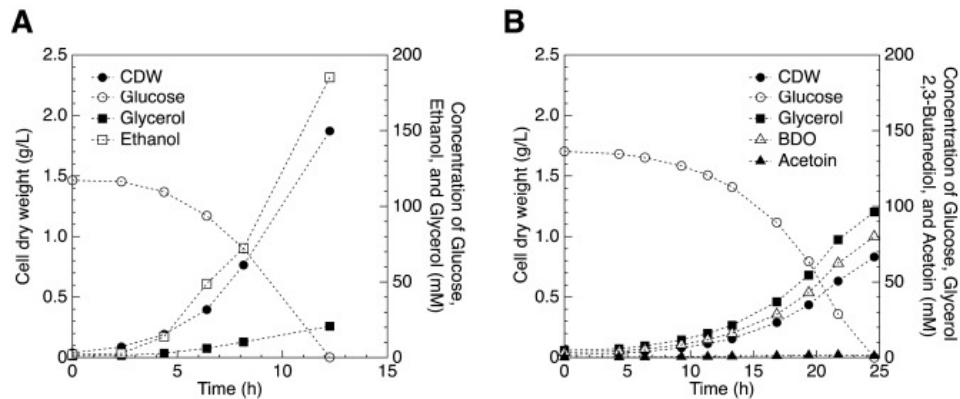


Figure S1. Anaerobic batch cultivation of *S. cerevisiae* in stirred tank reactors at 30 °C, pH 5.0. **A:** Representative cultivation profiles of reference strain GSY013. **B:** representative cultivation profiles of engineered 23BD-glycerol co-producing strain GSY014. The mean of six replicates is presented. Reproduced from Sjöberg et al., Metabolic Engineering 82,49-59, 2024, under a CC BY licence.

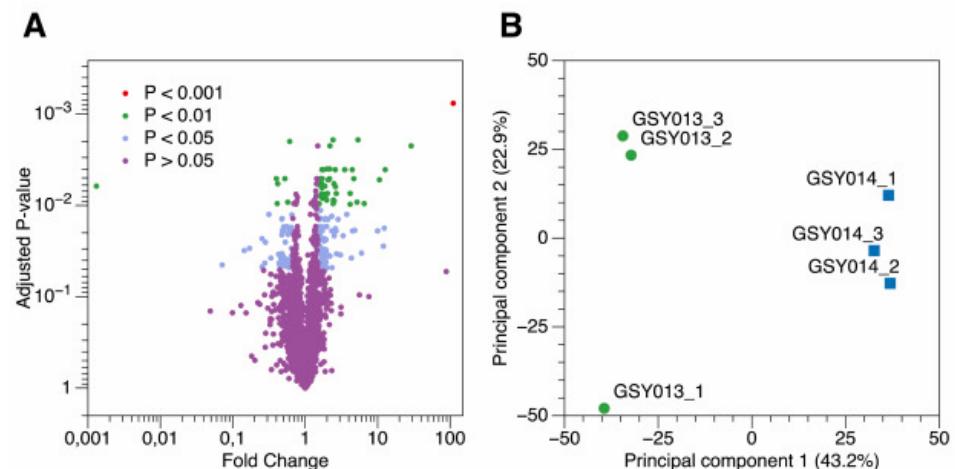


Figure S2. Absolute proteome quantification in *S. cerevisiae* anaerobic bioreactor batch cultures of the reference strain GSY013 and 23BD-glycerol co-producing strain GSY014. **A:** Volcano plot comparing absolute protein abundances of three replicates in GSY013 and GSY014 strains. **B:** Principal component analysis. P denotes adjusted p value according to Benjamini-Hochberg (Benjamini & Hochberg, 1995b). Reproduced from Sjöberg et al., Metabolic Engineering 82,49-59, 2024, under a CC BY licence.

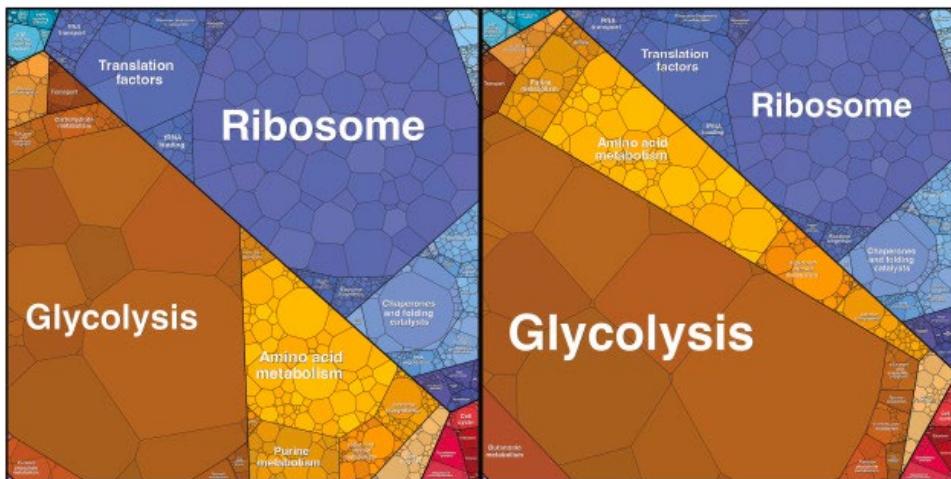


Figure S3. Proteomaps (Liebermeister et al., 2014) representing allocation of mean protein abundance of three replicates to major functional groups. Left: Proteome allocation of the reference strain GSY013. Right: Proteome allocation of 23BD-glycerol co-producing strain GSY014. Reproduced from Sjöberg et al., *Metabolic Engineering* 82, 49-59, 2024, under a CC BY licence.

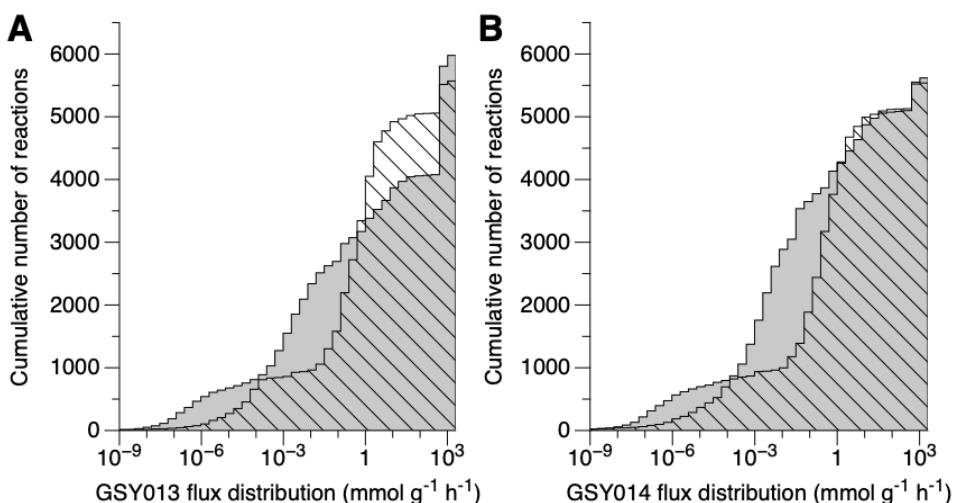
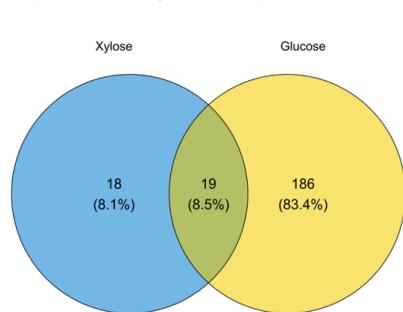


Figure S4. Flux variability analysis of ecGEMs. A: Cumulative flux distribution of the reference strain GSY013. B: Cumulative flux distribution of 23BD-glycerol co-producing strain GSY014. Grey denotes proteome-constrained ecGEM; dashed denotes total-protein ecGEM. Reproduced from Sjöberg et al., *Metabolic Engineering* 82, 49-59, 2024, under a CC BY licence.

Supplementary figures to Study 2

A. Exponential vs Nitrogen limitation phase



B. Exponential phase

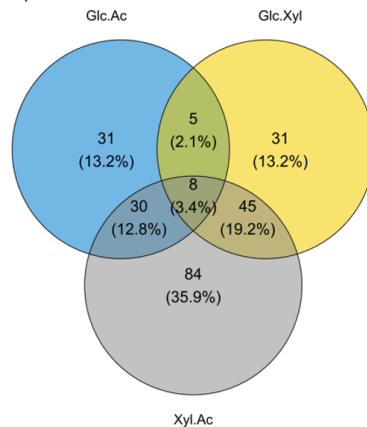
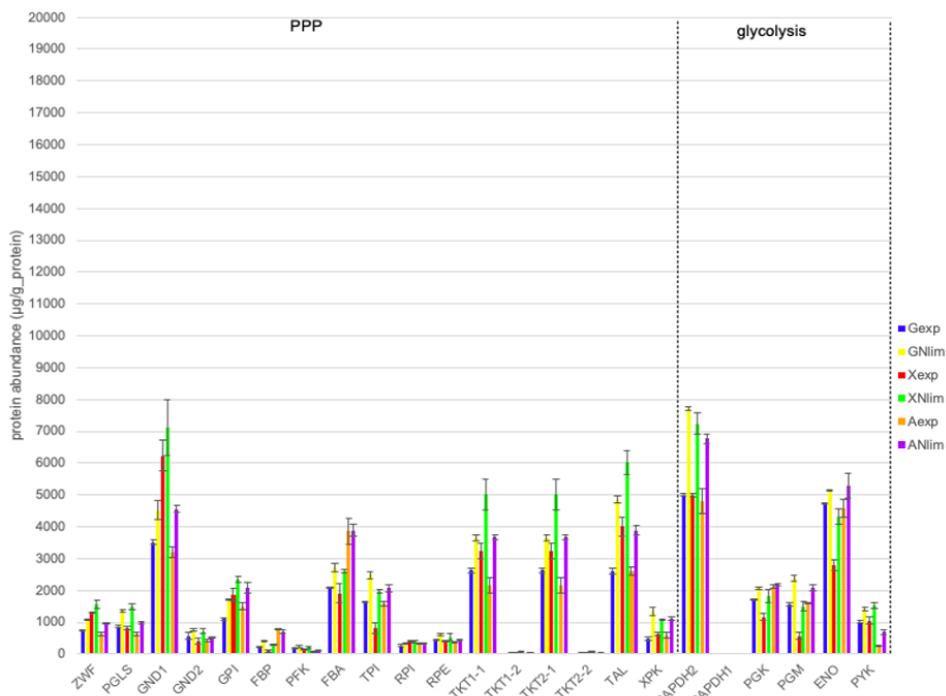


Figure S5. Absolute proteome quantification in *R. toruloides* wild-type strain CCT 7815 batch bioreactor cultures. **A.** Venn diagram comparing the number of significantly differentially expressed proteins of two replicates between nitrogen limitation and exponential growth phases. **B.** Venn diagram comparing the number of significantly differentially expressed proteins of two replicates during the exponential growth phase in media with different carbon sources. Reproduced from Rekēna et al., *PLOS Comput. Biol.* 19(4):e1011009, 2023, under a CC BY licence.



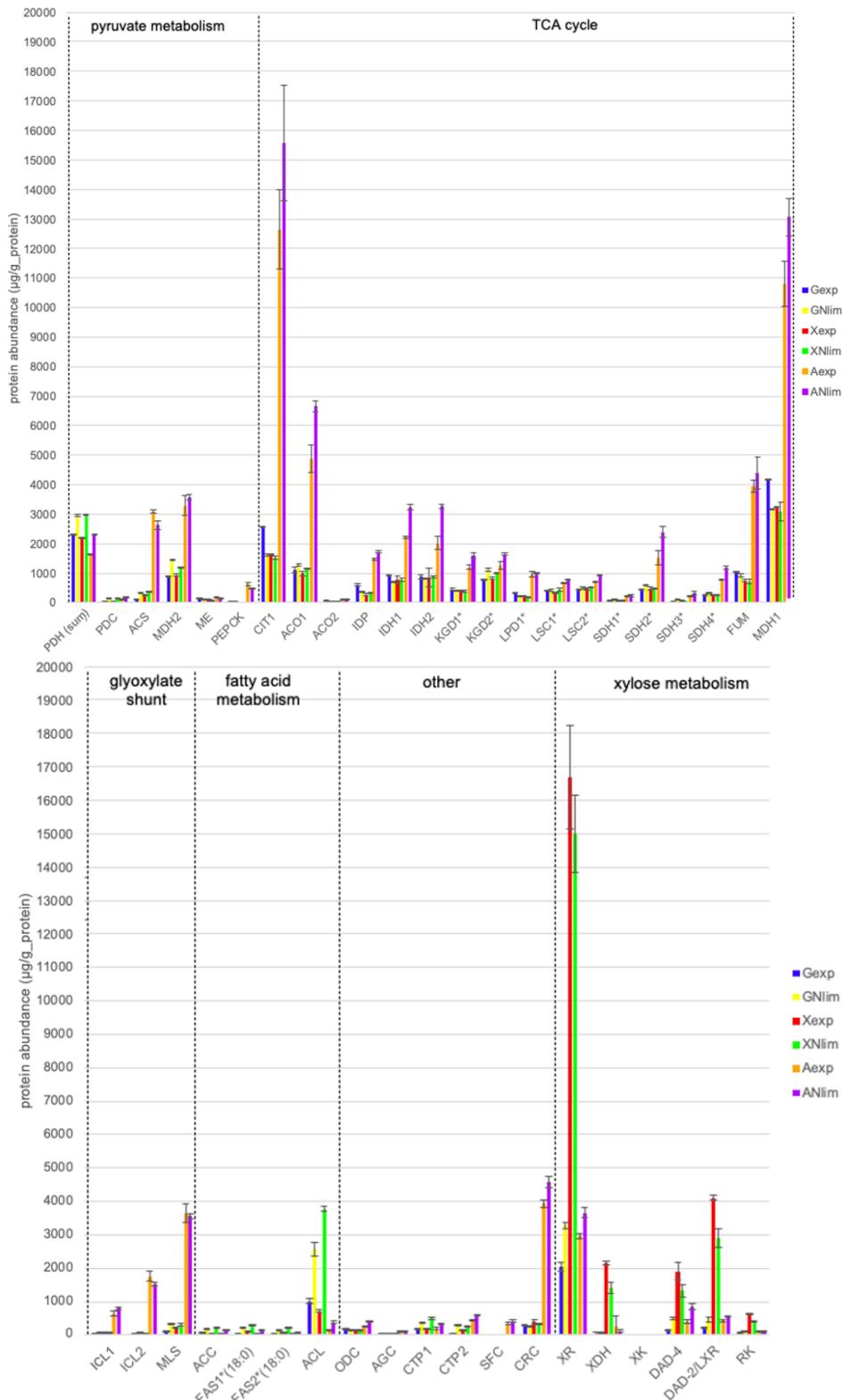


Figure S6. Absolute enzyme concentrations ($\mu\text{g g}_{\text{total_protein}}^{-1}$) (Sánchez et al., 2021) in *R. toruloides* wild-type strain CCT 7815 cultivated in batch bioreactors with low-nitrogen chemically defined media containing glucose (G), xylose (X), or acetate (A) as the carbon source during exponential growth (exp) and nitrogen-limited (Nlim) phases. Mean and standard deviation of duplicate experiments is presented. Abbreviations: PPP, pentose phosphate pathway; TCA, tricarboxylic acid. Gene-protein names are defined in **Table S3**. Reproduced from Rekēna et al., PLOS Comput. Biol. 19(4):e1011009, 2023, under a CC BY licence.

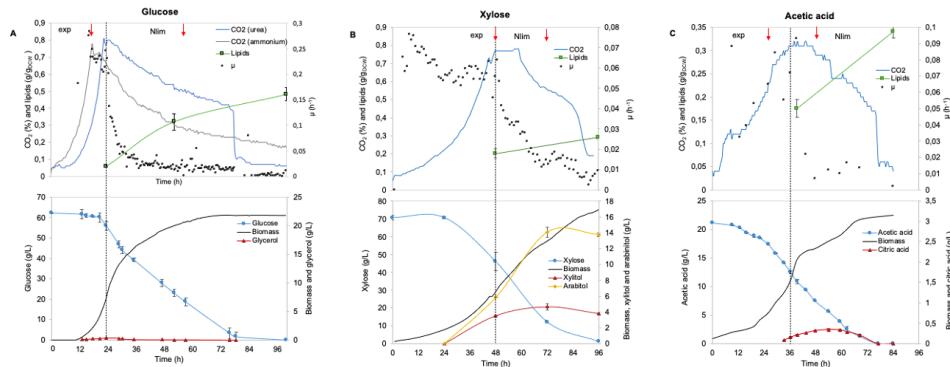


Figure S7. Batch cultivation of *R. toruloides* in aerobic stirred tank bioreactors at 30°C, pH 6.0. Representative cultivation profiles of wild type strain CCT 7815 grown in low-nitrogen chemically defined media during exponential growth (exp) and nitrogen-limited (Nlim) phases. **A:** Representative cultivation profiles of glucose-grown cells (C/N ratio of 68.6 mol mol⁻¹). **B:** Representative cultivation profiles of xylose-grown cells (C/N ratio of 80). **C:** Representative cultivation profiles of acetate-grown cells (C/N ratio of 80). The mean and standard deviation of duplicate experiments are presented. In cases where a single measurement is shown for clarity, replicate datasets were nevertheless obtained and used for data interpretation. Reproduced from Rekēna et al., PLOS Comput. Biol. 19(4):e1011009, 2023, under a CC BY licence.

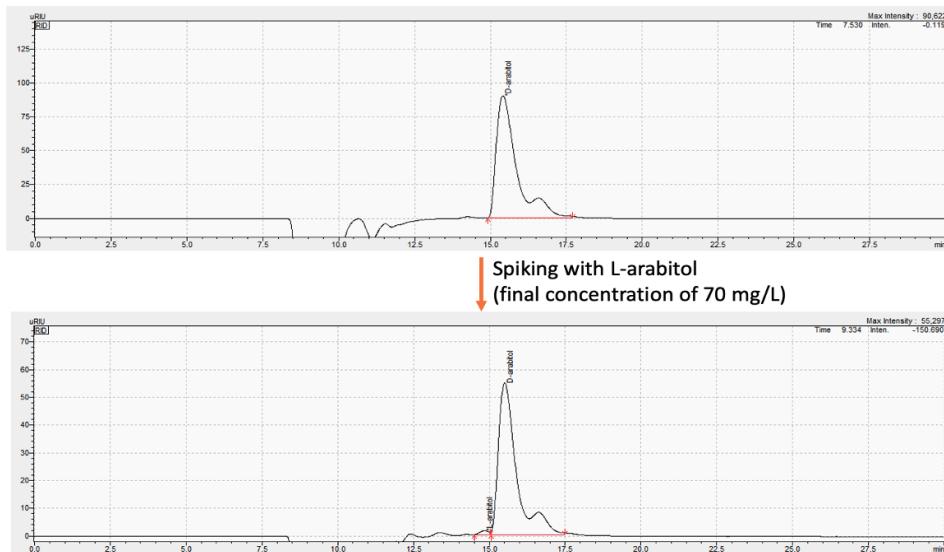


Figure S8. D-arabitol detection in the cultivation broth of *R. toruloides* wild-type strain CCT 7815 grown on xylose as the sole carbon source (70 g L⁻¹, C/N ratio of 80). The diagrams show HPLC chromatograms obtained using a refractive index detector (RID) and a Chiralpak column, with

hexane-ethanol (70:30, v/v) as the mobile phase at a flow rate of 0.3 mL min⁻¹ and a column temperature of 20 °C. Reproduced from Rekēna et al., PLOS Comput. Biol. 19(4):e1011009, 2023, under a CC BY licence.

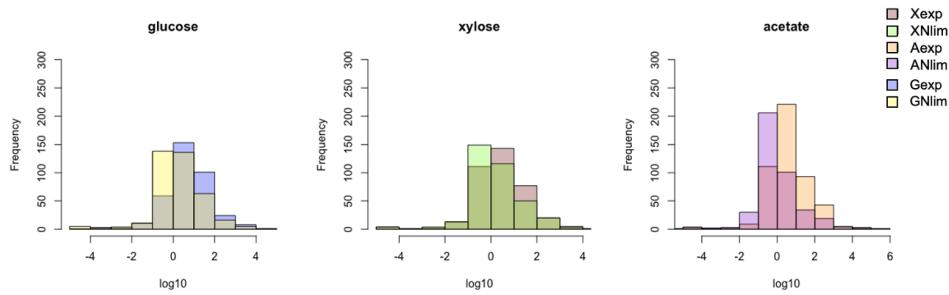
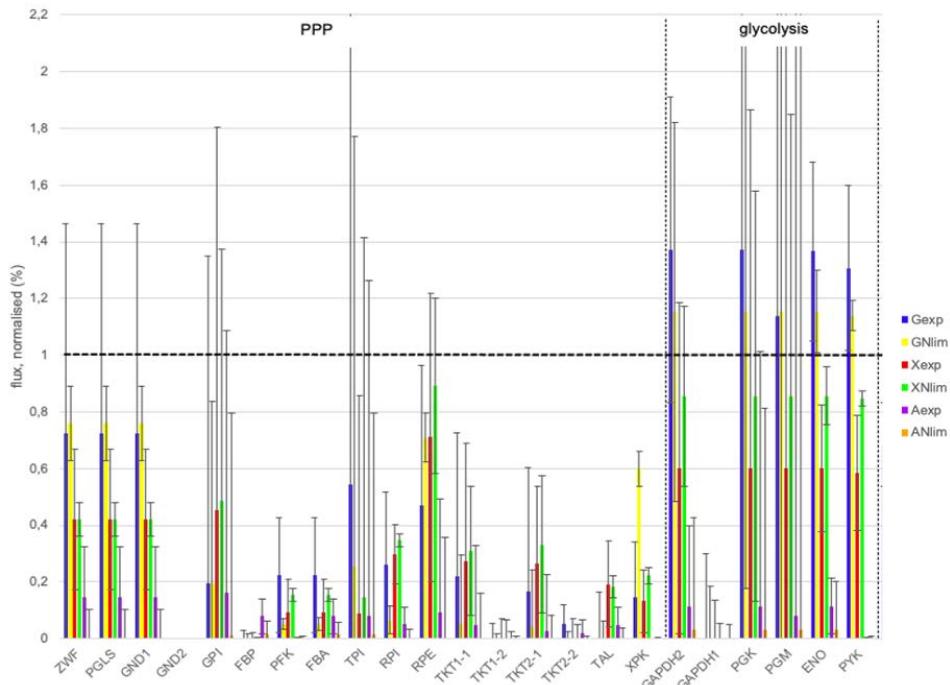
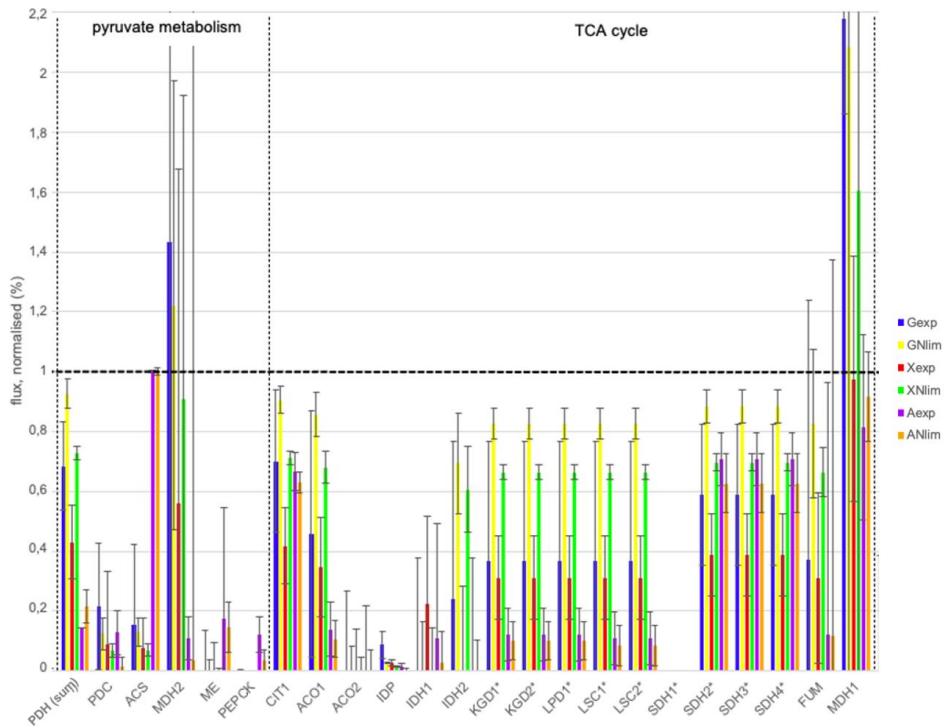


Figure S9. Histograms of apparent enzyme catalytic activities, k_{app} (s⁻¹) of *R. toruloides* grown in low-nitrogen chemically defined media with glucose (G), xylose (X), or acetate (A) as a carbon source during exponential growth (exp) and nitrogen-limited (Nlim) phases. k_{app} values were calculated as the ratio of metabolic flux (mmol (gCDW h)⁻¹) to the corresponding protein abundance (mmol gCDW⁻¹). Reproduced from Rekēna et al., PLOS Comput. Biol. 19(4):e1011009, 2023, under a CC BY licence.





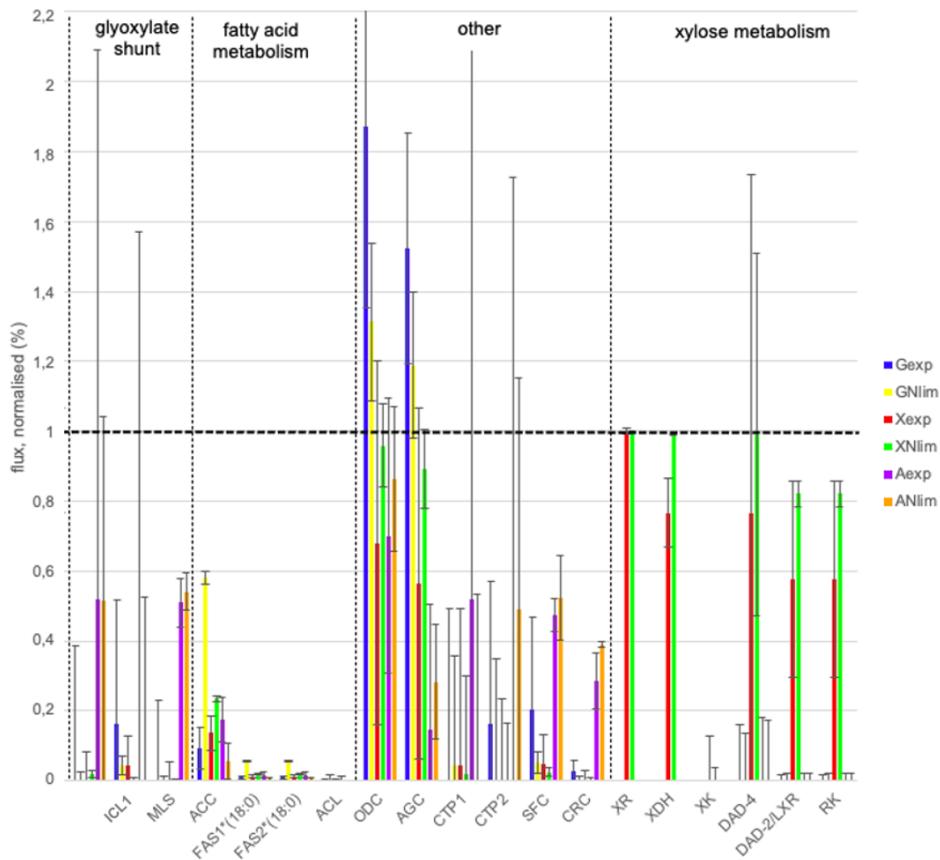


Figure S10. Predicted fluxes ($\text{mmol (gCDW h)}^{-1}$) of *R. toruloides* wild-type strain CCT 7815 cultivated in batch bioreactors with low-nitrogen chemically defined media containing glucose (G), xylose (X), or acetate (A) as the carbon source during exponential growth (exp) and nitrogen-limited (Nlim) phases, simulated using condition-specific, proteomics-constrained ecGEMs of *R. toruloides*. Fluxes were normalized to the specific substrate uptake rate, representing the percentage of carbon distribution. Median and standard deviation of 2000 flux samples are presented. Abbreviations: PPP, pentose phosphate pathway; TCA, tricarboxylic acid. Gene-protein names are defined in **Table S3**. Reproduced from Rekēna et al., *PLOS Comput. Biol.* 19(4):e1011009, 2023, under a CC BY licence.

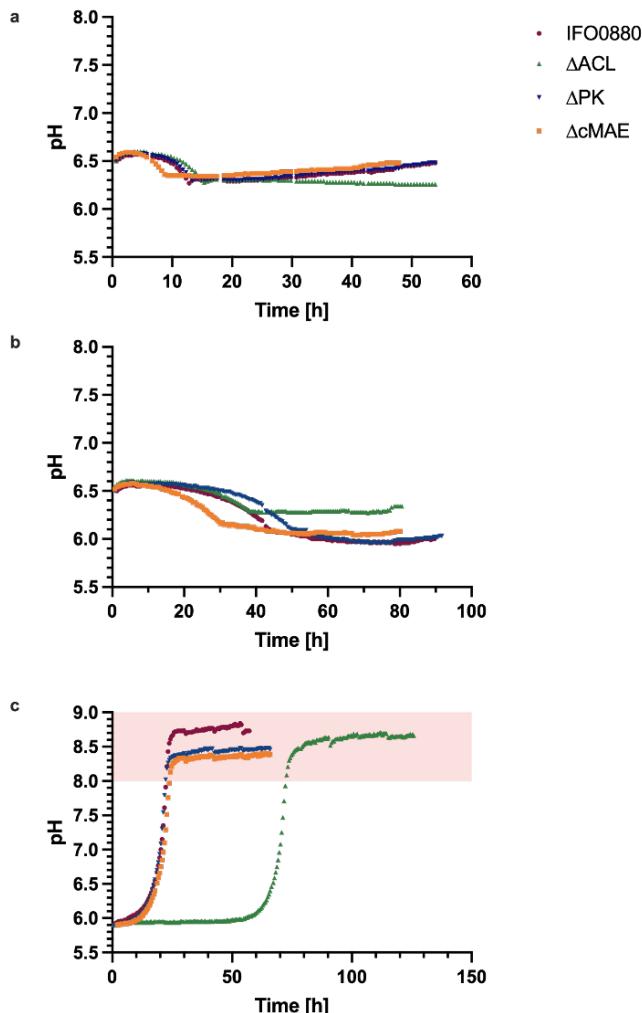


Figure S11. pH tracking data of *R. toruloides* NBRC 0880 (formerly known as IFO 0880) wild type and mutant strains cultivated in Falcon tube bioreactors at 30°C. **A:** Representative pH profiles of glucose-grown strains. **B:** Representative pH profiles of xylose-grown strains. **C:** Representative pH profiles of acetate-grown strains. Strains include ATP citrate lyase knockout (Δ ACL, green), phosphoketolase knockout (Δ PK, blue), and cytosolic malic enzyme knockout (Δ cMAE, orange). pH was measured real time using a non-invasive optical sensor. Pink shading denotes values outside the sensor's working range. Each strain was cultivated in three biological replicates, of which one tube was equipped with the pH sensor and used for recording shown profiles. Reproduced from Rekēna et al., *Appl Microbiol Biotechnol* 109:77, 2025, under a CC BY NC ND licence.

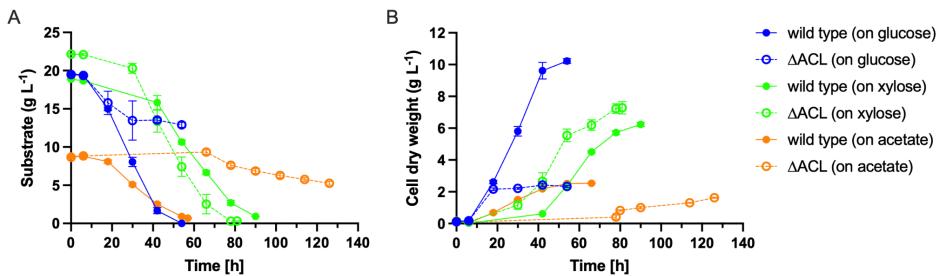


Figure S12. Batch cultivation of *R. toruloides* in Falcon tube bioreactors at 30 °C. Representative cultivation profiles of wild type (filled circle) and mutant strains (open circle), including ATP citrate lyase knockout (Δ ACL), phosphoketolase knockout (Δ PK), and cytosolic malic enzyme knockout (Δ cMAE) in low-nitrogen chemically defined media with glucose (blue), xylose (green), or acetate (orange) as a carbon source. **A:** Residual substrate concentration (g L^{-1}). **B:** Cell dry weight concentration (g L^{-1}). The mean and standard deviation of three replicates are presented. Layouts differ from the published versions for copyright compliance.

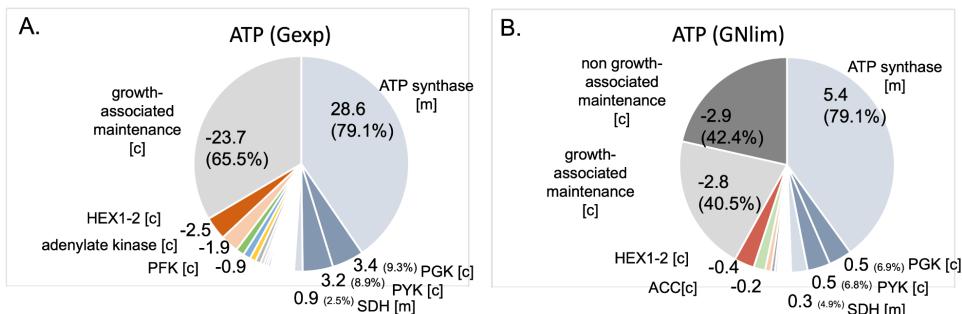


Figure S13. Predicted fluxes of ATP ($\text{mmol (gCDW h)}^{-1}$) of *R. toruloides* wild-type strain CCT 7815 cultivated in batch bioreactors with low-nitrogen chemically defined media containing glucose (G) as the carbon source, simulated using condition-specific, proteomics-constrained ecGEMs. **A:** ATP distribution during exponential growth (exp) phase. **B:** ATP distribution during nitrogen-limited (Nlim) phase. Negative flux denotes metabolite consumption, positive flux denotes metabolite production. Percentages indicate the relative contribution of each flux to the total ATP turnover under the given condition. [m] and [c] denote compartments of mitochondria and cytoplasm. Abbreviations: HEX, hexokinase; PFK, phosphofructokinase; ACC, acetyl-CoA carboxylase; SDH, succinate dehydrogenase; PYK, pyruvate kinase; PGK, phosphoglycerate kinase.

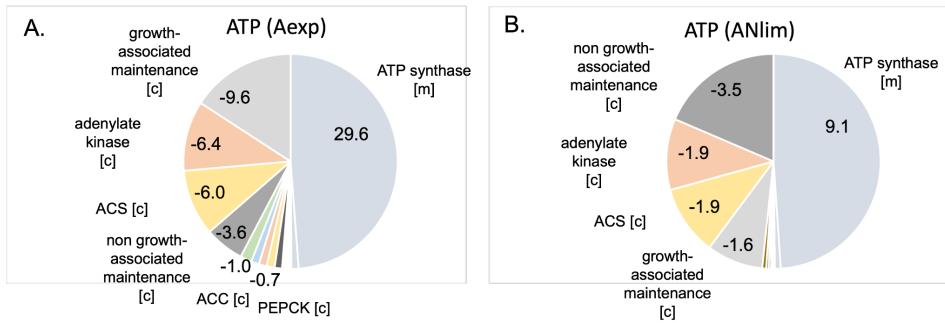
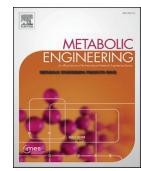


Figure S14. Predicted fluxes of ATP (mmol (gCDW h)⁻¹) of *R. toruloides* wild-type strain CCT 7815 cultivated in batch bioreactors with low-nitrogen chemically defined media containing acetate (A) as the carbon source, simulated using condition-specific, proteomics-constrained ecGEMs. **A:** ATP distribution during exponential growth (exp) phase. **B:** ATP distribution during nitrogen-limited (Nlim) phase. Negative flux denotes metabolite consumption, positive flux denotes metabolite production. [m] and [c] denote compartments of mitochondria and cytoplasm. Abbreviations: ACS, acetyl-CoA synthetase; ACC, acetyl-CoA carboxylase; PEPCK, phosphoenolpyruvate carboxykinase. Reproduced from Rekēna et al., PLOS Comput. Biol. 19(4):e1011009, 2023, under a CC BY licence.

Appendix 2

Publication I

Sjöberg, G., Rekēna, A., Fornstad, M., Lahtvee, P.-J., van Maris, A. J. A. (2024). Evaluation of enzyme-constrained genome-scale model through metabolic engineering of anaerobic co-production of 2,3-butanediol and glycerol by *Saccharomyces cerevisiae*. *Metabolic Engineering*, 82, 49–59.



Evaluation of enzyme-constrained genome-scale model through metabolic engineering of anaerobic co-production of 2,3-butanediol and glycerol by *Saccharomyces cerevisiae*

Gustav Sjöberg ^a, Alīna Reķēna ^b, Matilda Fornstad ^a, Petri-Jaan Lahtvee ^b, Antonius J.A. van Maris ^{a,*}

^a Department of Industrial Biotechnology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden

^b Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia



ARTICLE INFO

Keywords:
2,3-Butanediol
Enzyme-constrained genome-scale model
Anaerobic
Proteomics
Metabolic engineering

ABSTRACT

Enzyme-constrained genome-scale models (ecGEMs) have potential to predict phenotypes in a variety of conditions, such as growth rates or carbon sources. This study investigated if ecGEMs can guide metabolic engineering efforts to swap anaerobic redox-neutral ATP-providing pathways in yeast from alcoholic fermentation to equimolar co-production of 2,3-butanediol and glycerol. With proven pathways and low product toxicity, the ecGEM solution space aligned well with observed phenotypes. Since this catabolic pathway provides only one-third of the ATP of alcoholic fermentation (2/3 versus 2 ATP per glucose), the ecGEM predicted a growth decrease from 0.36 h^{-1} in the reference to 0.175 h^{-1} in the engineered strain. However, this <3 -fold decrease would require the specific glucose consumption rate to increase. Surprisingly, after the pathway swap the engineered strain immediately grew at 0.15 h^{-1} with a glucose consumption rate of $29 \text{ mmol (g CDW)}^{-1} \text{ h}^{-1}$, which was indeed higher than reference ($23 \text{ mmol (g CDW)}^{-1} \text{ h}^{-1}$) and one of the highest reported for *S. cerevisiae*. The accompanying 2,3-butanediol- ($15.8 \text{ mmol (g CDW)}^{-1} \text{ h}^{-1}$) and glycerol ($19.6 \text{ mmol (g CDW)}^{-1} \text{ h}^{-1}$) production rates were close to predicted values. Proteomics confirmed that this increased consumption rate was facilitated by enzyme reallocation from especially ribosomes (from 25.5 to 18.5 %) towards glycolysis (from 28.7 to 43.5 %). Subsequently, 200 generations of sequential transfer did not improve growth of the engineered strain, showing the use of ecGEMs in predicting opportunity space for laboratory evolution. The observations in this study illustrate both the current potential, as well as future improvements, of ecGEMs as a tool for both metabolic engineering and laboratory evolution.

1. Introduction

Genome-scale metabolic models are an important part of today's metabolic engineering toolbox (Duröt et al., 2009; Patil et al., 2004). To expand the predictive power of these models, thermodynamics, enzyme kinetics, signaling and transcriptomics can be integrated (Chung et al., 2021; Li et al., 2022; Niu et al., 2021; O'Brien et al., 2013; Oftadeh et al., 2021; Österberg et al., 2021; Salvy and Hatzimanikatis, 2020; Sánchez et al., 2017; Sinha et al., 2021; Yang et al., 2021). Enzyme constrained genome-scale models (ecGEMs) have now been constructed for several organisms to give mechanistic insight and guide metabolic engineering (Alter et al., 2021; Arend et al., 2022; Bujdos et al., 2023; Kerkhoven, 2022; Niu et al., 2022; Wu et al., 2023; Zhou et al., 2021). By using

turnover numbers for relevant metabolic enzymes and their molecular weights, ecGEMs can predict flux distribution, as well as the proteome allocation, providing a mechanistic explanation for fluxes (Sánchez et al., 2017). Recently, a deep learning algorithm to predict turnover numbers directly from protein sequences (without the need for *in vitro* data on each individual enzyme), and application of a Bayesian model to adjust turnover numbers from *in vitro* assays to conditions *in vivo*, were added (Li et al., 2022). This version of the yeast ecGEM has been shown to predict growth and metabolism across various carbon sources and growth rates, and has been able to predict behavior of a wide range of yeast species (Li et al., 2022). Given their relatively recent addition to the metabolic engineering toolbox, only a few articles have applied ecGEMs as part of their workflow (Ishchuk et al., 2022; Liu et al., 2019;

* Corresponding author.

E-mail address: tonvm@kth.se (A.J.A. van Maris).

Ye et al., 2020).

Anaerobic redox-neutral and ATP-yielding pathways from substrate to desired products provide the stoichiometrically highest possible yield. This not only holds for classical alcoholic and homolactic fermentation, but where thermodynamically possible also for metabolically engineered heterologous pathways (Cueto-Rojas et al., 2015). This provides a direct economic benefit by reducing substrate cost and decreases investment cost by enabling the use of simple(r) bioreactors without need for aeration, eliminates the need for air compression and decreases cooling requirements. Additionally, the simpler bioreactors are easier to clean and maintain aseptic. The coupling of product formation to ATP conservation enables use of laboratory evolution for improved growth and -product formation, and cell recycling during production can minimize biomass formation, whilst still allowing cellular maintenance. The coupling to growth and reliance on a high flux through central metabolism makes heterologous ATP-yielding, redox-neutral product formation pathways an interesting test case to evaluate ecGEMs and their ability to predict fluxes and enzyme allocation.

Heterologous production of (R,R)-2,3-butanediol (23BD) in yeast was chosen as target product due to industrial relevance and large body of prior knowledge, albeit under aerobic conditions (Kim et al., 2017; Ng et al., 2012). 23BD is an industrially relevant molecule with applications in solvents, flavoring, fuel (additives), printing inks, synthetic rubber, plastics and softeners (Celińska and Grajek, 2009; Garg and Jain, 1995). The fact that 23BD has previously been produced aerobically with *S. cerevisiae* to a high titer at a high rate (Kim et al., 2017), means that product export and toxicity, both aspects that are difficult to take into account in ecGEMs, will likely not influence the predictions. Starting from two molecules of pyruvate, 23BD can be produced through acetolactate synthase, acetolactate decarboxylase and 23BD dehydrogenase. Competing alcoholic fermentation can be eliminated by deletion of pyruvate decarboxylase encoding genes (*PDC1*, 5 and 6) and concomitant (over)expression of *MTH1ΔT* and *ACH1* (Chen et al., 2015; Oud et al., 2012). However, production of 23BD from glucose through glycolysis and the abovementioned pathway results in one surplus NADH. Under anaerobic conditions, yeast normally reoxidizes excess NADH through glycerol production (van Dijken and Scheffers, 1986). This implies that under anaerobic conditions, lacking other more efficient NADH reoxidizing mechanisms, the catabolic pathway will have to co-produce 23BD and glycerol at a one-to-one ratio.

The aim of this study was to investigate the predictive power of ecGEMs for fluxes, yields, as well as proteome distribution in yeast strains where the alcoholic fermentation was swapped for redox-neutral, ATP-yielding co-production of 23BD and glycerol. The recently published Bayesian GECKO model emodel *Saccharomyces cerevisiae*.Posterior_mean of *S. cerevisiae* (Li et al., 2022), was adjusted for comparison of model-based growth and product formation of the CRISPR-based engineered strain with *in vivo* anaerobic bioreactor batch cultures. Proteome analysis was performed to benchmark the ecGEM enzyme allocation predictions. Additionally, sequential transfer in serum tubes was used to investigate the opportunity space for laboratory evolution. Possible model improvements were investigated by constraining the ecGEM with the measured enzyme abundances.

2. Methods

2.1. Strains and media

The *S. cerevisiae* strains used in this study (Table 1) are derived from the CEN.PK lineage (Entian and Kötter, 2007). The strains were cultured in synthetic medium (Verduyn et al., 1992), with addition of suitable auxotrophic markers (Pronk, 2002), and a tween-ergosterol-ethanol solution in anaerobic conditions, as described by Verduyn et al. (1990). For growth without maintaining auxotrophic markers, YPD medium, containing 10 g L⁻¹ yeast extract, 20 g L⁻¹ peptone, and 20 g L⁻¹ glucose, was used. *Escherichia coli* strain DH5 α was used for plasmid

Table 1
Saccharomyces cerevisiae strains used in this study.

Strain name (accession no.)	Relevant genotype	Parental strain	Reference
CEN.PK113-7D			Entian and Kötter (2007)
IMX672 (Y40595)	MATA <i>ura3-52 trp1-289 leu2-3112 his3Δ can1Δ::cas9-natNT2</i>	CEN.PK2-1C	Mans et al. (2015)
GSY006	<i>proSH-mth1-pdc5+6 trp1-289 leu2-3112 his3Δ pdc6::pTEF1-ACH1; MTH1ΔT; pdc5Δ</i>	IMX672	This study
GSY008	<i>proSH-pdc1; ura3-52 trp1-289 leu2-3112 pdc6::pTEF1-ACH1; MTH1ΔT; pdc5Δ; pdc1::(alsD-tADH2 pTDH3-BDH1 pFBA1-budA)</i>	GSY006	This study
GSY013	<i>pTHUL</i>	IMX672	This study
GSY014	<i>pTHUL; pdc6::pTEF1-ACH1; MTH1ΔT; pdc5Δ; pdc1::(alsD-tADH2 pTDH3-BDH1 pFBA1-budA)</i>	GSY008	This study

production. It was cultured on LB medium containing 5 g L⁻¹ yeast extract, 10 g L⁻¹ peptone, and 10 g L⁻¹ NaCl. All strains were stored at -80 °C in their respective media, after addition of glycerol to a final concentration of a 25 % (w/w).

2.2. Plasmids and strain engineering

Starting from strain IMX672, integrations and deletions were performed as described by Mans et al. (2018), by transforming the parental strain with indicated plasmids and PCR fragments (Supplement S1). Colonies of the resultant strains were screened by PCR amplification of relevant sites (Supplement S1), before proceeding to the next round of strain construction. All plasmids (Table 2) were constructed by Gibson assembly (Gibson et al., 2009) from their indicated PCR fragments (primers and details in Supplement S1), followed by transformation to *E. coli* DH5 α . Selected colonies were screened by PCR and confirmed using Sanger sequencing of relevant regions (Eurofins genomics). The plasmids and PCR products were purified using the GeneJET Miniprep Kits (Thermo Fisher scientific). Colony PCR was performed using DreamTaq Green polymerase (Thermo Fisher scientific), while Phusion polymerase (Thermo Fisher scientific) was used for all strain construction.

2.3. Laboratory evolution

Triplicate 25 mL anaerobic tubes with 5 mL synthetic medium at pH 6.0 were inoculated with GSY014 glycerol stocks and incubated at 30 °C

Table 2
Plasmids used in this study.

Plasmid name	Relevant characteristics	Reference
pROS10	2 μ m ampR <i>URA3</i> gRNA-CAN1 gRNA-ADE2	Mans et al. (2015)
pROS16	2 μ m ampR <i>HIS3</i> gRNA-CAN1 gRNA-ADE2	Mans et al. (2015)
pBC414	<i>CEN6/ARS4</i> chlR <i>TRP1</i>	(Frazer and O'Keefe, 2007)
pROSU-mth1-pdc5+6	2 μ m ampR <i>URA3</i> gRNA- <i>MTH1</i> gRNA-[<i>PDC5</i> and <i>PDC6</i>]	This study
pROSH-pdc1	2 μ m ampR <i>HIS3</i> gRNA- <i>PDC1</i> [2 \times]	This study
pTHUL	<i>CEN6/ARS4</i> chlR <i>TRP1</i> <i>LEU2</i> <i>HIS3</i> <i>URA3</i>	This study
pYGS011	ampR <i>budA-tADH2</i>	This study
pYGS012	ampR <i>pTDH3-BDH1</i>	This study
pYGS013	ampR <i>pFBA1-alsS</i>	This study
pYGS017	ampR <i>pTEF1-ACH1</i>	This study

with 180 rpm shaking after purging with 20 cycles of alternating vacuum (VCP 130 pump, VWR) and nitrogen gas. Upon reaching stationary phase, cultures were reinoculated by a 50 times dilution in fresh tubes with purged medium. Glycerol stocks were stored at various time points throughout the experiment. The evolution experiment was terminated after 35 inoculation cycles (≈ 200 generations).

2.4. Batch cultivations in bioreactors

Batch cultivations were carried out in a multiparallel stainless steel bioreactor system with 6 reactors (GRETA, Belach Bioteknik, Skogås, Sweden) at 30 °C, in 800 mL of the synthetic medium described above. Anaerobic conditions were maintained by applying a 100–250 mL min⁻¹ nitrogen stream in the headspace and 500 rpm stirring. The pH was maintained at 5.0 by automatic titration with 2 M KOH. The outlet gas from each condenser was sampled on a 1313 Fermentation Monitor (LumaSense Technologies, Santa Clara, USA) to obtain the CO₂ concentrations. The cultures were sampled aseptically by withdrawing samples through rubber septa using a needle and syringe. Optical density at 600 nm (OD₆₀₀) and samples for HPLC analysis were withdrawn at each sample point, while cell dry weight (CDW) was sampled two times per batch. Samples for proteomics analysis were taken at the second to last sample point, while the cells were still growing exponentially, before the glucose was depleted. Inoculum for each bioreactor was prepared by first transferring the appropriate glycerol stock to a shake flask with synthetic medium at pH 6.0 and incubating it aerobically overnight at 30 °C with 180 rpm shaking (Minitron HT Infors, Bottmingen-Basel, Switzerland). The cells were then inoculated to multiple sealed 125 mL serum bottles containing 20 mL of synthetic medium at pH 6.0 and purged from oxygen by applying 20 cycles of alternating vacuum and nitrogen gas. The serum bottles were incubated until they reached an OD of 0.3–2.0. The bioreactors were inoculated to a target OD₆₀₀ of 0.05.

2.5. Cell mass analysis

OD₆₀₀ was measured with a spectrophotometer (Genesys 20, Thermo Fisher scientific) at 600 nm after dilutions to an OD₆₀₀ of 0.1–0.2 using a 9 g L⁻¹ NaCl solution. CDW was measured in triplicate by filtering 5 mL culture samples through pre-weighed 47 mm nitrocellulose filters (0.45 µm pore size, Cytiva, Marlborough, USA). The filters were washed with deionized water and then dried in a microwave oven (R-742, Sharp Corporation, Sakai City, Japan) at power setting 30 (≈ 300 W) for 20 min and weighed. A standard curve for converting OD₆₀₀ to CDW was established by linear regression to allow estimation of the CDW at all sample points.

2.6. Metabolite analysis

Samples for extracellular metabolite analysis were centrifuged at 20238g for 5 min (centrifuge 5424, Eppendorf). The supernatant was analyzed on an Alliance 2695 system (Waters, Milford, MA, USA) equipped with a 2414 refractive index detector (Waters), a 2996 photodiode array detector (PDA, Waters), and a TCM column heater (Waters). 10 µL samples were injected on an Aminex HPX-87H organic acid analytical column equipped with a guard column (Bio-Rad, Hercules, CA, USA) maintained at 60 °C and separated with a mobile phase of 5 mM H₂SO₄ at 0.6 mL min⁻¹. Glucose, glycerol, ethanol, 2,3-butanediol, and acetoin were all quantified on the refractive index detector, while trace amounts of organic acids were detected at 210 nm on the PDA. The organic acids contributed to less than 1 % of the degree of reduction- and carbon balance in GSY014 and were therefore excluded from modeling and presentation in the results section. They are included in the total degree of reduction- and carbon balances in Table 4. Degree of reduction was calculated as described by (Heijnen, 1994), a carbon content of 25 g CDW (mol C)⁻¹, and a degree of reduction of 4.2 (mol

C)⁻¹ was assumed for the biomass (Lange and Heijnen, 2001).

2.7. Proteome analysis

Samples for proteomics were withdrawn rapidly from the reactors using a syringe and several 2 mL samples were quickly added to pre-chilled 2 mL microcentrifuge tubes already placed in a pre-cooled centrifuge rotor (5424, Eppendorf). For samples of GSY013, 6 tubes were filled with 2 mL each (total 12 mL), and for samples of GSY014, 12 tubes were filled with 2 mL each (total 24 mL), to account for the lower CDW in the culture. The centrifuge was rapidly closed and run for 15 s before being stopped. As soon as it was possible to open the centrifuge, the supernatant was discarded from each tube by pouring, and any remaining drop of supernatant was wiped off on a paper napkin. Then, the tube was closed, and dropped into liquid nitrogen. The entire procedure took less than 3 min in total for each set of tubes. Once frozen, the samples were stored at -80 °C. Shipping of the samples from KTH Royal Institute of Technology to the Proteomics Core Facility at the University of Tartu, was carried out on dry ice.

2.8. Determination of protein content

Protein content was determined on the same frozen pellets that were also used for proteomics analysis with a Lowry method adapted from Verduyn et al. (1992). Briefly, frozen pellets were thawed and washed three times in 1 mL deionized water, followed by centrifugation for 5 min at 20238g (centrifuge 5424, Eppendorf) and withdrawal of the supernatant. The pellet was dissolved to 1.0 mL with deionized water, and 0.5 mL of 3M NaOH was added. A bovine serum albumin (BSA) standard curve (0.0–5.0 mg mL⁻¹) was prepared by addition of 0.5 mL 3M NaOH to 1 mL of each standard solutions. Both standards and samples were incubated for 10 min at 100 °C in a block heater and then immediately cooled on ice. 300 µL of 0.1 M CuSO₄ was added to each tube and incubated at room temperature for 5 min. After centrifuging at 20238g for 5 min, the supernatants were analyzed on a spectrophotometer (Cary 50 bio, Agilent) at 510 nm and the sample concentrations were estimated from a linear fit of the BSA standard curve.

2.9. Sample preparation for absolute proteomics

Cell pellets were suspended with 10 vol (relative to pellet volume) of chaotrope-based lysis buffer (6 M guanidine-HCl, 100 mM Tris-HCl pH 8.0, 50 mM dithiothreitol), heated at 95 °C for 10 min and sonicated with a Bioruptor (Diagenode) sonication device for 15 cycles (30 s ON, 60 s OFF; "High" setting) at 4 °C. Next, 0.5 vol (relative to lysate volume) of Silibeads TypZY-s 0.4–0.6 mm (Sigmund Lindner) beads were added to the lysate and bead beating was carried out with FastPrep24 (MP Biomedicals) device (2 × 40 s at 6 m s⁻¹). Samples were centrifuged at 17000g for 5 min and the supernatant was separated from the beads and transferred to a new tube. Small aliquots of the lysate were then precipitated with trichloroacetic acid-deoxycholate (TCA-DOC) precipitation and protein concentration determined with the Micro BCA assay (Thermo Fisher Scientific). Next, 20 µg of lysate protein was alkylated with 100 mM chloroacetamide by incubating for 1 h in the dark at room temperature. Samples were then processed to peptides by the SP3 protocol essentially as described elsewhere (Hughes et al., 2019).

2.10. LC/MS/MS analysis

1 µg of peptides were injected to an Ultimate 3500 RSLC nano system (Dionex) using a 0.3 × 5 mm trap-column (5 µm C18 particles, Dionex) and an in-house packed (3 µm C18 particles, Dr Maisch) analytical 50 cm × 75 µm emitter-column (New Objective). Columns were operated at 45 °C. Peptides were eluted at 250 nL min⁻¹ with an 8–40 % B 60 min gradient (buffer B: 80 % acetonitrile + 0.1 % formic acid, buffer A: 0.1 % formic acid) to a Q Exactive HF (Thermo Fisher Scientific) mass

spectrometer (MS) using a nano-electrospray source (spray voltage of 2.5 kV in positive mode).

The MS was operated with data-independent acquisition (DIA) method using variable isolation windows over a range of 400–1200 m/z. Briefly, one full range 400–1200 m/z MS1 scan was collected at a resolution setting of 60 000 (max ion injection time of 60 ms, max of 3e6 ions), followed by 25 overlapping (overlap of 1 m/z) variable size DIA windows. Each DIA scan was collected at a resolution setting of 30 000 (max ion injection time of 41 ms, max of 3e6 ions), default charge state was set to +3 and normalized collision energy to 27.

2.11. Raw DIA MS-data analysis

Raw data was processed with the DIA-NN (Demichev et al., 2020) software (v1.8) using library-free search. FASTA files used were UniProt (www.uniprot.org) *Saccharomyces cerevisiae* S288C reference proteome supplemented with *Bacillus subtilis* alsS and *Klebsiella aerogenes* budA sequences. Precursor m/z range was set to 400–1200, precursor intra-sample q-values and their associated intrasample protein group q-values were set to ≤ 0.01 , all other settings were default.

2.12. Bioinformatics

Absolute protein concentrations (Supplement S3) were derived from normalized sum of MS intensities using the total protein approach (Sánchez et al., 2021), which assumes its proportionality to the measured total protein content. Triplicate experiments were used in differential expression analysis. The Benjamini-Hochberg (Benjamini and Hochberg, 1995) method was used to adjust the p-values for multiple comparisons. Additional data analysis included PC analysis carried out using ClustVis tool (Metsalu and Vilo, 2015). Gene set analysis was performed with proteome-normalized proteomics data ($\mu\text{g} (\text{g protein})^{-1}$) using Piano (Väremo et al., 2013). All other settings were default, except gene sets smaller than 5 and larger than 300 genes were ignored.

2.13. Enzyme constrained genome-scale model

The recently published *S. cerevisiae* ecGEM emodel_Saccharomyces_cerevisiae_Posterior_mean fit to experimental data using a Bayesian method (Li et al., 2022) was chosen as a starting point for the modeling of GSY014, because it has been shown to be highly constrained and predictive of *S. cerevisiae* metabolism across various conditions. Model analysis and alteration was carried out with the toolboxes RAVEN 2.0 and COBRA 3.0 (Heirendt et al., 2019; Wang et al., 2018). All modeling was performed with unconstrained substrate uptake rate. Instead, the growth rate was constrained by the total protein abundance.

First, the model was adapted to anaerobic growth conditions by setting the bounds on the oxygen exchange reaction (r_1992) to zero and introducing exchange reactions for the metabolites pantethenate (s_0032), nicotinate (s_0032), ergosterol (r_1757), ergosta-5,7,22,24 (28)-tetraen-3beta-ol (r_2137), oleate (r_2189) and palmitoleate (r_1994). The heme requirement was removed from the biomass reaction, since it cannot be synthesized anaerobically, and it is not required in the absence of oxygen. The non-growth associated maintenance reaction (r_4046) was set to 1 mmol (g CDW) $^{-1}$ h $^{-1}$ based on previously published data on retentostat cultures of *S. cerevisiae* (Boender et al., 2009). It was not necessary to update the growth associated maintenance value of 55.2 mmol (g CDW) $^{-1}$ to match the phenotype of GSY013.

Second, the model was fit to the experimentally measured extracellular fluxes of GSY013 by iteratively altering the σ factor (average *in vivo* saturation of all enzymes) until the predicted growth rate was within 0.01 h $^{-1}$ of the observed average growth rate. The value for f (the mass fraction of enzymes in the model) was assumed to be 0.5. After adaption, a total protein pool pseudoreaction (EX_protein_pool) of 123 mg (g

CDW) $^{-1}$ was obtained that fit the observed growth rate of 0.36 h $^{-1}$. A reaction converting NADH to NAD $^{+}$ at 3 mmol (g CDW) $^{-1}$ was coupled to the growth reaction to give the correct ratio of glycerol production to glucose consumption, and to align the degree of reduction (as defined by (Heijnen, 1994)) of the modeled biomass to a published value for degree of reduction of *S. cerevisiae* biomass at 4.2 C-mol $^{-1}$ (Lange and Heijnen, 2001).

Third, the ecGEM was adapted to the genotype of GSY014 by introducing both metabolite and protein reactions for alsS and budA with k_{cat} values of 121 s $^{-1}$ (Atsumi et al., 2009) and 27 s $^{-1}$ (Choi et al., 2016) respectively. Pyruvate decarboxylase reactions (r_0959_*, r_0960_*, r_4651_*) were deleted. Cytoplasmic acetyl-CoA supply through acetate transport from the mitochondria as proposed by (Chen et al., 2015) was implemented by providing acetyl-CoA hydrolyzing activity of Ach1 and introducing acetate transport across the mitochondrial membrane by Ady2. Both reactions used the k_{cat} values already present for those proteins. A competing acetate transport mechanism through ethyl acetate (r_4650) was blocked, since it was deemed less feasible based on the evidence provided by (Chen et al., 2015). Solving for the maximum growth rate in this model yielded the predicted data for GSY014.

The total-protein-constrained models of GSY013 and GSY014 were further constrained by measured extracellular fluxes and proteomics data through the GECKO pipeline, essentially as described by Sánchez et al. (2017). To introduce this model constraint, individual protein pseudoreactions were still allowed to pull flux from the total protein pool, but now with an upper bound per enzyme as measured by proteomics. If required, the total protein pool pseudoreaction was adjusted to match the measured extracellular fluxes. Also individual protein constraints that limited the growth rate from reaching the observed value were flexibilized by adjusting the upper bound of those enzymes as initially described for GECKO models (Sánchez et al., 2017). Both the original total-protein-constrained models and the proteomics-constrained models were subjected to flux variability analysis and sampling as described for GECKO models (Sánchez et al., 2017).

3. Results

An (R,R)-2,3-butanediol-glycerol co-producing strain (GSY014) was constructed by overexpression of a 23BD production pathway constituting the native *S. cerevisiae* 2,3-butanediol dehydrogenase Bdh1, acetolactate decarboxylase budA from *Klebsiella aerogenes* and acetolactate synthase alsS from *Bacillus subtilis* (Choi et al., 2016). To eliminate pathways that compete for NADH, all genes encoding functional pyruvate decarboxylase alleles (PDC1, PDC5 and PDC6) were deleted, while the *MTH1ΔT* allele was introduced to alleviate glucose sensitivity caused by Pdc deficiency (Oud et al., 2012), and the native *ACH1* gene was overexpressed to alleviate expected cytosolic acetyl-CoA deficiency (Chen et al., 2015). A theoretical stoichiometric analysis of the anaerobic metabolism of the resulting strain predicts conversion of glucose to equimolar amounts of 2,3-butanediol and glycerol for redox-neutral ATP generation, whilst the engineered catabolic pathway in GSY014 has a theoretical ATP yield per glucose of 2/3 mol mol $^{-1}$, compared to 2 mol mol $^{-1}$ for the ethanol producing reference strain (GSY013). It should be noted that, to reoxidize the excess NADH from biosynthesis (van Dijken and Scheffers, 1986), for both strains (additional) conversion of some glucose to glycerol is expected.

3.1. Adaptation of enzyme constrained models and boundary conditions

To predict growth and metabolism of GSY014, the recently developed Bayesian GECKO model of *S. cerevisiae* was chosen as a starting point. It was chosen since it is highly constrained to experimental data of *S. cerevisiae* across a range of conditions and has been shown to predict changes in growth rate based on those conditions (Li et al., 2022). The ecGEM was adapted for anaerobic growth and the genotype of GSY014

was introduced (see Methods for specifics). To determine the upper bound of the protein pool pseudoreaction, which determines the growth rate of the model in batch conditions, a batch culture of the reference strain GSY013 was performed (Fig. 1A and Table 3). The experimentally determined growth rate and exchange rates were not significantly different from previously published values for similar reference strains (Papapetridis *et al.*, 2018), the degree of reduction and carbon balances closed reasonably well (Table 4), and the total protein content was experimentally determined to be 450 mg (g CDW)⁻¹. To match the experimentally determined growth rate, the protein pool pseudoreaction bound was set to 123 mg (g CDW)⁻¹ and the degree of reduction of biomass in the GSY013 ecGEM was reduced by 3 mmol (g CDW)⁻¹ to match the model fluxes to the experimentally determined values.

The same changes to the boundary conditions were applied to the ecGEM of the engineered GSY014 strain to obtain a prediction of the growth rate and metabolism of this strain (Tables 3 and 4). In line with the back-of-the-envelope calculated stoichiometry, the ecGEM predicted rates of 23BD (16.8 mmol (g CDW)⁻¹ h⁻¹) and glycerol (19.0 mmol (g CDW)⁻¹ h⁻¹) production were close to equal, with the latter being slightly higher (Table 3). As indicated above, the ATP yield of 23BD/glycerol production is only one-third of that of alcoholic fermentation, which resulted in a similar predicted 3-fold reduction in the biomass yield on glucose by the ecGEM for GSY014 (Table 4; bottom row). Interestingly, this yield decrease did not translate in a 3-fold reduction in the predicted specific growth rate (0.175 h⁻¹ versus 0.36 h⁻¹), as the enzyme constrained model predicted an increase of the modeled biomass specific glucose uptake rate from 23.0 mmol (g CDW)⁻¹ h⁻¹ in the reference strain to 27.8 mmol (g CDW)⁻¹ h⁻¹ in the engineered strain (Table 3).

3.2. In vivo evaluation of predicted rates and yields

To investigate growth and metabolism of the constructed 23BD and glycerol co-producing strain and to evaluate the model predictions, GSY014 was studied in anaerobic batch cultures (Fig. 1B). Degree of reduction and carbon balances closed well for both strains. Ethanol formation was successfully eliminated and replaced as catabolic pathway by near equimolar formation of 23BD and glycerol, close to model predictions (Tables 3 and 4). The biomass yield on glucose for GSY014 was about a third compared to the reference strain, while the glycerol yield increased 3.5-fold, in line with the model prediction (Table 4). Interestingly, although the observed maximum specific growth rate was slightly below the prediction, the measured specific

glucose uptake rate was about 25 % higher than the reference strain and even a bit higher than predicted. The only flux not predicted by the model was a small excretion of acetoin, which in turn could result in a slight increase in glycerol formation compared to predictions. No significant formation of other metabolites was observed.

3.3. Laboratory evolution

Since the observed maximum specific growth rate of GSY014 was close to, but slightly lower than predicted by the ecGEM, it was investigated if laboratory evolution could be used to improve the strain fitness. To place selective pressure on growth rate improvements, repeated batch cultures in anaerobic serum tubes were done for 35 transfers totaling approximately 200 generations. After 200 generations, no significant differences were observed for the evolved culture GSY014E compared to starting strain GSY014 on any of the measured rates, except for a slight increase in cell specific acetoin production rate (Table 3). There were no significant yield differences, not even in the acetoin yield per glucose (Table 4).

3.4. Proteome data analysis

To compare *in vivo* changes in protein concentrations with the ecGEM predicted changes in proteome allocation, the proteome of the batch cultivations was analyzed. Absolute proteome abundances were obtained with mass spectrometry-based data-independent acquisition (Demichev *et al.*, 2020) and total protein amount (Sánchez *et al.*, 2021) quantification methods, based on a measured protein content of 0.45 g (g CDW)⁻¹ for both strains. A total of 3797 absolute abundances were determined. The results in principal component analysis showed most of variance (43.2 %) in principal component 1, allowing a separation between the engineered and the control strain (Fig. 2B).

Proteome-normalized data (μg (g protein)⁻¹, Supplement S3) were used to analyze abundance changes in each strain. The differential expression analysis showed significant changes for 172 proteins, or 4.5 % of total proteins, based on a cut-off of a 1.5-fold change and Benjamini-Hochberg corrected P-value (P) < 0.05 (Fig. 2A and Supplement S3).

Subsequently, the proteomic results were analyzed based on protein functional groups representing different metabolic pathways and cellular functions. Functional groups were obtained from the proteomap database (Liebermeister *et al.*, 2014). In line with the decreased growth rate and increased glucose consumption rate in GSY014, the ribosome abundance decreased from 25.5 % in the reference strain to 18.5 % in

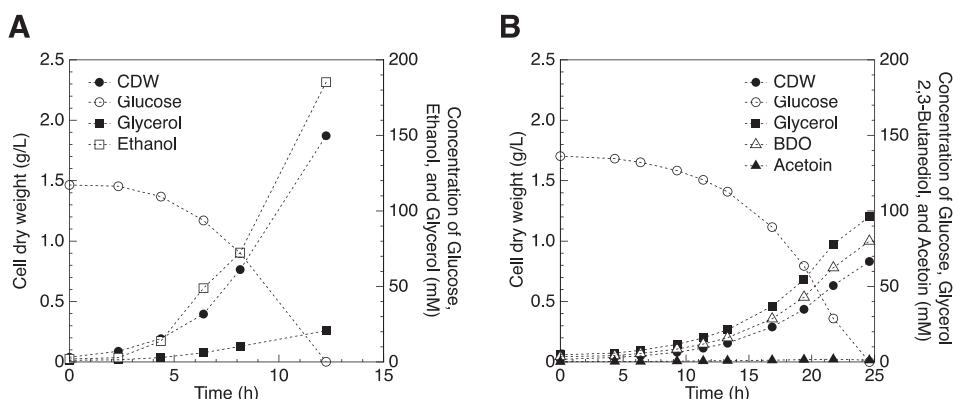


Fig. 1. Representative profiles of sextuplicate anaerobic batch cultivations of reference strain GSY013 (A) and engineered 23BD-glycerol co-producing strain GSY014 (B) in stirred tank reactors at 30 °C, pH 5.0. Symbols: Filled circles, Cell dry weight (CDW); open circles, glucose; open squares, ethanol; filled squares, glycerol; open triangles, 2,3-butanediol (23BD); filled triangles, acetoin. Data for all cultures is available in Supplement S2.

Table 3

Biomass specific rates observed in anaerobic batch cultivations in bioreactors.

Strain	μ (h ⁻¹)	Rate of consumption or production (mmol (g CDW) ⁻¹ h ⁻¹)					
		Glucose	Glycerol	Ethanol	23BD ^a	Acetoin	CO ₂
GSY013 ^b	0.36 ± 0.02	-23 ± 2	4.5 ± 0.4	31 ± 2	ND ^d	ND ^d	38 ± 10
Predicted GSY014	0.175	-27.8	19.0	0	16.8	0	34.9
GSY014 ^b	0.15 ± 0.01	-29 ± 1	19.6 ± 0.5	ND ^d	15.8 ± 0.4	0.39 ± 0.03	36 ± 2
GSY014E ^c	0.16 ± 0.01	-27 ± 1	18.6 ± 0.8	ND ^d	14.8 ± 0.7	0.48 ± 0.01	33 ± 9

^a 2,3-butanediol.^b Average of 6 independent cultivations ± standard deviation.^c Average of 2 independent cultivations ± mean deviation.^d Not detected.**Table 4**

Yields and carbon- and degree of reduction balances observed in anaerobic batch cultivations in bioreactors.

Strain	Yields on glucose (mg (g glucose) ⁻¹)						Recovery of consumed substrate, based on:	
	CDW	Glycerol	Ethanol	23BD ^a	Acetoin	CO ₂		
GSY013 ^c	89 ± 6	101 ± 2	340 ± 1	ND ^e	ND ^e	380 ± 100	93 ± 7 %	91 ± 1 %
Predicted GSY014	35	349	0	302	0	307	NA ^f	NA ^f
GSY014 ^c	30 ± 2	346 ± 1	ND ^e	274 ± 1	7 ± 1	300 ± 10	95 ± 1 %	94.5 ± 0.1 %
GSY014E ^d	33 ± 1	350 ± 3	ND ^e	273 ± 1	9 ± 1	300 ± 90	97 ± 7 %	95.7 ± 0.4 %

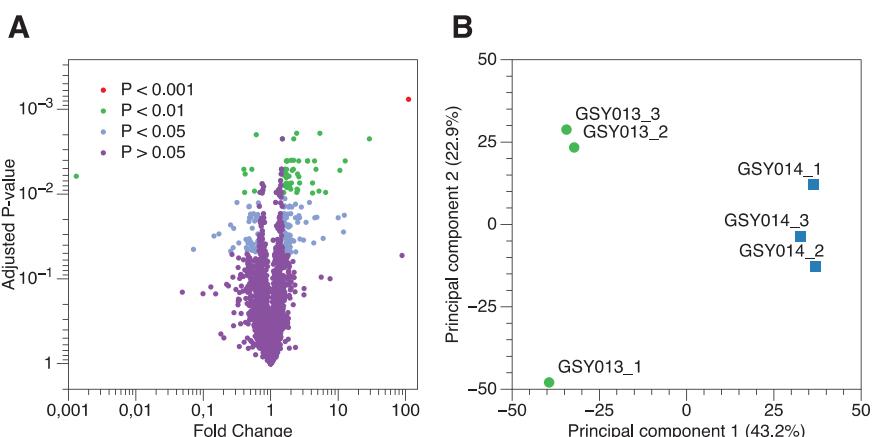
^a 2,3-butanediol.^b Degree of reduction as defined by (Heijnen, 1994).^c Average of 6 independent cultivations ± standard deviation.^d Average of 2 independent cultivations ± mean deviation.^e Not detected.^f Not applicable.

Fig. 2. Comparison of absolute protein abundances (μ (g protein)⁻¹) in anaerobic batch cultures in bioreactors of 2,3-butanediol-glycerol co-producing GSY014 and the ethanol producing reference strain GSY013. (A) Volcano plot of fold changes and false discovery rate (as indicated by the adjusted P-value) of absolute protein abundances in GSY014 compared to GSY013. The plot shows the magnitude of change (measured GSY014 value over GSY013) on the x-axis and the statistical significance on the y-axis. Proteins that are significantly differentially expressed in the engineered strain compared to control strain are shown in blue ($P \leq 0.05$), green ($P \leq 0.01$) or red ($P \leq 0.001$). The P-value was adjusted for multiple comparisons ($n = 3797$) using Benjamini and Hochberg (1995) method. The data used to generate the volcano plot is available in Supplement S3. (B) A principal component analysis plot indicating the distribution of protein abundances in two-dimensional space based on the first two principal components. Each dot represents protein abundances of a biological replicate, and replicates from the same strain are shown in the same color and shape.

the engineered strain (Fig. 3A), whilst glycolysis formed the largest protein group in the GSY014 with an increase from 28.7 % to 43.5 % of the total measured proteins (Fig. 3B). Other statistically significant differences in proteome allocation, including their directionality, were identified through gene set analysis ($P < 0.05$) (Väremo et al., 2013). In addition to glycolysis, expression levels of proteins involved in biosynthesis of cofactors, amino acid and lipid metabolism were significantly

upregulated in the engineered strain ($P < 0.05$) (Supplement S4). While in addition to proteins involved in ribosome and transcription, also proteins involved in cell wall- and glycogen metabolism, were significantly downregulated.

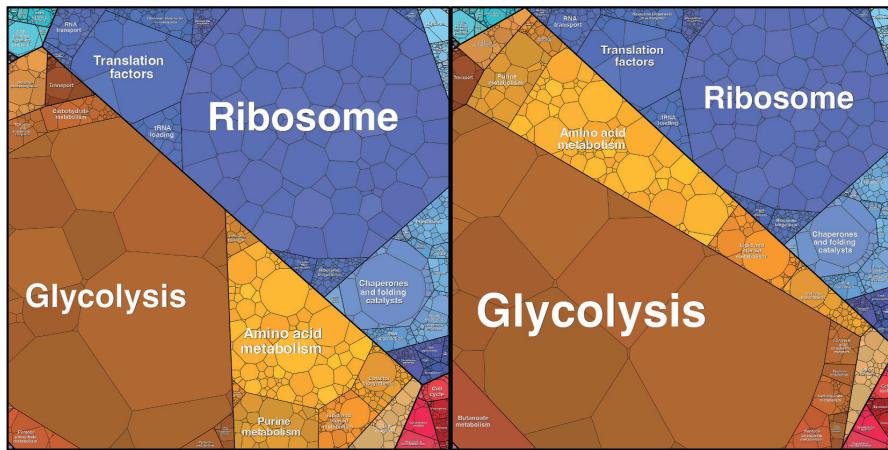


Fig. 3. Proteomaps (Liebermeister et al., 2014) of the reference strain GSY013 (left) and the 2,3-butanediol-glycerol co-producing strain GSY014 (right), showing the relative abundances of major protein groups of the respective strains.

3.5. Comparison of proteome data to ecGEM protein allocation predictions

To gain insight into how redistribution of protein allocation enabled the observed *in vivo* flux changes, a comparison was made between the total protein constrained ecGEM enzyme allocation prediction, where the algorithm was allowed to choose distribution of (iso)enzymes among metabolic pathways, and the measured proteomics (Fig. 4). The total protein constrained ecGEM prediction indicated a 30 % increase of proteins in upper glycolysis, while the lower part was reduced by 5 %. Although proteomics data showed that upper glycolysis was indeed upregulated by 30–70 %, the same upregulation was also seen for lower glycolysis (Fig. 4). This suggests a co-regulation of the expression of both parts of glycolysis. Another difference between proteomics and model is that the ecGEM only selects the most efficient isoenzyme for each catalytic step, in terms of k_{cat}/M_w . This is illustrated by model selection of Err1 instead of the catalytically active enolases. Presumably, this is because Err1 is annotated as having enolase activity with the same k_{cat} , but with a lower molecular weight, than the main enzymes Eno1 and Eno2.

Whereas both total protein constrained ecGEM and proteomics showed upregulation of the glycerol 3-phosphate dehydrogenases and -phosphatases, the increase in the proteomics data was lower than predicted. This likely reflects an overcapacity of the glycerol producing proteins in the reference strain to be able to deal with disturbances in redox metabolism. Below pyruvate, aside from the expected absence of pyruvate decarboxylase and the introduction of the heterologous 23BD-pathway genes, a strong upregulation of acetaldehyde- and alcohol dehydrogenases was seen in the proteomics. The fact that these proteins were predicted to be absent by the ecGEM, illustrates the difference between the modeling environment and *in vivo* reality. Outside these changes in central metabolism, the ecGEM also predicted that the expression of biosynthetic genes should be reduced in proportion to the roughly 50 % reduction in growth rate. This decrease was however not observed, and in contrast biosynthetic-protein functional groups, excluding translation, increased from 10.7 % to 13.8 % of the total protein pool.

3.6. Refining model predictions by introduction of individual protein constraints

All ecGEM predictions generated above were constrained by the total

amount of enzyme. To further compare the combined results of the ecGEM and the proteomics data, the measured individual protein abundances and extracellular fluxes were used to constrain the ecGEM. In the fitting process, the abundance of 46 proteins, constituting a total of 50 mg (g CDW)⁻¹, had to be flexibilized to obtain solutions for the proteomics-constrained ecGEM of GSY014. Among these proteins were enzymes in glycolysis, some required for biosynthesis and the heterologous BudA and AlsS (Supplement S6). For the proteomics-constrained ecGEM of reference strain GSY013, 72 protein abundances had to be flexibilized in the fitting process (Supplement S6), which totals to 54 mg (g CDW)⁻¹. These GSY013 proteins were mainly biosynthetic enzymes with annotated low k_{cat} values. This requirement for increased flexibility means that the measured protein concentration was too low to explain the observed flux, given the assumed k_{cat} and molecular weight in the model.

The addition of individual protein abundance constraints to the ecGEMs reduced the flux variability of the GSY014 model (Supplement S7, Fig. S9A), and the total protein pool pseudoreaction was adjusted to 382 mg (g CDW)⁻¹. In total, 750 measured enzyme abundances were applied as upper bounds to reactions in the model. This accounted for 352 mg (g CDW)⁻¹, or 92 % of the protein constraint in the model and 72 % of the total measured protein mass. The proteomics-constraints on the ecGEM of GSY013 had similar effects on the distribution of the flux variability of the model (Supplement S7, Fig. S9B), and on the distribution of sampled fluxes (Supplement S8).

To assess the correlation between the primary total-protein-constrained model and the refined proteomics-constrained ecGEM, sampled mean deviations were compared (Fig. 5A, Supplement S8). Overall, there is a large spread in mean fluxes, both through pseudoreactions related to enzyme supply (red) and metabolic reactions (black). This was mainly explained by decreased fluxes in the most optimal (as indicated by k_{cat}/M_w) enzymes and increased flux through less efficient isoenzymes and alternate pathways. This was especially evident for the fluxes through central metabolism (Fig. 5B), as illustrated by the switch between isoenzymes Pyk1 and Pyk2. The increased diversification in enzyme usage was also reflected by a decrease in the number of non-zero fluxes by 11.2 % in the proteomics-constrained ecGEM. Additionally, the measured extracellular fluxes were slightly lower than predicted by the total-protein-constrained model.

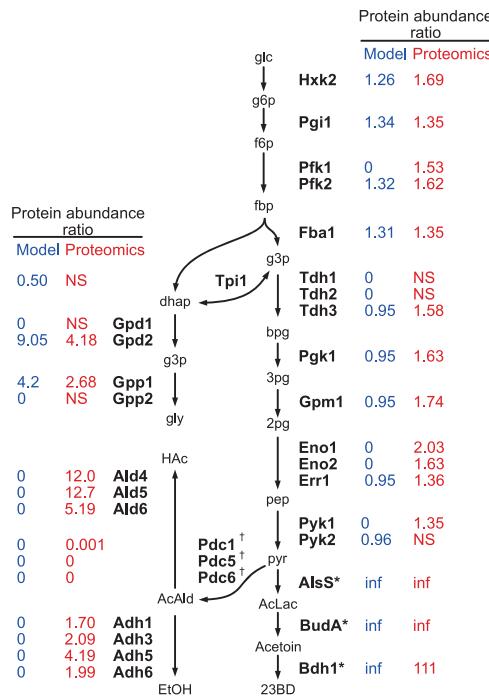


Fig. 4. Ratio of protein abundance in central metabolism between engineered 23BD-glycerol coproducing strain and the reference strain. The left column (blue) indicates the ecGEM predicted ratio of the indicated pseudoreaction related to enzyme supply for the engineered strain over the reference strain (Supplement S8). The right column (red) indicates the ratio as measured by proteomics at the 0.05 significance level ($n = 3$ for each strain, Supplement S5). Abbreviations: glc, glucose; g6p, glucose-6-phosphate; f6p, fructose-6-phosphate; fbp, fructose-1,6-bisphosphate; g3p, glyceraldehyde-3-phosphate; bpg, 2,3-bisphosphoglycerate; 3 pg, 3-phosphoglycerate; 2 pg, 2-phosphoglycerate; pep, phosphoenolpyruvate; pyr, pyruvate; AcLac, acetolactate; 23BD, 2,3-butanediol; dhap, dihydroxyacetone phosphate; g3p, glycerol-3-phosphate; gly, glycerol; HAc, acetate; AcAld, acetaldehyde; EtOH, ethanol. NS, no significant change; inf, infinite due to 0 flux in reference strain; * overexpressed genes; † deletion.

4. Discussion

The metabolic engineering approach in this study shows the power of thermodynamically feasible, redox neutral and ATP yielding pathways for anaerobic product formation (Cueto-Rojas et al., 2015). Although benefitting from previously proven pathway genes, apparently easy and ATP neutral product export and relatively low product toxicity (Choi et al., 2016; Mizobata et al., 2021), solely the introduction of the 23BD pathway and elimination of pathways competing for NADH, directly resulted in coupled anaerobic 23BD and glycerol formation with growth rates and specific production rates close to model predictions by the ecGEM. Importantly, this study also provides insights into the limits of thermodynamic potential and ATP yield on glucose required for such anaerobic pathways to be successful. The increase in the biochemical Gibbs free energy change (ΔG_r°) from $-215 \text{ kJ} (\text{mol glucose})^{-1}$ to $-147 \text{ kJ} (\text{mol glucose})^{-1}$, and the accompanying larger decrease in ATP yield from $2 \text{ mol ATP} (\text{mol glucose})^{-1}$ to $2/3 \text{ mol ATP} (\text{mol glucose})^{-1}$, were apparently not a hurdle for implementation of the engineered pathway. To the contrary, higher substrate consumption rates, lower biomass yield and high product formation rates were observed. The economic

potential of anaerobic production pathways is of course higher if such a catabolic pathway has only one product, rather than equimolar co-products. That scenario increases the concentration of the sole product, decreases downstream processing investment cost from two to one product and avoids that the economic success of a factory relies on beneficial market conditions for two products.

The fact that these large changes in flux distribution did occur in GSY014, without a need to apply laboratory evolution, demonstrated a remarkable flexibility of the central metabolism of *S. cerevisiae*. The ecGEM-predicted increase, which would not have been possible with classical genome-scale models, and *in vivo* confirmed increase in glycolytic flux, was realized by a large increase in the enzyme allocation to glycolytic proteins to 43.5 % of total measured proteins. However, where the ecGEM only predicted increases of the key enzymes, the proteome data indicated a general and coordinated upregulation of glycolysis. This suggests that global regulators or transcription factors underlie a coordinated upregulation of the glycolytic enzymes. Indeed, several transcription factors, such as Gat1, Mig1, Swi1 and Whi5 were down regulated (Coffman et al., 1995; Jorgensen and Tyers, 2004; Kayikci and Nielsen, 2015; Sudarsanam et al., 2000). It should be noted that, aside from a need for a higher glycolytic flux to optimize the biomass specific ATP production rate, the engineered strain also contains the *Mth1ΔT* allele (Oud et al., 2012). An unexpected observation in the proteomics data was an increase of aldehyde- and alcohol dehydrogenases, despite the complete absence of pyruvate decarboxylases. This upregulation could be a response to changed NADH/NAD⁺ ratios upon changing from alcohol-to-glycerol/2,3-butanediol dehydrogenases as NADH reoxidizing enzymes, or alternatively a response to intracellular levels of acetoin. Alternatively, the downregulation of Swi1, which is known to repress expression of *Adh1* and *Adh2* (Peterson and Herskowitz, 1992; Taguchi and Young, 1987), might have had this upregulation as a side effect. Although 200 generations of laboratory evolution did not result in improved phenotypes in bioreactor characterization, this upregulation of unneeded protein, and the excretion of acetoin, suggests that there is room for (evolutionary) improvement, albeit potentially requiring much longer time spans.

In the current generation ecGEMs, the biomass composition is assumed constant across all growth rates, which is a known simplification of the *in vivo* reality (Lange and Heijnen, 2001). A possible improvement would be to enable ecGEMs to automatically adjust biomass composition even in the absence of experimental data to, for instance, accommodate the fact that ribosome numbers, and thereby proteome allocation towards ribosomes, varies with growth rate (Björkeroth et al., 2020; Scott and Hwa, 2011). Care should however be taken to ensure that the benefits of introduction of variable biomass compositions (Elseman et al., 2022; O'Brien et al., 2013; Pramanik and Keasling, 1997), are not out-weighed by the need for demanding optimization algorithm beyond linear programming. This study showed that the maximum allowed protein abundance had to be flexibilized for a surprising number of proteins to enable proteomics-constrained ecGEM solutions. Assuming that the underlying deviation is in the model and not in the proteomic analysis, this indicates that the Bayesian procedure used in the creation of the model (Li et al., 2022) underestimated k_{cat} of the involved enzymes. Indeed, inspection of key glycolytic enzymes that needed flexibilization revealed that the k_{cat} values were much lower than previously published (Nilsson and Nielsen, 2016), and that flexibilization would not have been required if the higher literature values had been used in the model (Supplement S6). Interestingly, also the manually added k_{cat} values for the heterologously expressed enzymes AlsS and BudA required flexibilization, indicating they are more efficient than the current available *in vitro* data suggests. This makes efforts to improve correlation between *in vivo* and *in vitro* data (García-Contreras et al., 2012; Li et al., 2022), and manual curation of the model, as was done for the original GECKO model ecYeast7 (Sanchez et al., 2017), very beneficial for future ecGEM applications.

This study clearly showed the potential of ecGEMs in both metabolic

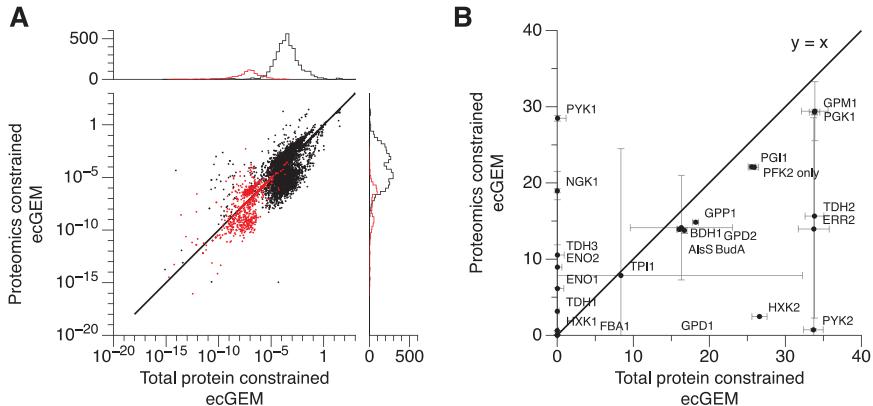


Fig. 5. Correlation between sampled mean flux in the proteomics-constrained ecGEM compared to the original total protein constrained ecGEM. A) A comparison of all the reactions in the model. On the outside axes, histograms of the corresponding fluxes are displayed as counts. Red denotes protein pseudoreactions while black indicates metabolic reactions. B) A comparison of the reactions displayed in Fig. 4, including isoenzymes. Names of the enzymes contributing to each flux are indicated, and the error bars are for a single standard deviation. All points are based on 10 000 random samples of the respective models.

engineering and laboratory evolution. Future improvements, as discussed above, will only increase this potential further. In metabolic engineering studies, ecGEMs can set goals for further pathway improvements, identify needs for improved information on important pathway enzymes, inform design of more efficient enzymes and/or identify if other factors, such as product export and toxicity, are influencing the outcome. For laboratory evolution, ecGEMs can help identify the opportunity space and indicate when (further) phenotype improvements are unlikely or will take a disproportionate amount of time. The use and further improvement of ecGEMs in modeling various organisms can result in faster design-build-test cycles, shorter time-to-market and improve efficacy of experiments.

Author contributions

GS and AJAvM conceived the study. GS and MF constructed the strains. MF performed the laboratory evolution. GS designed, performed, and analyzed the cultivation experiments. GS and AR constructed and analyzed the ecGEM. AR and PJL performed the bioinformatics analysis. The manuscript was written by GS, AJAvM and AR. All authors read and approved the final version of the manuscript. (CRediT author statement).

GS: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing- Original Draft, Visualization. AR: Methodology, Software, Formal analysis, Writing- Original Draft. MF: Investigation, Formal analysis, Writing - Review & Editing. PJL: Methodology, Writing - Review & Editing, Supervision. AJAvM: Conceptualization, Writing- Original Draft, Supervision.

Data availability

All research data is available in the supplementary materials

Acknowledgements

The authors acknowledge Robert Mans for advice on CRISPR-based genetic engineering and Sergo Kasvandik for his assistance with the proteomics analysis. The work of GS, MF and AVM was supported by grant 2019-05391 from the Swedish Research Council. The work of AR and PJL was supported by grant PRG1101 from the Estonian Research Council.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymben.2024.01.007>.

References

- Alter, T.B., Blank, L.M., Ebert, B.E., 2021. Proteome regulation patterns determine *Escherichia coli* wild-type and mutant phenotypes. *mSystems* 6, e00625-20. <https://doi.org/10.1128/mSystems.00625-20>.
- Arned, M., Zimmer, D., Xu, R., Sommer, F., Mühlhaus, T., Nikoloski, Z., 2022. Proteomes and constraint-based modeling reveal enzyme kinetic properties of *Chlamydomonas reinhardtii* on a genome scale. Preprint. *bioRxiv*. <https://doi.org/10.1101/2022.11.06.515318>.
- Atsumi, S., Li, Z., Liao, J.C., 2009. Acetolactate synthase from *Bacillus subtilis* serves as a 2-ketoisovalerate decarboxylase for isobutanol biosynthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* 75, 6306–6311. <https://doi.org/10.1128/AEM.01160-09>.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Björkeroth, J., Campbell, K., Malina, C., Yu, R., Di Bartolomeo, F., Nielsen, J., 2020. Proteome reallocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. *Proc. Natl. Acad. Sci. U.S.A.* 117, 21804–21812. <https://doi.org/10.1073/pnas.1921890117>.
- Boender, L.G.M., de Hulster, E.A.F., van Maris, A.J.A., Daran-Lapujade, P.A.S., Pronk, J. T., 2009. Quantitative physiology of *Saccharomyces cerevisiae* at near-zero specific growth rates. *Appl. Environ. Microbiol.* 75, 5607–5614. <https://doi.org/10.1128/AEM.00429-09>.
- Bujdös, D., Popelárová, B., Volke, D.C., Nikel, P.I., Sonnenschein, N., Dvorák, P., 2023. Engineering of *Pseudomonas putida* for accelerated co-utilization of glucose and cellobiose yields aerobic overproduction of pyruvate explained by an upgraded metabolic model. *Metab. Eng.* 27, 29–46. <https://doi.org/10.1016/j.ymben.2022.10.011>.
- Celińska, E., Grajek, W., 2009. Biotechnological production of 2,3-butanediol—current state and prospects. *Biotechnol. Adv., Biotechnology for the Sustainability of Human Society* 27, 715–725. <https://doi.org/10.1016/j.biotechadv.2009.05.002>.
- Chen, Y., Zhang, Y., Siewers, V., Nielsen, J., 2015. Achi1 is involved in shuttling mitochondrial acetyl units for cytosolic C2 provision in *Saccharomyces cerevisiae* lacking pyruvate decarboxylase. *FEMS Yeast Res.* 15, fov015. <https://doi.org/10.1093/femsy/fov015>.
- Choi, M.-H., Kim, S.-J., Kim, J.-W., Park, Y.-C., Seo, J.-H., 2016. Molecular cloning and expression of *Enterobacter aerogenes* α -acetolactate decarboxylase in pyruvate decarboxylase-deficient *Saccharomyces cerevisiae* for efficient 2,3-butanediol production. *Process Biochem* 51, 170–176. <https://doi.org/10.1016/j.procbio.2015.11.023>.
- Chung, C.H., Lin, D.-W., Eames, A., Chandrasekaran, S., 2021. Next-generation genome-scale metabolic modeling through integration of regulatory mechanisms. *Metabolites* 11, 606. <https://doi.org/10.3390/metabo1109066>.
- Coffman, J.A., Rai, R., Cooper, T.G., 1995. Genetic evidence for Gln3p-independent, nitrogen catabolite repression-sensitive gene expression in *Saccharomyces cerevisiae*. *J. Bacteriol.* 177, 6910–6918. <https://doi.org/10.1128/jb.177.23.6910-6918.1995>.

Cueto-Rojas, H.F., van Maris, A.J.A., Wahl, S.A., Heijnen, J.J., 2015. Thermodynamics-based design of microbial cell factories for anaerobic product formation. *Trends Biotechnol.* 33, 534–546. <https://doi.org/10.1016/j.tibtech.2015.06.010>.

Demichev, V., Messner, C.B., Vernaldis, S.I., Lilley, K.S., Ralser, M., 2020. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* 17, 41–44. <https://doi.org/10.1038/s41592-019-0638-x>.

Durot, M., Bourguignon, P.-Y., Schachter, V., 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol. Rev.* 33, 164–190. <https://doi.org/10.1111/j.1574-6976.2008.00146.x>.

Elseman, I.E., Rodriguez Prado, A., Grigaitis, P., Garcia Albornoz, M., Harman, V., Holman, S.W., van Heerden, J., Bruggeman, F.J., Bisschops, M.M.M., Sonnenseel, N., Hubbard, S., Beynon, R., Daran-Lapujade, P., Nielsen, J., Teusink, B., 2022. Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *Nat. Commun.* 13, 801. <https://doi.org/10.1038/s41467-022-28467-6>.

Entian, K.-D., Köttler, P., 2007. 25 yeast genetic strain and plasmid collections. In: Stansfield, I., Stark, M.J. (Eds.), *Methods in Microbiology, Yeast Gene Analysis*. Academic Press, pp. 629–666. [https://doi.org/10.1016/S0580-9517\(06\)36025-4](https://doi.org/10.1016/S0580-9517(06)36025-4).

Frazer, L.N., O'Keefe, R.T., 2007. A new series of yeast shuttle vectors for the recovery and identification of multiple plasmids from *Saccharomyces cerevisiae*. *Yeast* 24, 777–789. <https://doi.org/10.1002/yea.1509>.

García-Contreras, R., Vos, P., Westerhoff, H.V., Boogerd, F.C., 2012. Why in vivo may not equal in vitro – new effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions. *FEBS J.* 279, 4145–4159. <https://doi.org/10.1111/febs.12007>.

Garg, S.K., Jain, A., 1995. Fermentative production of 2,3-butanediol: a review. *Bioresour. Technol.* 51, 103–109. [https://doi.org/10.1016/0960-8524\(94\)00136-O](https://doi.org/10.1016/0960-8524(94)00136-O).

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Iii, C.A.H., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. <https://doi.org/10.1038/nmeth.1318>.

Heijnen, S.J., 1994. Thermodynamics of microbial growth and its implications for process design. *Trends Biotechnol.* 12, 483–492. [https://doi.org/10.1016/0167-7799\(94\)90056-6](https://doi.org/10.1016/0167-7799(94)90056-6).

Heirendt, L., Arrecks, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., Magnusdóttir, S., Ng, C.Y., Preciat, G., Zagare, A., Chan, S.H.J., Auriach, M.K., Clancy, C.M., Modamio, J., Sauls, J.T., Noronha, A., Bordbar, A., Cousins, B., El Assal, D.C., Valcarcel, L.V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Ben Guebila, M., Kostromins, A., Sompairac, N., Le, H.M., Ma, D., Sun, Y., Wang, L., Yurkovich, J.T., Oliveira, M.A.P., Vuong, P.T., El Assal, L.P., Kuperstein, I., Zinov'yev, A., Hinton, H.S., Bryant, W.A., Aragón Artacho, F.J., Planes, F.J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M.A., Maranas, C.D., Lewis, N.E., Sauter, T., Palsson, B.O., Thiele, I., Fleming, R.M.T., 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* 14, 639–702. <https://doi.org/10.1038/s41596-018-0098-2>.

Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., Krijgsveld, J., 2019. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* 14, 68–85. <https://doi.org/10.1038/s41596-018-0082-x>.

Ishchuk, O.P., Domenzain, I., Sánchez, B.J., Muñiz-Paredes, F., Martínez, J.L., Nielsen, J., Petranovic, D., 2022. Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2108245119. <https://doi.org/10.1073/pnas.2108245119>.

Jørgensen, P., Tyers, M., 2004. How cells coordinate growth and division. *Curr. Biol.* 14, R1014–R1027. <https://doi.org/10.1016/j.cub.2004.11.027>.

Kayikci, Ö., Nielsen, J., 2015. Glucose repression in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15, fov068. <https://doi.org/10.1093/femsyr/fov068>.

Kerkhoven, E.J., 2022. Advances in constraint-based models: methods for improved predictive power based on resource allocation constraints. *Curr. Opin. Microbiol.* 68, 102168. <https://doi.org/10.1016/j.mib.2022.102168>.

Kim, S.-J., Kim, J.-W., Lee, Y.-G., Park, Y.-C., Seo, J.-H., 2017. Metabolic engineering of *Saccharomyces cerevisiae* for 2,3-butanediol production. *Appl. Microbiol. Biotechnol.* 101, 2241–2250. <https://doi.org/10.1007/s00253-017-8172-1>.

Lange, H.C., Heijnen, J.J., 2001. Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 75, 334–344. <https://doi.org/10.1002/bit.10054>.

Li, F., Yuan, L., Lu, H., Li, G., Chen, Y., Engqvist, M.K.M., Kerkhoven, E.J., Nielsen, J., 2022. Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction. *Nat. Catal.* 5, 662–672. <https://doi.org/10.1038/s41929-022-00798-z>.

Liebermeister, W., Noor, E., Flamholz, A., Davidi, D., Bernhardt, J., Milo, R., 2014. Visual account of protein investment in cellular functions. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8488–8493. <https://doi.org/10.1073/pnas.1314810111>.

Liu, Q., Yu, T., Li, X., Chen, Y., Campbell, K., Nielsen, J., Chen, Yun, 2019. Rewiring carbon metabolism in yeast for high level production of aromatic chemicals. *Nat. Commun.* 10, 4976. <https://doi.org/10.1038/s41467-019-12961-5>.

Mans, R., van Rossum, H.M., Wijnsman, M., Backx, A., Kuijpers, N.G.A., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J.A., Daran, J.-M.G., 2015. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15. <https://doi.org/10.1093/femsyr/fov004>.

Mans, R., Wijnsman, M., Daran-Lapujade, P., Daran, J.-M., 2018. A protocol for introduction of multiple genetic modifications in *Saccharomyces cerevisiae* using CRISPR/Cas9. *FEMS Yeast Res.* 18. <https://doi.org/10.1093/femsyr/fov063>.

Metsalu, T., Vilo, J., 2015. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* 43, W566–W570. <https://doi.org/10.1093/nar/gkv468>.

Mizobata, A., Mitsui, R., Yamada, R., Matsumoto, T., Yoshihara, S., Tokumoto, H., Ogino, H., 2021. Improvement of 2,3-butanediol tolerance in *Saccharomyces cerevisiae* by using a novel mutagenesis strategy. *J. Biosci. Bioeng.* 131, 283–289. <https://doi.org/10.1016/j.jbiolsc.2020.11.004>.

Ng, C., Jung, M., Lee, J., Oh, M.-K., 2012. Production of 2,3-butanediol in *Saccharomyces cerevisiae* by in silico aided metabolic engineering. *Microb. Cell Factories* 11, 68. <https://doi.org/10.1186/1475-2859-11-68>.

Nilsson, A., Nielsen, J., 2016. Metabolic trade-offs in yeast are caused by F1FO-ATP synthase. *Sci. Rep.* 6, 22264. <https://doi.org/10.1038/srep22264>.

Niu, J., Mao, Z., Mao, Y., Wu, K., Shi, Z., Yuan, Q., Cai, J., Ma, H., 2022. Construction and analysis of an enzyme-constrained metabolic model of *Corynebacterium glutamicum*. *Biomol.* 12, 1499. <https://doi.org/10.1390/biom12101499>.

Niu, P., Soto, M.J., Yoon, B.-J., Dougherty, E.R., Alexander, F.J., Blaby, I., Qian, X., 2021. TRIMER: transcription regulation integrated with metabolic regulation. *iScience* 24. <https://doi.org/10.1016/j.isci.2021.103218>.

O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., Palsson, B.O., 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* 9, 693. <https://doi.org/10.1038/msb.2013.52>.

Oftadeh, O., Salvy, P., Masid, M., Curvat, M., Miskovic, L., Hatzimanikatis, V., 2021. A genome-scale metabolic model of *Saccharomyces cerevisiae* that integrates expression constraints and reaction thermodynamics. *Nat. Commun.* 12, 4790. <https://doi.org/10.1038/s41467-021-25158-6>.

Österberg, L., Domenzain, I., Münch, J., Nielsen, J., Hohmann, S., Cvijovic, M., 2021. A novel yeast hybrid modeling framework integrating Boolean and enzyme-constrained networks enables exploration of the interplay between signaling and metabolism. *PLoS Comput. Biol.* 17, e1008891. <https://doi.org/10.1371/journal.pcbi.1008891>.

Oud, B., Flores, C.-L., Gancedo, C., Zhang, X., Trueheart, J., Daran, J.-M., Pronk, J.T., van Maris, A.J., 2012. An internal deletion in MTH1 enables growth on glucose of pyruvate-decarboxylase negative, non-fermentative *Saccharomyces cerevisiae*. *Microb. Cell Factories* 11, 131. <https://doi.org/10.1186/1475-2859-11-131>.

Papapetridis, I., Goudriaan, M., Vázquez Vitali, M., de Keijzer, N.A., van den Broek, M., van Maris, A.J.A., Pronk, J.T., 2018. Optimizing anaerobic growth rate and fermentation kinetics in *Saccharomyces cerevisiae* strains expressing Calvin-cycle enzymes for improved ethanol yield. *Biotechnol. Biofuels* 11. <https://doi.org/10.1186/s13068-017-1001-z>.

Patil, K.R., Åkesson, M., Nielsen, J., 2004. Use of genome-scale microbial models for metabolic engineering. *Curr. Opin. Biotechnol.* 15, 64–69. <https://doi.org/10.1016/j.copbio.2003.11.003>.

Peterson, C.L., Herskowitz, I., 1992. Characterization of the yeast SW11, SW12, and SW13 genes, which encode a global activator of transcription. *Cell* 68, 573–583. [https://doi.org/10.1016/0092-8674\(92\)90192-F](https://doi.org/10.1016/0092-8674(92)90192-F).

Pramanik, J., Keasling, J.D., 1997. Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* 56, 398–421. [https://doi.org/10.1002/\(SICI\)1097-0290\(19971120\)56:4<398::AID-BIT6>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0290(19971120)56:4<398::AID-BIT6>3.0.CO;2-J).

Pronk, J.T., 2002. Auxotrophic yeast strains in fundamental and applied research. *Appl. Environ. Microbiol.* 68, 2095–2100. <https://doi.org/10.1128/AEM.68.5.2095-2100.2002>.

Salvy, P., Hatzimanikatis, V., 2020. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat. Commun.* 11, 30. <https://doi.org/10.1038/s41467-019-13818-7>.

Sánchez, B.J., Lahtvee, P.-J., Campbell, K., Kasvandik, S., Yu, R., Domenzain, I., Zelezník, A., Nielsen, J., 2021. Benchmarking accuracy and precision of intensity-based absolute quantification of protein abundances in *Saccharomyces cerevisiae*. *Proteome* 21, 2000093. <https://doi.org/10.1002/pmic.202000093>.

Sánchez, B.J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E.J., Nielsen, J., 2017. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* 13. <https://doi.org/10.15252/msb.20167411>.

Scott, M., Hwa, T., 2011. Bacterial growth laws and their applications. *Curr. Opin. Biotechnol.* 22, 559–565. <https://doi.org/10.1016/j.copbio.2011.04.014>.

Sinha, N., van Schothorst, E.M., Hooiveld, G.J.E.J., Keijer, J., Martins dos Santos, V.A.P., Suárez-Díez, M., 2021. Exploring the associations between transcript levels and fluxes in constraint-based models of metabolism. *BMC Bioinf.* 22, 574. <https://doi.org/10.1186/s12859-021-04488-8>.

Sudarsanan, P., Iyer, V.R., Brown, P.O., Winston, F., 2000. Whole-genome expression analysis of snf2 mutants of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 3364–3369. <https://doi.org/10.1073/pnas.97.7.3364>.

Taguchi, A.K.W., Young, E.T., 1987. The cloning and mapping of ADR6, a gene required for sporulation and for expression of the alcohol dehydrogenase II isozyme from *Saccharomyces cerevisiae*. *For. Genet.* 116, 531–540. <https://doi.org/10.1093/genetics/116.4.531>.

van Dijken, J.P., Scheffers, W.A., 1986. Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiol. Rev.* 1, 199–224. <https://doi.org/10.1111/j.1574-6968.1986.tb01194.x>.

Väremo, L., Nielsen, J., Nookaew, I., 2013. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res.* 41, 4378–4391. <https://doi.org/10.1093/nar/gkt111>.

Verduyn, C., Postma, E., Scheffers, W.A., Van Dijken, J.P., 1992. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8, 501–517. <https://doi.org/10.1002/yea.320080703>.

Verduyn, C., Postma, E., Scheffers, W.A., van Dijken, J.P., 1990. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *Microbiology* 136, 395–403. <https://doi.org/10.1099/00221287-136-3-395>.

Wang, H., Marcisauskas, S., Sánchez, B.J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., Kerkhoven, E.J., 2018. Raven 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput. Biol.* 14, e1006541 <https://doi.org/10.1371/journal.pcbi.1006541>.

Wu, K., Mao, Z., Mao, Y., Niu, J., Cai, J., Yuan, Q., Yun, L., Liao, X., Wang, Z., Ma, H., 2023. eCBSU1: a genome-scale enzyme-constrained model of *Bacillus subtilis* based on the ECMPy workflow. *Microorganisms* 11, 178. <https://doi.org/10.3390/microorganisms11010178>.

Yang, X., Mao, Z., Zhao, X., Wang, R., Zhang, P., Cai, J., Xue, C., Ma, H., 2021. Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models. *Metab. Eng.* 67, 133–144. <https://doi.org/10.1016/j.ymben.2021.06.005>.

Ye, C., Luo, Q., Guo, L., Gao, C., Xu, N., Zhang, L., Liu, L., Chen, X., 2020. Improving lysine production through construction of an *Escherichia coli* enzyme-constrained model. *Biotechnol. Bioeng.* 117, 3533–3544. <https://doi.org/10.1002/bit.27485>.

Zhou, J., Zhuang, Y., Xia, J., 2021. Integration of enzyme constraints in a genome-scale metabolic model of *Aspergillus niger* improves phenotype predictions. *Microb. Cell Factories* 20, 125. <https://doi.org/10.1186/s12934-021-01614-2>.

Publication II

Rekēna, A., Pinheiro, M. J., Bonturi, N., Belouah, I., Tammekivi, E., Herodes, K., Kerkhoven, E. J., Lahtvee, P.-J. (2023). Genome-scale metabolic modelling reveals metabolic trade-offs associated with lipid production in *Rhodotorula toruloides*. *PLOS Computational Biology*, 19(4), e1011009.

RESEARCH ARTICLE

Genome-scale metabolic modeling reveals metabolic trade-offs associated with lipid production in *Rhodotorula toruloides*

Alīna Reķēna¹, Marina J. Pinheiro², Nemailla Bonturi¹, Isma Belouah¹, Eliise Tammekivi³, Koit Herodes³, Eduard J. Kerkhoven⁴, Petri-Jaan Lahtvee^{1*}

1 Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia,

2 Department of Materials and Bioprocess Engineering, University of Campinas, Campinas, Brazil,

3 Institute of Chemistry, University of Tartu, Tartu, Estonia, **4** Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

* lahtvee@taltech.ee



Abstract

Rhodotorula toruloides is a non-conventional, oleaginous yeast able to naturally accumulate high amounts of microbial lipids. Constraint-based modeling of *R. toruloides* has been mainly focused on the comparison of experimentally measured and model predicted growth rates, while the intracellular flux patterns have been analyzed on a rather general level. Hence, the intrinsic metabolic properties of *R. toruloides* that make lipid synthesis possible are not thoroughly understood. At the same time, the lack of diverse physiological data sets has often been the bottleneck to predict accurate fluxes. In this study, we collected detailed physiology data sets of *R. toruloides* while growing on glucose, xylose, and acetate as the sole carbon source in chemically defined medium. Regardless of the carbon source, the growth was divided into two phases from which proteomic and lipidomic data were collected. Complemental physiological parameters were collected in these two phases and altogether implemented into metabolic models. Simulated intracellular flux patterns demonstrated the role of phosphoketolase in the generation of acetyl-CoA, one of the main precursors during lipid biosynthesis, while the role of ATP citrate lyase was not confirmed. Metabolic modeling on xylose as a carbon substrate was greatly improved by the detection of chirality of D-arabinitol, which together with D-ribulose were involved in an alternative xylose assimilation pathway. Further, flux patterns pointed to metabolic trade-offs associated with NADPH allocation between nitrogen assimilation and lipid biosynthetic pathways, which was linked to large-scale differences in protein and lipid content. This work includes the first extensive multi-condition analysis of *R. toruloides* using enzyme-constrained models and quantitative proteomics. Further, more precise k_{cat} values should extend the application of the newly developed enzyme-constrained models that are publicly available for future studies.

Author summary

Transition towards a biobased, circular economy to reduce the industrial dependence on fossil-based resources requires new technologies. One of the options is to convert available

[alnarekena/ecRhtoGEM](https://doi.org/10.1371/journal.pcbi.1011009.g001). LC-MS/MS data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD037281.

Funding: This work was supported by the Estonian Research Council grants PUT1488P, PRG1101 (AR, MJP, NB, IB, PJL) and PRG1198 (ET) and NordForsk grant 103506 (AR, NB, PJL). MJP would additionally like to acknowledge Coordination for the Improvement of Higher Education Personnel (Capes), São Paulo Research Foundation (FAPESP, grant 2016/10638-8). EJK acknowledges funding by the Novo Nordisk Foundation (grant NNF10CC1016517), and the Research Council for Environment, Agricultural Sciences, and Spatial Planning (Formas) (grant 2018-00597). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: Authors declare no competing interests.

biomass feedstocks into valuable chemicals using microbes as biocatalysts. *Rhodotorula toruloides* is a nonpathogenic, nonconventional yeast that has recently emerged as one of the most promising yeasts for sustainable production of chemicals and fuels due to its natural ability to synthesize large amounts of lipids. However, its unique metabolic properties are not yet fully understood. We have computationally predicted metabolic fluxes in *R. toruloides* while growing in economically viable growth conditions inducing lipid accumulation and analyzed them together with absolute proteome quantification. Our holistic approach has highlighted metabolic pathways important for lipid biosynthesis and revealed metabolic trade-offs associated with NADPH allocation during lipogenesis. In addition, our work highlighted the necessity for accurate computational approaches in characterizing enzymatic kinetic properties that would improve the metabolic studies of *R. toruloides*.

Introduction

R. toruloides is a red basidiomycota known for its ability to accumulate high amounts of intracellular lipids [1] and consume different carbon substrates [2,3]. It has been studied for its ability to consume complex biomass substrates, including from the lignocellulosic origin [4–6] that would make it interesting for a biorefinery concept. However, studies aimed at fundamental investigation of *R. toruloides* metabolism have been mainly conducted using a single carbon source as substrate, such as xylose [1,7–9], glucose [8–11], glycerol [7], acetate [7], L-arabinose and *p*-coumarate [9], in a chemically defined mineral medium and occasionally rich cultivation medium (YP) [3]. Secondary nutrient limitation induces lipid accumulation [12]. In nitrogen limitation, 65% of lipids of dry cell weight were reached in a batch cultivation regime [1].

Metabolic pathways producing intracellular metabolite acetyl-CoA and a cofactor NADPH in *R. toruloides* have been the main focus of metabolic studies due to their central role in lipid biosynthesis. Fatty acids, which mainly accumulate in the form of triacylglycerols (TAGs), are produced via the sequence of four enzymatic reactions that require 1 ATP and 2 NADPH molecules per 1 acetyl-CoA added to the fatty acid chain [13]. To study lipid metabolism in *R. toruloides*, previous studies have taken the systems biology approach. Genome sequencing and high-throughput multi-omics analysis facilitated the reconstruction of the metabolic networks. Based on a genome sequence of *R. toruloides* strain NP11, the first metabolic network of *R. toruloides* included its central carbon metabolism and lipid biosynthetic pathways [10]. *R. toruloides* possesses several enzymatic pathways that differ from the model yeast *Saccharomyces cerevisiae* and which specifically facilitate the generation of lipid precursors. The key differences included the synthesis of acetyl-CoA from citrate by ATP citrate lyase (ACL), synthesis of acetyl-CoA from xylulose 5-phosphate by phosphoketolase (XPK), and the conversion of S-malate into pyruvate by malic enzyme (ME) that provides for NADPH [10,14]. Proteomics analysis has suggested NADPH regeneration primarily through the pentose phosphate pathway on xylose and glucose, but the role of malic enzyme is not clearly understood [8,10]. The role of XPK in the generation of acetyl-CoA has not been acknowledged previously, whereas ACL has been demonstrated to be upregulated during lipid accumulation [10], especially in presence of xylose [8].

It has been reported that on xylose *R. toruloides* is growing 2 to 3 times slower compared to glucose [8], but the underlying mechanisms are yet to be identified. In our previous proteomics study [1], we discovered from proteomics quantification that xylulokinase, encoded in

the genome as the third step in the currently known xylose assimilation pathway, is not present in the proteomic data set, suggesting potential limitation in xylose metabolism. Later, a similar finding was reported by Jagtap et al. 2021 [3] and Kim et al. 2021 [2] using a different *R. toruloides* strain, IFO 0880. The latter proposed an alternative xylose assimilation pathway for this species.

A holistic view on metabolism can be provided by genome-scale metabolic models (GEMs). GEMs are metabolic networks reconstructed from a genome sequence of a specific organism. They contain all known biochemical reactions of the organism. GEMs allow the calculation of metabolic fluxes that represent activity of different metabolic pathways under specified conditions, e.g., an uptake of a particular carbon source. GEMs of *R. toruloides* were built based on the genome sequences of strains NP11 [15] and IFO 0880 [11]. Flux balance analysis predicted that up to 87% of NADPH was regenerated from xylose through the oxidative part of pentose phosphate pathway (oxPPP) [1]. Phosphoketolase was suggested as the main supplier of acetyl-CoA during lipogenesis in xylose-grown cells [1]. On the other hand, TCA cycle related enzymes were suggested for NADPH production on acetate-grown cells [7], demonstrating that metabolic operations can vary significantly with the carbon source uptake. Models have also been used to study metabolism during cell growth on glucose [11] and glycerol [7].

A better understanding of how different metabolic pathways contribute to lipid accumulation under different substrates would help to design better metabolic engineering strategies. GEMs can be a powerful and helpful tool in metabolic studies, if their predictive power is good. Enzyme-constrained GEMs integrate additional constraints on enzyme capacity and their total abundances (as thoroughly reviewed by Chen and Nielsen 2021 [16]). Phenomenological constraint is imposed on metabolic flux (v ; mmol/gDCW/h), formulated as enzyme kinetics (Eq 1)

$$v \leq E \cdot k_{cat} \quad (1)$$

where E is protein abundance (mmol/gDCW) and k_{cat} is the enzyme's turnover number (1/s), provided with an upper limit on individual or total protein abundances. The integration of enzymatic constraints in *S. cerevisiae* significantly improved phenotype prediction [17]. The strength of proteome constraints has also been demonstrated by predicting overflow metabolism in *E. coli* [18] and metabolic shift in arginine catabolism in *L. lactis* [19]. A similar coarse-grained approach that allowed the prediction of maximal growth without constraining the model with any exchange fluxes in *S. cerevisiae* was demonstrated by applying a global thermodynamics constraint [20].

In addition to a curated annotation, the quality of the predicted fluxes depends on accuracy of physiological data, notably on the biomass composition specificity. The tuning of *R. toruloides* biomass reaction in the prior and current models improved the condition-wise specificity of predicted fluxes.

In the present study, we created condition-specific enzyme-constrained genome-scale metabolic models of *R. toruloides*, ecRhtoGEMs, and used them to predict intracellular fluxes. Flux bounds to constrain the model were obtained from bioreactor (1 L) experiments with yeast cultivation in chemically defined medium, with three carbon sources studied individually—glucose, xylose and acetate. These very detailed physiological data sets enabled us to precisely characterize metabolism at exponential growth and lipid accumulation phase. In all conditions, we performed mass spectrometry (MS) based absolute proteome quantification. Also, biomass macromolecular composition in regard to lipids and proteins was determined, including lipid profiling by gas chromatography (GC) analysis. Using this data, we generated 6 different versions of the *R. toruloides* model with enzyme constraints and biomass composition

specificity, where we were able to demonstrate trade-offs in NADPH requirements for the cells growing exponentially versus in nitrogen limitation. To our knowledge, this is the first proteomics analysis of acetate-grown *R. toruloides* cells and the first detailed GEM analysis combined with proteome analysis of acetate and glucose conditions for this strain.

Results

Differences in physiological parameters under glucose, xylose or acetate as a sole carbon source

Here we present production profile, specific growth rate, lipidomics and total protein measurements of batch cultivation of *R. toruloides* strain CCT 7815 growing in a chemically defined medium on three substrates as a sole carbon source—glucose (63 g/L), xylose (70 g/L) or acetate (20 g/L). *R. toruloides* CCT 7815 is a tolerant strain developed during a short-term adaptation of strain CCT 0783 (Coleção de Culturas Tropicais, Fundação André Tosello, Campinas, Brazil) in sugarcane bagasse hemicellulosic hydrolyzate, demonstrating an increased lipid production without impacting growth and substrate consumption as a result of induction of hydrolysate-tolerance- and lipid accumulation-related genes [21]. Cultures were grown at a starting molar C/N ratio of 69 (glucose/urea) and 80 (xylose- or acetate/ammonium sulfate), which will result in nitrogen limitation that is known to induce lipid accumulation [12]. Cell growth was monitored by online biomass measurements and CO₂ production data. Experiments were run until complete substrate depletion. Regardless of the carbon source, the results demonstrated two distinct growth phases: (i) exponential growth (exp) phase where all substrates were in excess, and (ii) nitrogen-limited (Nlim) phase, associated with nitrogen depletion (Fig 1A). For lipidomics, the first sample was analyzed at the end of exp phase and the second sample was analyzed at the end of Nlim phase (Fig 1A). For intracellular protein content analysis, biomass samples were analyzed at the late or end exp and mid-Nlim phases (Fig 1A). Physiological parameters are available in S1 Table.

The highest amount of intracellular lipids was accumulated while cells were growing on glucose, resulting in 0.48 ± 0.04 g/gDCW, while the lipid yield was approximately 15% and 20% lower on acetate and xylose, respectively (Fig 1B). On glucose, lipid accumulation started later than on xylose and acetate, where up to 20% and 18% lipid yield, respectively, was reached already during the late exp phase (S1 Fig). In a similar study using a different *R. toruloides* strain NP11 [8], less lipids were quantified in xylose at the late exp phase, while a higher final lipid yield was reached compared to our study. On glucose, the final lipid yield was comparable with previous studies [8]. In acetate condition, the final lipid content was 0.34 ± 0.01 g/gDCW, which was in line with previous experiments by our group measured in continuous cultivation experiments [7].

Maximum specific growth rate was the highest on glucose, 0.19 ± 0.025 h⁻¹, while it was at least 2-fold lower on acetate and xylose (Fig 1C). Two-fold difference in maximum growth rate on xylose and glucose conditions has been reported previously [8]. Nlim growth phase, where most of the lipid accumulation occurred in all studied conditions, was characterized by significantly lower specific growth rate (Fig 1C), specific substrate uptake rate (Fig 1D) and total protein content (Fig 1E).

Lipid composition was similar in all studied conditions, with oleate (C18:1) as the dominant fatty acid (Fig 1F). During the Nlim phase, the relative amount of oleate (monounsaturated fatty acids, MUFAs) further increased, while polyunsaturated fatty acids (PUFAs)—linoleate (C18:2) and linolenate (C18:3)—decreased. Interestingly, on glucose at late exp phase the amount of PUFAs was higher than MUFAs, but it significantly changed during the Nlim phase when total lipid amount increased almost 10-fold (Fig 1B). Our results demonstrated that the

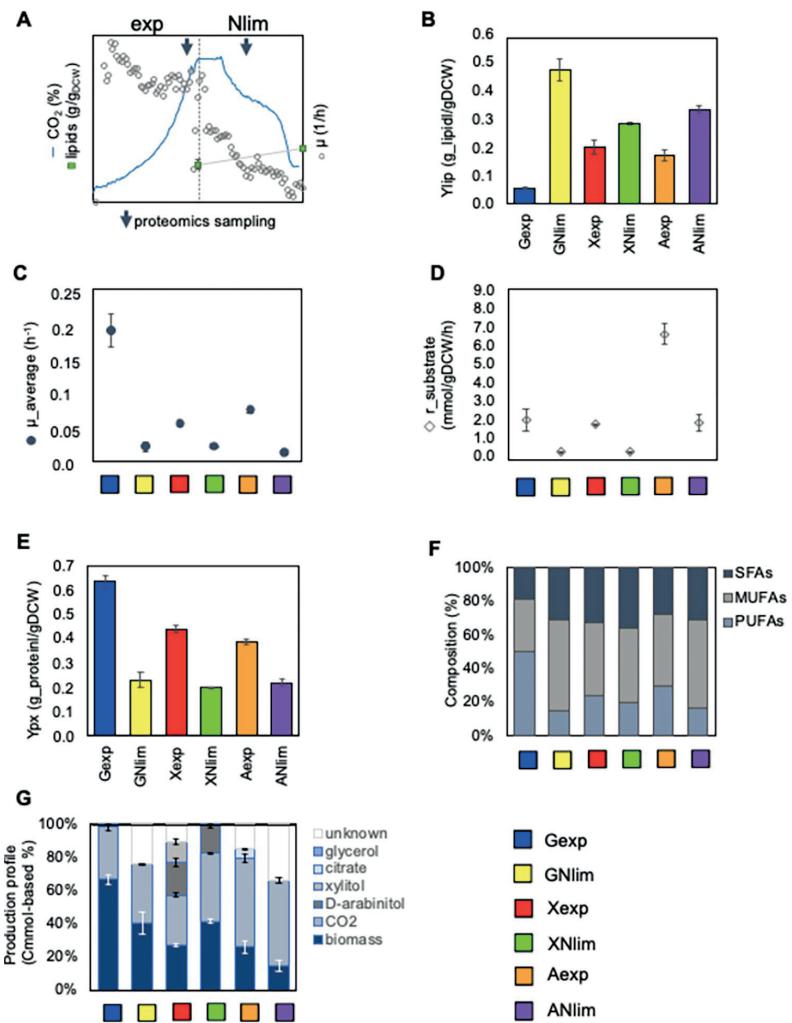


Fig 1. Growth characterization on three carbon sources—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases. (A) Batch cultivation growth curve and sampling timepoints for lipidomics and proteomics on all tested carbon sources. (B) Lipid yield, Y_{lip} (g/gDCW). (C) Average specific growth rate, μ_{average} (h⁻¹). (D) Substrate uptake rate, $r_{\text{substrate}}$ (mmol/gDCW/h). (E) Protein content, Y_{p} (g/gDCW). (F) Fatty acid profiles (% of total lipid). (G) Carbon balance (% of total substrate uptake). SFAs: saturated fatty acids; MUFA: monounsaturated fatty acids; PUFA: polyunsaturated fatty acids. Average of duplicate experiments with SD is illustrated.

<https://doi.org/10.1371/journal.pcbi.1011009.g001>

degree of fatty acid saturation for C18 dynamically increased throughout cultivation. The distribution of different types of fatty acids was in agreement with the previous data reported on glucose and xylose [8], whereby oleate and palmitate were the most abundant at the end of batch cultivation. Notwithstanding the general agreement between both studies, in our study the PUFA, mainly linoleate (C18:2) increased more during Nlim, while in the previous study

[8] palmitate (C16:0) increased more during lipid accumulation. It might possibly reflect the fact that different *R. toruloides* strains were used in these studies.

Final biomass titers were similar on xylose and glucose, respectively 18 g/L and 22 g/L ([S1 Fig](#)), but the highest biomass yield, 0.32 gDCW/g_substrate, was reached on xylose during Nlim phase ([S1 Table](#)). On xylose, 32% of substrate was excreted as byproducts xylitol and D-arabinitol during the exp growth phase ([Fig 1G](#)). For arabinitol, a stereoselective analysis was done using high-performance liquid chromatography (HPLC) separation with a chiral column (Chiralpak, Daicel Technologies, Japan), similarly as described in Lopes and Gaspar 2008 [22] ([S2 Fig](#)). Although Jagtap and Rao [23] already assumed the production of D-arabinitol, we were able to validate it. At low growth rates (during the Nlim phase), xylitol and D-arabinitol were not excreted but rather co-consumed. All byproducts were consumed at the end of the experiment at 168 h. On acetate, the amount of byproducts other than CO₂ increased during the Nlim phase to 31% ([Fig 1G](#)). These byproducts remain to be identified. On glucose, no byproduct other than CO₂ was detected. However, we were able to measure only 68.5% of carbon during the Nlim phase ([Fig 1G and S1 Table](#)). Likely, it was because *R. toruloides* strain CCT 7815 was making cell aggregates when grown in the chemically defined glucose-based medium. Mass balance calculation took into account glucose uptake, carbon dioxide production and biomass (in C-mol). As our biomass measurements were based on optical density, which relies on the assumption that cells are evenly distributed and of equal size [24,25], it may underestimate the actual cell concentration in liquid culture when aggregates are formed. To solve aggregation problem, we switched the nitrogen source from ammonium sulfate to urea in glucose condition. It helped to reduce aggregate formation but did not eliminate it. Based on the comparison of growth curves when using ammonium sulfate or urea, the results were highly similar ([S1A Fig](#)). Further analysis with glucose was carried out using urea as the nitrogen source. Therefore, results exclusively in the glucose condition (both exp and Nlim growth phases), including the proteomics and metabolic flux data presented in [Figs 1–5](#) and [S1–S13 and S1–S6 Tables and S1–S6 Datasets](#), belong to experiments in which we used urea as a nitrogen source.

Proteomics data shows a significant allocation into ribosomes

We also present a high quality dataset with absolute proteome abundances of *R. toruloides* measured at the late exp and mid-Nlim phases during growth on xylose, glucose and acetate. Proteins were measured and quantified with mass spectrometry-based TPA (total protein amount) quantification method [26], and we were able to determine the absolute abundances of 3160 proteins across 6 conditions ([S1 Dataset](#)). Principal component (PC) analysis showed coherency in our proteome data ([Fig 2A](#)). High similarity between acetate exp and Nlim data was detected, while showing significant differences with other studied conditions (separated on the PC1, describing 49% variation in the data). PC analysis has previously been done for *R. toruloides* strain IFO 0880 comparing gene expression during the exponential growth phase on rich medium containing sole carbon substrate, similarly as in our study [3]. PC1 using transcriptomics revealed distinct expression patterns on acetate-grown as compared to glucose- and xylose-grown cells, agreeing with the proteomics results obtained our study. The only noticeable difference was that PC2 in [3] separated exp phase from glucose to xylose. In our study, PC2, describing 34% of the variation in our data, separated mainly exp and Nlim conditions in the same way on glucose and xylose ([Fig 2A](#)).

Significant variation in proteome between the two growth phases, exp and Nlim, was also observed by differential expression analysis. We found 204 differentially expressed proteins in Nlim (lipid accumulation) versus exp growth phase on glucose, 37 on xylose and none on

acetate using a cut-off of $|\log_{2}FC| > 1$ and Benjamini-Hochberg corrected p-value < 0.05 ([S3 Fig](#) and [S2 Dataset](#)). Proteome profiles on xylose and glucose were more similar in comparison to the growth on acetate. Comparison of protein levels between carbon sources revealed the largest difference between xylose and acetate at exp growth phase, resulting in 159 differentially expressed proteins ([S3 Fig](#) and [S2 Dataset](#)). We then analyzed protein levels based on Gene Ontology (GO) group relations that represent different metabolic pathways present in *R. toruloides*. GO groups were obtained from the Uniprot database (*R. toruloides* NP11) and genome-scale model, rhto-GEM [15] (for a full list see [S2 Table](#)). GO relations from both sources were combined to provide possibly the most accurate information on different metabolic pathways present in *R. toruloides*.

We discovered that ribosomes formed the largest GO group of the proteome (data were represented per gram of total proteome, $\mu\text{g/g_protein}$) ([Fig 2B](#)). Interestingly, the ribosomal abundance in *R. toruloides* up to 46% of proteome was higher than observed previously in *S. cerevisiae* (around 37%) [27]. Expression levels in glycolytic pathways were largely unchanged during the lipid accumulation in xylose and acetate conditions, while upregulation was observed on glucose ([Fig 2B](#)). On glucose- and xylose-grown cells, proteome allocation to TCA cycle was considerably lower compared to glycolytic metabolic pathways ([Fig 2B](#)). On acetate, protein levels of the TCA cycle were almost 3-fold higher than on glucose and xylose ([Fig 2B](#)). Higher TCA cycle activity was expected as acetate assimilation directly produces TCA cycle-related metabolites. The electron transfer chain (ETC) was the only metabolic pathway, in which protein levels increased significantly during lipid production in all the studied carbon substrates ([Fig 2B](#)).

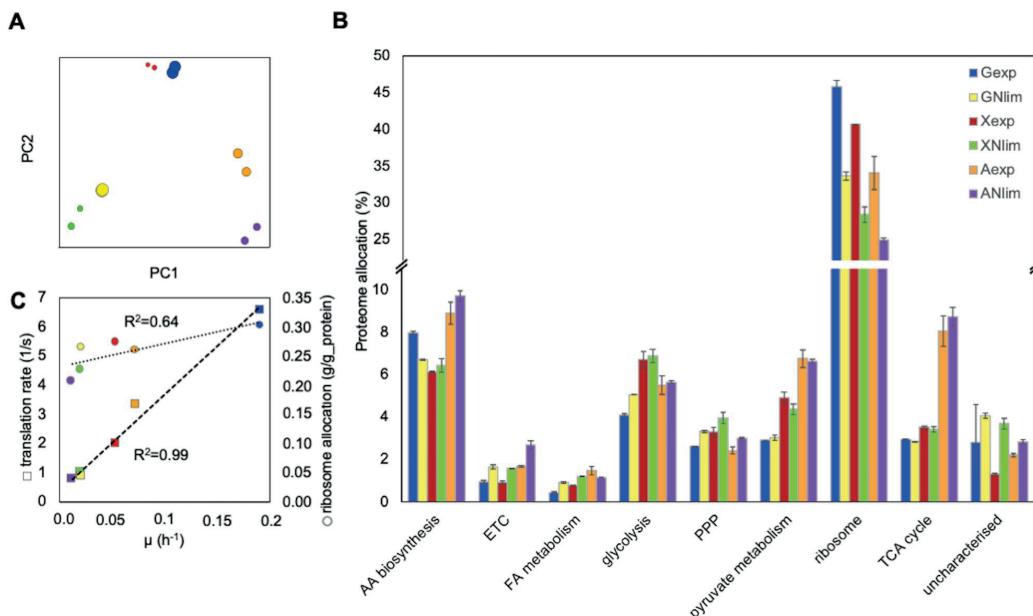


Fig 2. Absolute proteomics data. (A) Principal component analysis ($\mu\text{g/g_protein}$). (B) Proteome allocation (% of $\mu\text{g/g}$ total protein) to metabolic pathways associated with amino acid (AA) biosynthesis, electron transport chain (ETC), fatty acid (FA) metabolism, glycolysis, pentose phosphate pathway (PPP), pyruvate metabolism, ribosome, tricarboxylic acid (TCA) cycle, and uncharacterized proteins. (C) Ribosomal translation rate (s^{-1}) and ribosome allocation (g/g_protein). Average of duplicate experiments with SD is illustrated. Proteins in each GO group are shown in [S2 Table](#).

<https://doi.org/10.1371/journal.pcbi.1011009.g002>

During lipid accumulation, the amounts of uncharacterised proteins, especially in xylose condition, increased (Fig 2B), indicating the importance of discovering unknown protein functions for future research. Interestingly, on glucose the least amount of proteins were allocated to FA metabolism, while the highest total lipid content was measured experimentally. The highest expression levels of proteins in the fatty acid metabolic pathways were detected on acetate (Fig 2B). It was mostly due to high expression levels of beta-oxidation proteins (S1 Dataset).

Relation between ribosomal content, growth rate and translation

We used absolute quantification of the proteome and the ribosomal content to calculate the rate of protein synthesis per ribosome, also known as ribosome efficiency or protein translation rate (for instructions see S3 Table). The ribosome of *R. toruloides* strain NP11 was characterized by 178 structurally distinct proteins reported in Uniprot.org, from which 147 were identified in CCT 7815 strain and quantified (S2 Table). The calculated translation rates varied from 0.8 to 6.6 aa/s (Fig 2C), which was very similar as observed in *S. cerevisiae* (between 2.8 and 10 aa/s) [28]. Among the 6 conditions analyzed, we observed a linear correlation between the translation rate and specific growth rate μ ($R^2 = 0.99$, p-value < 0.001). The mass-wise ribosome content of proteome (g/g_protein) (Fig 2C) had no such distinct correlation with the μ ($R^2 = 0.68$, p-value = 0.043). Interestingly, the lowest ribosome content in proteome was detected during growth on acetate as compared to other substrates.

Integrating fluxomic and proteomic analysis using an enzyme-constrained genome-scale model

Genome-scale models allow an *in silico* simulation of intracellular flux patterns in accordance with exchange fluxes obtained from cultivation experiments. To improve the predictive power and consider the capacity constraints imposed by enzymatic catalytic capacities and their protein levels, we developed an enzyme-constrained GEM using the GECKO Toolbox [17]. In lieu of a strain-specific model, we used the NP11-based GEM [15] to represent the CCT 7815 strain used in this study. The genome of its parental strain CCT 0783 possesses two versions of the same gene, one presenting >90% identity and the other version presenting >70% identity to the genome of haploid strain NP11 [29]. We integrated individual protein concentrations with their corresponding catalytic activities (k_{cat}) in the model to constrain individual metabolic fluxes. We created separate models for exp and Nlim phases on xylose (X), glucose (G) and acetate (A), respectively. Hence, 6 different versions of the proteome constrained model with modified biomass composition, fatty acid profiles and flux bounds from the experimental data were constructed. Proteome constraints included the concentrations of 773 different enzymes across all conditions (S4A Fig and S4 Table), which were applied to 1515 metabolic reactions (30% of all reactions) (S5 Table). The coverage of these constraints was greatly improved by manually assigning EC numbers to 461 *R. toruloides* enzymes (S4 Table), which enabled GECKO Toolbox to assign their k_{cat} values. At first, BRENDA was queried for exact matching reaction, substrate and organism. But as kinetic parameter data for non-model organisms such as *R. toruloides* were not readily available, GECKO Toolbox step-wise relaxes the stringency when matching EC number, organism and substrate, to assign reasonable estimates of k_{cat} values [30]. Mass-wise, the proteome constraints of the measured fraction of enzymes covered between 14% (Gexp) to 25% (ANlim) of the quantified proteome (S4B Fig). Aside from enzyme concentrations, proteome constraints contained 535 unique k_{cat} values automatically queried from the BRENDA database (S3 Dataset). Models, data sets and scripts are hosted on a dedicated Github repository ecRhtoGEM ([www.github.com/alinarekena/ecRhtoGEM](https://github.com/alinarekena/ecRhtoGEM)).

Next we used Flux Balance Analysis [31] to simulate intracellular flux patterns and random sampling of the solution space [32] with 2000 sampling iterations to evaluate flux variability ([S4 Dataset](#)). To constrain the set of feasible solutions during sampling, we fixed the upper bound and lower bounds on the observed exchange fluxes, ATP hydrolysis (non-growth related maintenance) and protein pool exchange (see [Methods](#)). The average flux variability, estimated as a percentage of SD divided by the flux values, was 19% (median value of all conditions) ([S4 Dataset](#)).

From the simulated flux values, we calculated apparent enzyme catalytic activities (k_{app}) as ratio of model-predicted fluxes and measured protein concentration. k_{app} represents the apparent *in vivo* enzyme turnover which drives the biological processes in the environment, in contrast to k_{cat} representing maximum enzyme capacity. As we used model-predicted fluxes that were constrained by k_{cat} values in the ec-model, the k_{app} values that we obtained cannot be higher than the k_{cat} value, which means that we cannot capture any potential *in vivo* enzyme activity enhancement effect. Regardless, in case of high k_{app} values, high reaction rates are catalyzed by low protein concentration, and vice versa. This study is the first report on the *in silico* k_{app} values in *R. toruloides*. Calculated k_{app} values in all growth conditions are available in [S3 Dataset](#). Vast majority of k_{app} values were in the range from 0.1 to 100 (s^{-1}) ([S5 Fig](#)), which is in the range of “average enzyme” k_{cat} of $10\ s^{-1}$ reported by Bar-Even et al. [33]. Some of the lowest k_{app} values in acetate condition were associated with fatty acid degradation and beta oxidation metabolic pathways. We found that during the Nlim phase, when lipid accumulation occurs, the number of enzymes with relatively low k_{app} values ($0.1\ to\ 1\ s^{-1}$) was increased ([S5 Fig](#)). This reflects the fact that absolute fluxes decreased more than protein concentrations during the Nlim phase in comparison to exp growth phase, suggesting that for many reactions downregulation of the enzyme did not affect its reaction rate directly.

Growth on glucose

In our analysis of integrated flux and proteomics data, we focused on the major carbon fluxes and corresponding enzymes in the central carbon and lipid metabolism where acetyl-CoA, ATP and NADPH, the main precursors for lipid biosynthesis, are generated ([Fig 3](#)). For a better comparison, fluxes were normalized to the substrate uptake rate of the respective condition, providing percentage values of carbon distribution in the metabolic pathways ([S6 Fig](#), for a full list see [S4 Dataset](#)). During the exponential growth phase on glucose, 72% of the carbon was directed via the PPP, while only 20% went through the Embden-Meyerhoff glycolytic pathway. When comparing fluxes at exp versus Nlim phase, we did not observe any significant changes in normalized fluxes through the oxPPP in glucose condition ([Fig 4A](#)), which was also the main source of NADPH regeneration ([S7 Fig](#)) (reaching 76% glucose-derived carbon). During the exp phase, the majority of NADPH was consumed by glutamate dehydrogenase (GDH) which converts ammonium and oxoglutarate (AKG) to glutamate, while during the Nlim phase majority of NADPH was consumed in lipid biosynthesis by FAS1-2 ([S7 Fig](#)).

The flux via phosphoketolase pathway, which converts D-xylulose 5-phosphate to glyceraldehyde 3-phosphate and acetyl-CoA, increased more than 4-fold from 14% to 60% during the transition from exp growth to Nlim phase, consistent with a significant upregulation of phosphoketolase (XPK) on proteome level (apval. 0.043, [S2 Dataset](#)). While it is not known, which route of XPK enzyme in combination with a phosphotransacetylase (PTA) or an acetate kinase (ACK) is used in *R. toruloides* strain CCT 7815, we compared the fluxes of both possible scenarios ([S4](#) and [S5 Datasets](#)). As the results were highly similar, further flux analysis was carried out based on a metabolic route where PTA is active. XPK pathway was also the main

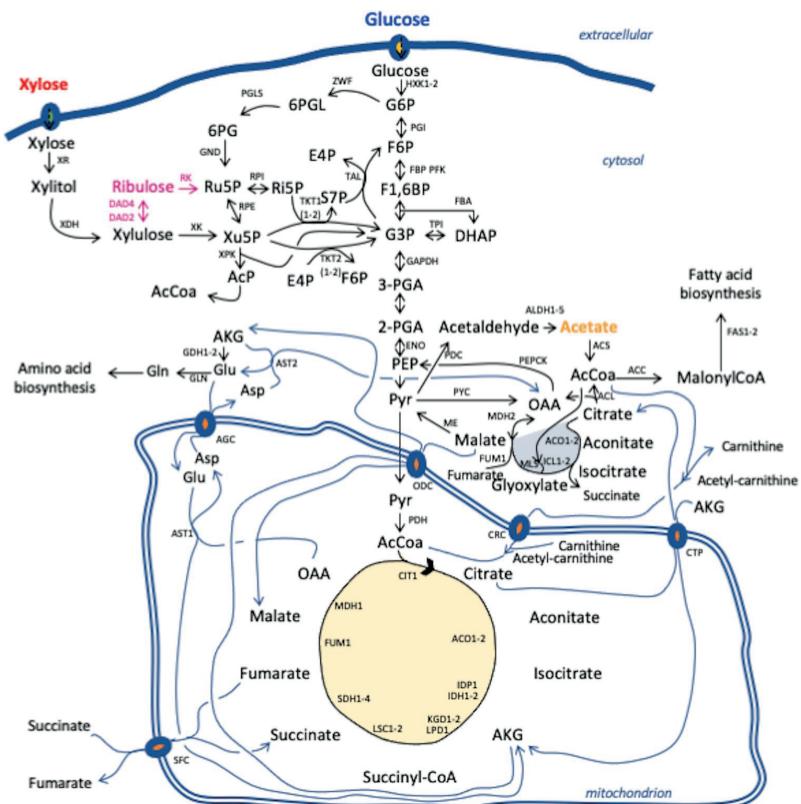


Fig 3. Main metabolic pathways present in *R. toruloides*. Blue arrows are used to denote mitochondrial carrier proteins and enzymes in shuttling pathways. Pink arrows are used to denote the alternative xylose assimilation pathway. Names, protein names and corresponding metabolic reaction IDs of genes are shown in [S2 Table](#).

<https://doi.org/10.1371/journal.pcbi.1011009.g003>

source of cytosolic acetyl-CoA during lipogenesis, which activity has never been reported on glucose, but is in line with previous findings in xylose condition [1,7]. The pyruvate decarboxylase and ACL, which exist as alternative pathways for producing cytosolic acetyl-CoA during lipid accumulation, were activated only when we blocked the XPK pathway ([S6 Dataset](#)). During the exp phase, cytosolic acetyl-CoA was not fully used for fatty acid biosynthesis, but 3% of carbon from glucose was transferred to TCA cycle via carnitine carrier (CRC) via acetylation reaction and in exchange of carnitine. The transfer of acetyl-CoA to mitochondria likely reflects that there was sufficient availability of cytoplasmic acetyl-CoA during the exp phase on glucose. At Nlim phase, majority of cytoplasmic acetyl-CoA was consumed by acetyl-CoA carboxylase (ACC), the first step in lipid biosynthesis, as more than 5-fold increase between 9% to 58% of carbon was observed via ACC during the transition from exp phase.

The main flux from the pyruvate branching point was channeled to the TCA cycle via pyruvate dehydrogenase (PDH), reaching 69% of carbon from glucose during exp phase. During exp phase, 37–46% of glucose-derived carbon was channeled from aconitase (ACO1-2) to fumarase (FUM1), while the flux through malate dehydrogenase (MDH1) appeared to

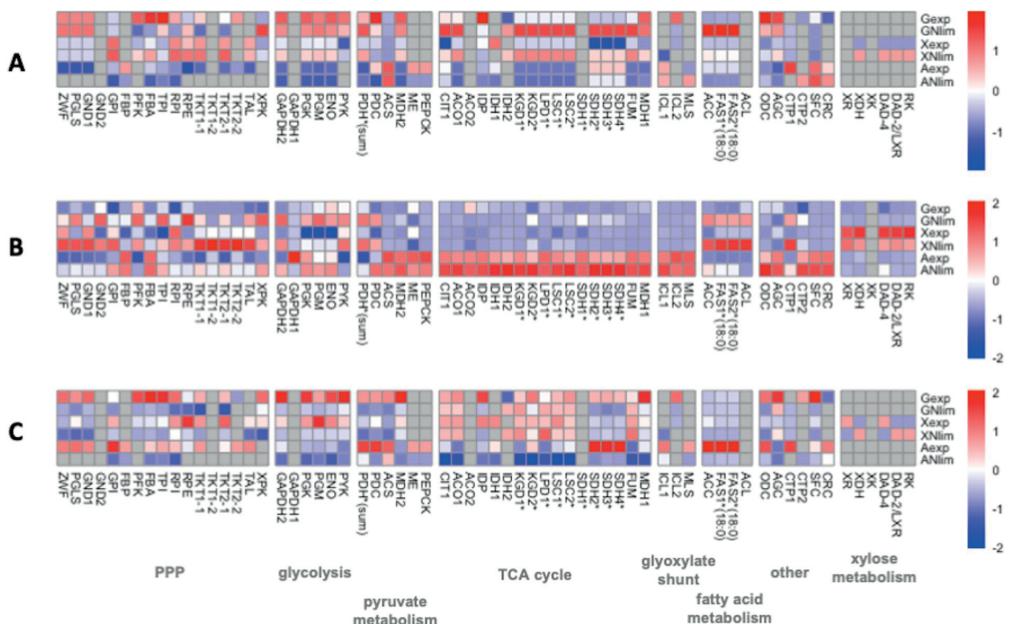


Fig 4. Heatmap of Z-scores for enzymes in central carbon metabolism. (A) Fluxes (mmol/gDCW/h) calculated as median from 2000 iterations of random sampling of the solution space [Bordel et al. 2010] [32], normalized by substrate uptake rate. (B) Proteomics data (µg/g of total protein) calculated as the average of duplicate experiments. (C) Apparent catalytic activities, k_{app} (s⁻¹), calculated as fluxes divided by protein concentrations. Gray color is used to denote missing values. Asterisks (*) are used to denote isoenzymes. PPP: pentose phosphate pathway. Names, protein names and corresponding metabolic reaction IDs of genes are shown in [S2 Table](#).

<https://doi.org/10.1371/journal.pcbi.1011009.g004>

correspond to 218% of glucose-derived carbon, indicating that an internal cycling of carbon was taking place. Similar recycling has been noticed also before in *R. toruloides* genome-scale models [1,7,11], however we assume this cycling to be artificial caused by a high demand of mitochondrial NADH. During the Nlim phase, ATP yield increased (**S8B Fig**), consistent with a significant upregulation of ETC on proteome level (**Fig 2B**). Since the TCA cycle and the respiratory chain are metabolically connected, cells require mitochondrial NADH to drive respiration. Ultimately, the yield of NADH during the Nlim phase slightly increased (**S9B Fig**). Fluxes through the TCA cycle significantly increased (**Fig 4A**), reaching 83% of carbon from glucose (**S6 Fig**), while the biomass yield decreased (**S1 Table**).

With respect to intracellular protein levels, the concentration of ACL was also high (1010 ± 75 $\mu\text{g/g_protein}$) at exp phase and was 2.6-fold upregulated (apval. 0.039) during the Nlim phase (**Fig 4B** and **S2 Dataset**), which was consistent with results from previous proteomics studies suggesting PDH-CIT1-ACL path for producing cytosolic acetyl-CoA [8,10], but not supported by the model flux results. The concentrations of PDH (2304 ± 11 $\mu\text{g/g_protein}$) and CIT1 (2573 ± 6 $\mu\text{g/g_protein}$) (**S10 Fig**) at exp phase were higher than on average for the TCA cycle enzymes (coinciding with high k_{cat} values of 486 s^{-1} and 540 s^{-1} for PDH and CIT1, **S3 Dataset**), which was also consistent with their role in the PDH-CIT1-ACL route, suggested by previous omics studies. However, the expression levels of mitochondrial membrane carrier proteins at both growth phases were low. At relatively low protein levels, high flux through these transporter proteins resulted in higher k_{app} values (**S11 Fig**).

Growth on xylose

Next, we explored *R. toruloides* metabolism during growth on xylose. Xylose is metabolized by xylose reductase (XR, NADPH-dependent), which reduces xylose to xylitol, xylitol dehydrogenase (XDH) and xylulokinase (XK), and further assimilated into central carbon metabolism via transketolase (TKT1-2) or XPK pathway. The expression of XK was not detected on proteome level in any of the conditions studied, suggesting an alternative pathway to the known fungal xylose assimilation pathway (Fig 3). The experimental detection of D-arabinitol isoform suggested the conversion of D-xylulose to D-arabinitol. This mechanism was supported by the presence of two genes in the *R. toruloides* genome encoding D-arabinitol dehydrogenase, RHTO_07844 and RHTO_07702. Only protein RHTO_07844 was detected in our proteomics analysis (1913 $\mu\text{g/g_protein}$) (S10 Fig), suggesting its role as D-arabinitol 4-dehydrogenase (DAD-4), converting D-xylulose to D-arabinitol. Arabinitol dehydrogenase could also convert arabinitol to ribulose (D-arabinitol 2-dehydrogenase) [34]. L-xylulose reductase (LXR) of fungal *A. monospora* has been reported to reversibly convert D-ribulose to D-arabinitol [35]. In support of this mechanism, protein levels of L-xylulose reductase (EC 1.1.1.10, RHTO_00373) were 10-fold upregulated during growth on xylose versus other substrates. Therefore, RHTO_00373 was selected as D-arabinitol 2-dehydrogenase (DAD-2) (converting D-arabinitol to D-ribulose). Arabinitol dehydrogenase is known to use NADH as cofactor [34]. LXR is mostly known for NADP(+)/NADPH specificity [35]. D-ribulose can enter the non-oxidative part of PPP via phosphorylation by D-ribulokinase (RK). An equivalent pathway was recently reported by Kim et al. 2021 [9]. One gene in *R. toruloides* IFO 0880 GEM (version 4.0) was annotated as D-ribulokinase (ID 14368) and we used it to identify potential RK in NP11 strain, which is more similar to the strain CCT 7815 used in this study [29]. Gene RHTO_00950 was identified as an ortholog of protein ID 14368 by a BLAST search, which found a match with 98.5% identity. Interestingly, in strain IFO 0880 orthologs of both genes RHTO_07844 and RHTO_07702 were identified as DAD-2 and DAD-4, respectively, and both were using NAD/NADH as the cofactor [9]. While it is not known, which cofactor of DAD-2/LXR enzyme is operational in *R. toruloides* strain CCT 7815, we analyzed the fluxes of both possible scenarios. Both simulation results were highly similar, with a difference in where NADPH was regenerated. The alternative pathway through DAD was preferred even when XK was not constrained with proteome.

In a scenario when DAD-2/LXR was NADP-dependent, during both exp and Nlim growth phases between 46–49% of carbon derived from xylose was directed via glucose 6-phosphate isomerase (GPI) in a reverse direction to the glycolytic flux. In a combination with that, 42% of carbon was directed via the oxPPP and returned to the Ru5P branching point, indicating that a loop associated with NADPH recycling is taking place. Alternatively, up to 88% of xylose-derived carbon was directed via oxPPP (S7 Dataset). In the first scenario, ZWF and GND provided more NADPH than LXR/DAD-2 during the exp phase (S12 Fig). During the Nlim phase, when the yield of NADPH slightly increased (S13 Fig), the flux of ZWF and GND remained unchanged, while the flux of LXR/DAD-2 increased (Fig 4A). XR consumed at least 2-fold more NADPH than any other NADP(+)-dependent enzyme during both growth phases. However during Nlim, more NADPH consumed by FAS1-2 was spent on lipid biosynthesis (S12 Fig).

From the proteomics analysis, the concentrations of enzymes involved in the xylose pathway were 1.1 to 1.6-fold downregulated during Nlim phase versus exp phase (Fig 4B), consistent with the decrease in xylose uptake rate (S1 Table). Lower concentration of RK (644 \pm 8 $\mu\text{g/g_protein}$) compared to other enzymes involved in xylose assimilation was measured (S10 Fig), suggesting enzyme limitation in the XK bypass pathway. At relatively low protein levels, high flux through RK resulted in relatively higher k_{app} values (S11 Fig).

Aside from enzymes directly involved in xylose assimilation, the intracellular flux patterns on xylose were the closest to growth on glucose, in comparison to growth on acetate. The flux of XPK was upregulated 1.7-fold between 13% to 22% during the Nlim phase ([Fig 4A](#)), which was also the main source of acetyl-CoA during lipogenesis. At the Nlim phase, the yields of ATP and NADH significantly increased ([S8](#) and [S9](#) Figs). The additional mitochondrial NADH during the Nlim phase was provided via internal cycling of MDH1 ([S14](#) Fig).

Growth on acetate

Lastly, we explored *R. toruloides* metabolism during growth on acetate. Acetate can cross the plasma membrane to enter the cells via simple or facilitated diffusion, but at pH below neutral (< pH6) the diffusion of the undissociated form of the acid induces the stress response or causes negative effect on metabolic activity [36]. In *R. toruloides*, two permeases have been found upregulated during growth on acetate-based rich medium in comparison to glucose-based rich medium [3], suggesting that facilitated diffusion is taking place. Once inside the cells, acetate is assimilated via ACS that directly provides acetyl-CoA ([Fig 3](#)), one of the main precursors for lipid biosynthesis. From acetyl-CoA branching point, the flux is channeled into the central metabolic pathways via isocitrate lyase (ICL1-2) and malate synthase (MLS), which are predicted to be located in cytosol, but no experimental evidence is available. Metabolic model predicted that at acetyl-CoA branching point, 18% of carbon from acetate during exp growth phase was directed to lipid biosynthesis via ACC, while the majority of carbon (51%) entered glyoxylate shunt. In addition, a significant amount of carbon from acetyl-CoA (29%) was channeled via CRC carrier, which was predicted to have a minor activity on glucose condition. The CRC route was preferred over the PDH pathway towards mitochondrial acetyl-CoA (MLS-ME-PDH), which channeled only 18% of carbon from acetate at exp phase. Metabolic model predicted 5% of carbon from acetate excreted as succinate from the glyoxylate shunt, in addition to 2% of carbon excreted as citrate, which was confirmed by HPLC. During the Nlim phase, the main fluxes demonstrated different regulation ([Fig 4A](#)). The increase in flux via CRC (1.4-fold) reflects that more carbon entered the TCA cycle during the Nlim phase. Interestingly, the flux of ACC was downregulated 3.1-fold at Nlim compared to the exp growth phase. Using the rate of lipid production, which decreased during Nlim phase, it can be explained that the lipid production in absolute amounts was higher during the exp phase to sustain the growth together with moderate lipid production ([S1](#) Table).

On acetate, fluxes of the TCA cycle were the lowest, while measured protein levels were the highest among all conditions analyzed ([Fig 4A](#) and [4B](#)). At Nlim phase, 28% of carbon from acetate was predicted to be excreted as OAA, while the biomass yield decreased ([S1](#) Table). Flux levels of the TCA cycle indicated that an internal cycling of carbon similar as in other conditions was taking place ([S6](#) Fig). It involved different transporter proteins,—the citrate-oxoglutarate (CTP) and succinate-fumarate (SFC) transport —, which allow channeling of the flux from the TCA cycle to glyoxylate shunt ([Fig 3](#)). During the Nlim phase, ATP turnover, produced entirely via ETC ([S15](#) Fig), and NADH turnover, produced almost entirely via the TCA cycle ([S16](#) Fig), both decreased ([S8](#) and [S9](#) Figs), unlike observed in glucose or xylose conditions, where ca 80% of the ATP originated from ETC, while the rest came mainly from glycolysis.

Aside from enzymes directly involved in the TCA cycle, cytosolic ME was the sole supplier of NADPH during the Nlim phase only in acetate condition ([S17](#) Fig). This is supported also by the measured protein levels of ME, which were significantly higher under acetate conditions, although the absolute levels of ME were relatively low under all studied conditions (189 ±3 µg/g_protein) ([Figs 4B](#) and [S6](#)). Only during the growth on acetate the NADPH yield

decreased during the Nlim phase ([S13 Fig](#)). During the Nlim phase, more NADPH was consumed by FAS and spent on lipid biosynthesis ([S17 Fig](#)).

We also observed few significant changes in fluxes of enzymes involved in gluconeogenesis, which is an important pathway during growth on acetate to provide xylose phosphate-based precursors for ribonucleotide synthesis. The normalized flux towards gluconeogenesis, channeled via MDH2, carried 11% of carbon exp growth phase. It may reflect the fact that PEPCK, the first enzyme in the gluconeogenesis pathway, consumed ATP, but we found that PEPCK was consuming only 2.4% of ATP during exp phase ([S15 Fig](#)). The concentration of PEPCK (627 $\mu\text{g/g_protein}$) ([S10 Fig](#)) and its k_{cat} value (38 s^{-1}) ([S2 Dataset](#)) were low, suggesting that PEPCK could have been a rate-limiting step of gluconeogenesis during the exp phase.

From proteomics analysis, the concentration of enzymes involved in fatty acid beta oxidation (RHTO_04957, RHTO_00300, RHTO_02848, RHTO_07118, RHTO_00476) at higher concentrations (from 163 ± 9 to $531 \pm 7 \mu\text{g/g_protein}$) as compared to cells grown on other substrates at exp phase ([S2 Dataset](#)), suggesting this pathway might be more active in *R. toruloides* during growth on acetate.

Intracellular flux patterns point to metabolic trade-offs associated with lipid production

Multi-layer data provided in this study allows us to analyze metabolic trade-offs and compare the resource allocation between different metabolic pathways present in *R. toruloides*. From the metabolic modeling results, we analyzed the NADPH allocation between nitrogen assimilation and fatty acid biosynthetic pathways in glucose- and xylose-grown cells ([Fig 5](#)). When the yeast had an abundant source of nitrogen, either by conversion of urea to ammonia (glucose condition) or by growth on ammonia itself, NADP-dependent glutamate dehydrogenase (GDH1) converted ammonia with the TCA cycle intermediate AKG into glutamate, which was then used for the amino acid biosynthesis. Thus, during the exp phase on glucose-grown cells 46% of NADPH turnover was channeled via GDH1, while 13% was consumed via FAS1-2. On the contrary, during the Nlim phase 12% of NADPH was channeled via GDH1, while 75% was consumed via FAS1-2. This was supported also by the measured protein levels of GDH1, which were significantly higher during exp growth phase ($4294 \pm 183 \mu\text{g/g_protein}$), as compared to Nlim phase ($2948 \pm 135 \mu\text{g/g_protein}$) ([S2 Dataset](#)). In this metabolic trade-off, less (almost 4-fold) cytosolic NADPH was consumed by GDH1 during the Nlim phase ([S7 Fig](#)) when the protein content reduced 2.8-fold ([S1 Table](#)). And vice versa, more NADPH (almost 6-fold) was consumed by FAS1-2 when total lipids increased 8.5-fold.

In acetate condition, the metabolism during lipogenesis at Nlim phase might be influenced by the beta-oxidation, a metabolic process of lipid degradation that can return carbon back into central metabolism, which was detected on proteome level in acetate condition. But the activity of this pathway could not be simulated with our current model.

Discussion

In this study, we presented detailed analysis of physiological characterization of *R. toruloides* CCT 7815 during growth on glucose, xylose or acetate as a sole carbon source. It was an important part of the study as the collected data together with the quantitative proteomics analysis was used to constrain the newly developed enzyme-constrained metabolic models. Cultivation experiments were carried out at a C/N ratio, which allowed nutrient excess conditions during the first part of the batch cultivation and resulted in nitrogen limitation during the second part of the experiment, a growth phase known to induce lipid accumulation [[12](#)]. Enabled by bioreactors equipped with online monitoring sensors, we were able to accurately

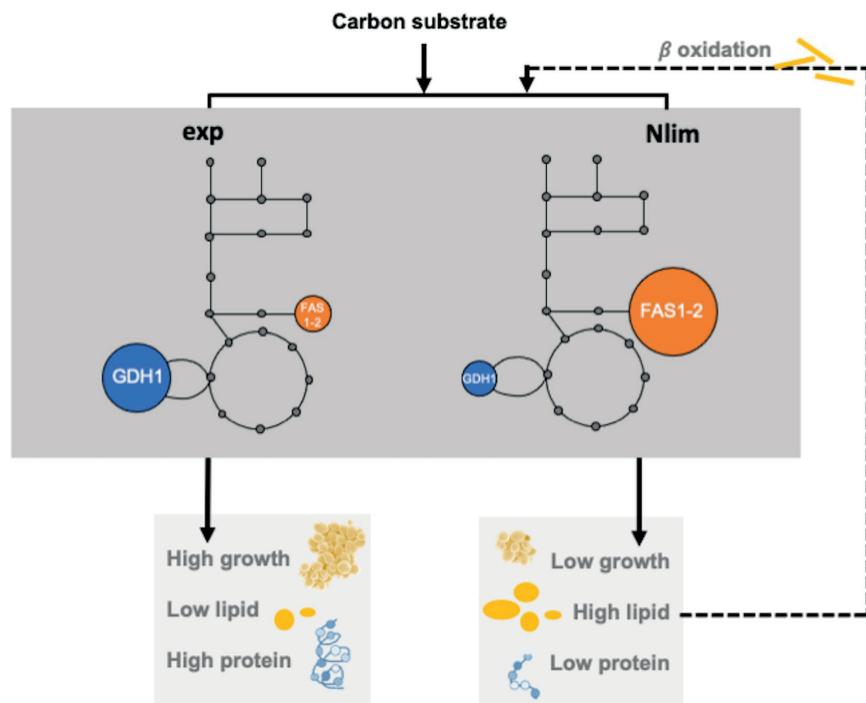


Fig 5. The trade-off between NADPH expenditure in *R. toruloides*. Blue circles represent nitrogen assimilation and orange circles represent lipid biosynthetic pathways of *R. toruloides* cultivated on glucose during exponential growth (exp) and nitrogen limitation phase (Nlim). Circle size (blue and orange) represents the % of NADPH turnover. GDH1, glutamate dehydrogenase (NADP+); FAS1-2, fatty acid synthase (α and β).

<https://doi.org/10.1371/journal.pcbi.1011009.g005>

measure two distinct growth phases during the batch culture (S1 Fig), characterized by different lipid content in biomass and other physiological parameters (S1 Table), which were later implemented into metabolic models.

Experimental results showed a slightly higher final biomass yield (0.32 gDCW/g_substrate) during growth on xylose as compared to Tiukova et al. 2019, [8] (0.28 gDCW/g_substrate), where significantly lower starting sugar concentration (40 g/L) and a different *R. toruloides* strain were used. Differences between strains were demonstrated also in their lipid composition (Fig 1E). Lower final biomass yield during growth on glucose (0.24 g/g) as compared to xylose condition (0.32 g/g) was difficult to explain by any other reason than the formation of cell aggregates on glucose, as no byproducts in this condition were detected. It has been reported earlier, and for future research aggregate formation in *R. toruloides* could be mitigated with the increased salt concentrations [37]. The possible explanation could be the production of exopolysaccharides that has been found in other *Rhodotorula* species [4]. However, the final lipid content in biomass during growth on glucose (48%) was the same as what has been reported in Tiukova et al. 2019, [8].

Absolute proteome quantification helped to improve the understanding of metabolism during lipid accumulation and on various substrates. Interestingly, we found that the proteome was largely unchanged during the Nlim versus exp phase on acetate, while up to 204 differently

regulated proteins were found on other tested substrates. Low proteome allocation to TCA cycle as compared to glycolytic metabolic pathways in the presence of glucose and xylose (Fig 2B) could possibly signal about citrate accumulation and subsequent transportation event from mitochondria to cytoplasm developed for lipid production.

Metabolic modeling confirmed previous modeling findings obtained on PPP pathway and ME as the main suppliers of NADPH, and XPK pathway as the primary source of acetyl-CoA during lipid biosynthesis in *R. toruloides* [1,7]. We found that these are carbon source dependent and to close proximity of the carbon substrate uptake. We also found that the pathways involved in synthesis of lipid precursors were not changed during the Nlim versus exp growth phase. To enable this analysis, we modified models' biomass reactions to reflect the measured lipid and protein content, based on precise physiology data of two distinct growth phases. For further work, it would be interesting to understand if the predicted fluxes via oxPPP while cells were growing on glucose and xylose are thermodynamically feasible.

Metabolic modeling also helped to explain *R. toruloides* physiological characteristics and byproduct excretion. Changes in biomass yield during growth on glucose were associated with increased fluxes through phosphoketolase (XPK) pathway and the TCA cycle, resulting in more carbon entering the TCA cycle during lipid accumulation. Predicted acetate kinase activity demonstrated certain robustness towards the preference for the XPK pathway (S5 Dataset). XPK pathway is considered an attractive option for generating cytosolic acetyl-CoA because it is more energy and carbon efficient. It circumvents one molecule of CO_2 lost per pyruvate and two equivalents of ATP consumed compared to the PDH bypass route (PDC-ALDH and ACS). In earlier studies using heterologous expression of XPK pathway in *S. cerevisiae* [38,39] it was found that the engineered strain had an increased flux towards TCA cycle and lower flux from the pyruvate branching point towards acetate formation. In a study by Bergman et al. 2019, [40], it was found that the activity of XPK pathway increases acetate flux and ATP requirement in *S. cerevisiae*, leading to an increased production of CO_2 and negative growth effects. Apart from the fact that there is no acetate excretion in *R. toruloides*, it would be interesting to further understand if use of XPK pathway in lipogenesis may also explain the carbon losses on glucose as observed in our study.

During growth on acetate, it was suggested that the byproduct formation was associated with energy metabolism, as the predicted excreted metabolites were TCA cycle-related. Using the metabolic models, we predicted a higher ATP turnover in acetate as compared to xylose condition (S8 Fig), while the specific growth rates were comparable. Our simulation results could not explain why Nlim phase, when more carbon entered the TCA cycle, was associated with the increase in internal cycling for NADH transport, also known as malate-aspartate NADH shuttle [41]. Stoichiometry in the oxidative phosphorylation pathway in *R. toruloides* in rhto-GEM has been adopted from experimental data in *S. cerevisiae* and was not changed in the present study. Moreover, rhto-GEM and ecRhtoGEM are based on *R. toruloides* exhibiting not only proton-pumping complex I (t_0001), but also external NADH dehydrogenase (r_0770) that do not pump protons. However, presence of complex I creates an electron competition process, which might have consequences on the yield of oxidative phosphorylation, as experimentally demonstrated in another Crabtree negative yeast *Candida utilis* [42]. As no experimental phosphate/oxygen (P/O) ratio has been yet reported in *R. toruloides*, it might be that the mitochondrial shuttling loop observed in our simulations in reality could correspond to a higher ATP requirement.

Internal carbon cycling was suggested not only in acetate, but also in glucose and xylose conditions. This mitochondrial shuttling loop was also noticeable from the results of earlier modeling studies [1,7,11]. In Dihm et al. 2019, [11], it was called “NADH shuttle”, but it is yet to be confirmed if the shuttle could be an artificial loop to feed NADH to mitochondria.

Despite progress in understanding the physiology and genetics of *R. toruloides*, very little is known about mitochondrial carrier (MC) proteins in this strain. From experimental studies in *S. cerevisiae*, oxodicarboxylate (ODC) and aspartate/glutamate (AGC) carriers are important to export AKG (in exchange of malate) for nitrogen assimilation and also for the malate-aspartate NADH shuttle [43]. In support of this mechanism, our simulations on glucose-grown cells showed that the NADH produced via glycolysis was transferred to mitochondria for electron transport using enzymes MDH2 and ODC. Alternatively, it might point to an artifact of a different P/O ratio.

Metabolic models on xylose were greatly improved by the detection of chirality of D-arabinitol. We presented an alternative xylose assimilation pathway, which was favored in our model simulations over the known xylose pathway in fungi that involved D-xylulose 5-phosphate. Our results were consistent with recent reports in strain IFO 0880 [9], but we also detected some differences in the pathway, which were related to the fact that we used a different *R. toruloides* strain, CCT 7815. To explain byproduct formation upstream glycolysis while cells were growing on xylose, we presented several ideas associated with the energy and lipid metabolism. In comparison to our previous work [1], the flux of PPP could be compensated by the amount of carbon channeled via the alternative xylose pathway.

The fact that no carbon was directed via PDH-CIT-ACL pathway might also point to lack of alternative routes of NADPH regeneration in *R. toruloides*. In our results, the activity of oxPPP was coupled to an active XPK pathway supplying the vast majority of cytosolic acetyl-CoA during the Nlim phase. Our proteomics data showed a significant increase in the amount of uncharacterized proteins during lipid accumulation, especially in xylose condition (Fig 2B). Hypothetically, CO₂-decoupled NADPH regeneration would reduce the fluxes through oxPPP and XPK pathways (but not eliminate them), the flux of glycolysis would remain the same as shown in this study, but more carbon would be channeled via IDP and ACL. It has been demonstrated that GAPDH contributes to NADPH supply in filamentous fungi *Mortierella alpina* [44]. CO₂-decoupled NADPH synthesis has been engineered in *S. cerevisiae*, demonstrating significant phenotypic changes [45].

Modeling results revealed metabolic trade-offs associated with NADPH allocation between nitrogen assimilation and lipid biosynthetic pathways. In *S. cerevisiae* and *E. coli*, a clear specific growth rate dependence of ribosomal proteins has been demonstrated [27,46,47]. In the present study, we demonstrated a similar significant correlation for specific growth rate dependent ribosomal content (Fig 2C). Moreover, we were able to demonstrate a trade-off in NADPH demand. Although NADPH regeneration was dependent on the carbon source, NADPH demand was shifting from protein production at higher growth rates to lipid biosynthesis at lower growth rates in Nlim phase (Fig 5). Enzyme-constrained metabolic models developed in this study used not fully matched k_{cat} values that can notably increase the prediction errors [48]. Characterizing enzymatic properties using physically based models requires enormous experimental work, therefore accurate computational approaches are needed to address this gap. Deep learning algorithms have demonstrated outstanding success in predicting protein structures based on their sequence information [49–51], and the method has also been applied in predicting enzyme k_{cat} values for yeast *S. cerevisiae* [52].

Conclusion

In this study, enzyme-constrained genome-scale metabolic models were generated for *R. toruloides*, where metabolic modeling together with proteome data gave a detailed interpretation of how flux patterns are changing in *R. toruloides* on different substrates during the exponential growth and in lipid accumulation. The results were consistent with previous knowledge on

the main pathways involved in lipid biosynthesis in *R. toruloides*, revealed by genome-scale modeling and multi-omics analyses. While detailed analysis of simulated intracellular flux patterns allowed us to explain some physiological parameters during growth on glucose, many observations require further validation. This work contributes to improving the knowledge about *R. toruloides* metabolism.

Materials and methods

Strain, media and growth conditions

R. toruloides CCT 7815 (Coleção de Culturas Tropicais, Fundação André Tosello, Campinas, Brazil) from a previous study [21] was used in the cultivation experiments. The same study identified increased lipid production, induction of hydrolysate-tolerance and lipid accumulation genes without physiological changes regarding growth and substrate consumption in *R. toruloides* strain CCT 7815 after a short-term adaptation in sugarcane bagasse hemicellulosic hydrolyzate. Seed cultures were grown on chemically defined medium according to Verduyn (3.0 g/L KH₂PO₄, 0.5 g/L MgSO₄·7H₂O, 15 mg/L EDTA, 4.5 mg/L ZnSO₄·7H₂O, 0.3 mg/L CoCl₂·6H₂O, 1 mg/L MnCl₂·4H₂O, 0.3 mg/L CuSO₄·5H₂O, 4.5 mg/L CaCl₂·2H₂O, 3 mg/L FeSO₄·7H₂O, 0.4 mg/L Na₂MoO₄·2H₂O, 1 mg/L H₃BO₃, 0.1 mg/L KI, 0.05 mg/L biotin, 1 mg/L calcium pantothenate, 1 mg/L nicotinic acid, 25 mg/L inositol, 1 mg/L thiamine HCl, 1 mg/L pyridoxine HCl, 0.2 mg/L *para*-aminobenzoic acid [53]) supplemented with a sole carbon source of 18.2 g/L glucose, 20 g/L xylose or 20.0 g/L acetic acid and 5 g/L (NH₄)₂SO₄ in duplicate shake flasks at 200 rpm and 30°C for 24 h. The carbon/nitrogen (C/N) molar ratio of the medium in seed cultures was 8.8. To obtain seed cultures, cells were pre-cultured in YPD media, and pelleted and washed twice with 0.9% (m/v) NaCl solution before inoculation. Seed cultures were used to inoculate 900 mL of chemically defined medium supplemented with either 63.6 g/L glucose and 0.9 g/L urea, or 70 g/L xylose and 2 g/L (NH₄)₂SO₄ or 20.0 g/L acetic acid and 0.6 g/L (NH₄)₂SO₄, and 0.1 mL/L antifoam 204 (Sigma-Aldrich, St. Louis, MO, United States) in duplicate bioreactors with a starting OD₆₀₀ of 0.4 at 400–600 rpm, 30°C, pH 6.0. At the start of cultivation, the (C/N) molar ratio of the media in bioreactors was set to 69 (glucose/urea) and 80 (xylose- or acetate/(NH₄)₂SO₄). Note, that xylose condition was carried out in Pinheiro et al. 2020, [1].

Cells were grown in 1-L bioreactors (Applikon Biotechnology, Delft, the Netherlands) in a batch cultivation regime. pH was controlled by the addition of 2 mol/L KOH. Dissolved oxygen was maintained not lower than 25% at 1-vvm airflow by regulating the stirring speed. CO₂ and O₂ outflow gas composition were measured using an online gas analyzer (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).

For dry cell weight measurement, samples were collected every 6 hours during the exponential growth phase and every 24 hours during the nitrogen limitation phase. For other analyses, samples were collected every 3 hours during the exponential growth phase and every 24 or 48 hours during the nitrogen limitation phase.

For extracellular metabolites, lipidomics and protein content analyses, samples were taken from bioreactors to 2-mL tubes, centrifuged 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 further analyses.

Analytical methods

For dry cell weight (DCW) measurement, culture samples were taken from bioreactors to 2-mL tubes, passed through a 0.3 μ m filter, dried and analyzed by gravimetric method. Biomass optical density data were calibrated by gravimetric cell mass measurements. For extracellular metabolites measurements, high-performance liquid chromatography (HPLC) separations were performed with Shimadzu instruments (LC-2030C Plus, Shimadzu, Kyoto, Japan) equipped with a refractive index detector (RID-20A, Shimadzu, Kyoto, Japan). Glucose, xylose, organic acids and glycerol concentrations were measured using a Rezex ROA Organic Acid column (Phenomenex, Torrance, United States). Separations were performed at 45°C and the mobile phase for isocratic elution was 5 mmol/L H₂SO₄. The flow rate was 0.6 mL/min. Stereoselective HPLC analysis of arabinitol isomers was done using a Chiralpak column (Daicel Technologies, Japan) and the mobile phase for isocratic elution was a mixture of hexane and ethanol (70:30, v/v) at 20°C; the flow was 0.3 mL/min. Chiralpak column of arabinitol standards gave different retention times for each enantiomer (D and L) ([S18 Fig](#)). Yields and specific consumption and production rates represent exp and Nlim phases separately, not cumulatively.

For intracellular total protein quantification, cell pellets were thawed on ice and resuspended in 0.9% (m/v) NaCl solution to a concentration of 1 g/L. Then 600 μ g of biomass was mixed with a commercially available protein extraction solution (Y-PER, Thermo Fisher) in a 2-mL tube and incubated at 30°C for 45 minutes. After incubation, samples were transferred to screw cap 2-mL tubes with glass beads. Cell lysis was performed using a FastPrep-24 device for 4 cycles (4 m/s for 20 s) with a 5 min interval after each cycle. After cell lysis, the tubes were centrifuged at 14800 rpm for 10 min at 4°C. Supernatant was collected to a new 2-mL tube and the leftover biomass sample was subjected to a repeated extraction cycle (without 45 min incubation interval) until no proteome was detected in supernatant. Before quantification, all fractions of supernatant were combined. Proteome was quantified using a commercially available colorimetric assay kit (Micro BCA Protein Assay Kit, Thermo Fisher Scientific, Waltham, MA, United States). Protein concentration was determined using the calibration curve of bovine serum albumin (BSA) standard of linear range dilutions from 0.5 to 200 μ g/mL. Assay was performed in triplicate for each sample. Samples chosen for analysis corresponded to 17 and 57 h in glucose, 48 and 72 h in xylose, and 26 and 44 h in acetate. Assay results represent cumulative proteome during each growth phase of yeast.

Lipidomics

To quantify lipids and determine their fatty acid composition, quantitative gas chromatography–mass spectrometry (GC-MS) analysis with the internal standard method was used, similar as described in Tammekivi et al. 2019, [\[54\]](#). Before analysis, cell pellets were lyophilized and derivatized by using acid-catalyzed methylation. This derivatization procedure produces methylated fatty acids from both free and bonded fatty acids. The quantitative analysis and derivatization procedure of the TAGs and free fatty acids was based on Tammekivi et al. 2021, [\[55\]](#). From the lyophilized cells, 10–12 mg of was weighed into a 15 mL glass vial. An analytical balance (Precisa, Dietikon, Switzerland, resolution of 0.01 mg) was used to weigh all components that influence the quantitative analysis (samples, solvents, internal standard). To the cells, 2 mL of MeOH (\geq 99.9%, Honeywell, Charlotte, NC, USA) was added and the vial was sonicated for 15 min. Then, 0.4 mL of conc. H₂SO₄ (98%, VWR Chemicals, Radnor, PA, USA) was carefully added to the solution and the derivatization mixture was heated for 3 h at 80°C. After, the mixture was extracted 3 x 2 mL with hexane (\geq 97.0%, Honeywell) and the extracts were pipetted through a layer of K₂CO₃ (99.5%, Reakhim) on top of a glass wool (Supelco,

Bellefonte, PA, United States) layer. The combined extracts were evaporated to dryness and the residue was redissolved in 2 mL of toluene ($\geq 99.9\%$, Honeywell). Depending on the expected fatty acid concentration, toluene and internal standard (hexadecane, $\geq 99\%$, Honeywell) were added so that the results would stay in the range of the calibration curve.

The solutions containing the fatty acid methyl esters and internal standard were analyzed with an Agilent (Santa Clara, CA, USA) 7890A GC connected to an Agilent 5975C inert XL mass spectrometric detector (MSD) with a triple-axis detector and an Agilent G4513A autosampler. The column was an Agilent DB-225MS capillary column (30 m x 0.25 mm diameter, 0.25 μm film thickness) with a (50%-cyanopropylphenyl)-methylpolysiloxane stationary phase. The temperatures of the mass spectrometer transfer line and ion source were 280°C and 230°C, respectively. The temperature of the inlet was 300°C, injection volume 0.5 μL , and splitless mode was used, where the split was opened after 2 min. The oven's temperature program was the following: isothermal for 2 min at 80°C, increased 20°C/min to 200°C, isothermal for 4 min, increased 5°C/min to 220°C, isothermal for 5 min, increased 10°C/min to 230°C, isothermal for 12 min. The total run time was 34 min. Electron ionization (EI) with 70 eV was used and the solvent delay was 5.6 min. Helium 6.0 was used as the carrier gas (flow rate 1.5 mL/min). Qualitative analysis was performed in the scan mode (mass range of 27–400 m/z) and quantitative analysis was performed in the selected ion monitoring (SIM) mode, which were both measured during the same GC-MS run. For data analysis, Agilent MSD ChemStation and NIST Mass Spectral Library Search 2.0 were used.

Commercial standard mixture of fatty acid methyl esters (FAME, C8–C24, Supelco) was used to confirm the identity (based on retention times, in addition to the mass spectral comparison) and to quantify the fatty acids. Seven calibration solutions were made from the FAME mixture in toluene and the same internal standard (hexadecane) was added. All calibration solutions were measured in random order in the same GC-MS sequence with the derivatized sample solutions. For each methylated fatty acid, a calibration curve was constructed based on the data obtained from the GC-MS analysis of the calibration solutions— $S_{\text{AD}}/S_{\text{IS}}$ vs. $C_{\text{AD}}/C_{\text{IS}}$ —where S represents the peak area, C the concentration, AD the derivatized fatty acid, and IS the internal standard. Knowing the $S_{\text{AD}}/S_{\text{IS}}$ and C_{IS} for the sample solution, if was possible to calculate the derivatized fatty acid concentration (C_{AD}). Finally, the obtained value was recalculated to represent the concentrations of particular fatty acids or homotriglycerides. Also the derivatization efficiency (for more information see Tammekivi et al. 2019, [51]) was taken into account by applying the same derivatization procedure and quantitative analysis for the analysis of five fatty acid standards (C16:0, C18:0, C18:1, C18:2 and C18:3) and their corresponding TAG standards. The obtained yield (%) from the weighed quantity of the corresponding standard) was used to correct the result of the sample analysis. The sum of the quantified TAGs was presented as the total lipid content. Samples chosen for analysis corresponded to 24, 52 and 100 h in glucose, 48 and 96 h in xylose, and 39 and 84 h in acetate. Analysis results represent cumulative lipidome during each growth phase of the yeast.

Experimental procedure for absolute proteomics

Absolute proteome quantification was performed using a nanoscale liquid chromatography with tandem mass spectrometry (Nano-LC/MS/MS), similar as described in Sanchez et al. 2021, [26]. Experimental procedure for cell lysis and sample preparation was done as described in the same study. Briefly, cell pellets were lysed using a pH 8.0 buffer (6 M guanidine HCl, 100 mM Tris-HCl, 20 mM dithiothreitol) and homogenized using the FastPrep-24 device (2x at 4 m/s for 30 s). After centrifugation and overnight precipitation (10% trichloroacetic acid, at 4°C), protein concentration was measured as described above in the total protein content

section. For absolute quantification, proteome samples were mixed heavy-labeled *R. toruloides* grown in previously described minimal medium supplemented with heavy 15N, 13C-lysine (Silantes, Munich, Germany), which was used as an internal standard [1]. Further sample preparation and Nano-LC/MS/MS analysis was similar to previous descriptions [26]. Samples chosen for analysis corresponded to 17 and 57 h in glucose, 48 and 72 h in xylose, and 26 and 44 h in acetate, same as in the total proteome analysis.

Proteomics data analysis

The raw data obtained from the Nano-LC/MS/MS analysis was processed using MaxQuant v1.6.1.0 software package [56] with similar settings as described in Sanchez et al. 2021, [26]. Data search was performed against the Uniprot (www.uniprot.org) *R. toruloides* NP11 reference proteome database [10]. Raw data quantification was similar to previous descriptions, except that the MS intensities were normalized with the average internal standard abundance (reverse Ratio H/L normalized). MS intensities were calculated from the internal standard abundance using the number of theoretically observable peptides (iBAQ, intensity Based Absolute Quantification; iBAQ H) feature in MaxQuant, the reverse Ratio H/L normalized of the sample, and reverse Ratio H/L. The resulting MS intensities were adjusted for 80% recovery of the sample injected. Finally, absolute protein concentrations were derived from the normalized sum of MS intensities assuming its proportionality to the measured total protein content, also known as the total protein approach [26].

LC-MS/MS data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [57] with the dataset identifier PXD037281. Processed quantitative data are presented in [S1 Dataset](#). Duplicate experiments were used in differential expression analysis. *p*-values were adjusted for multiple comparisons using Benjamini-Hochberg (1995) method [58].

Enzyme-constrained model reconstruction

Enzyme-constrained genome-scale metabolic model of *R. toruloides* was generated using the metabolic network *rhto-GEM* version 1.3.0 [15]. The workflow was based on a semi-automatic algorithm of the GECKO toolbox version 2.0.2 [17] operating on MATLAB (The MathWorks Inc., Natick, MA, United States). Model development was tracked on a dedicated Github repository: <https://github.com/alinarekena/ecRhtoGEM/>.

Firstly, functions *addMets*, *addGenesRaven* and *addRxns* from RAVEN [59] were used to add the alternative xylose assimilation pathway to *rhto-GEM*, as provided in *ecRhtoGEM/edit_rhtoGEM*. Next, pipelines *enhanceGEM* and *generate_protModels* from the GECKO Toolbox were used to generate ec-models, as provided in *ecRhtoGEM/reconstruct_ecRhtoGEM*.

During the *enhanceGEM* pipeline, enzyme kinetic parameters were relaxed to overcome model over constraint using the *manualModifications* function from the GECKO Toolbox. The enzymes subject to manual k_{cat} value curation were identified by running *enhanceGEM* pipeline initially with the physiology data of the xylose condition, as provided in *ecRhtoGEM/customGECKO/getModelParameters*. The *relative_proteomics.txt* and *uniprot.tab* input data were used to match enzymes with the model and retrieve their k_{cat} values from the BRENDA database. The data for *uniprot.tab* were downloaded from Uniprot.org with *R. toruloides* strain NP11 as query, while *relative_proteomics.txt* contained average protein abundances of enzymes detected in our proteomics analysis (in mmol/gDCW). The *getModelParameters* function was used to ensure that the newly generated ec-model was constrained with experimental data. The GECKO Toolbox automatically performed the initial sensitivity analysis on

the objective function (ie. maximize cell growth) with respect to the individual k_{cat} values by identifying the top limiting value and by iteratively replacing it with the maximum value available in BRENDA. According to the reported information, we adjusted k_{cat} values identified as limiting to reasonably higher values found in literature (for a detailed description see ecRhtoGEM/manualModifications). In the next step, we used the *topUsedEnzymes* function from GECKO Toolbox to calculate the top ten enzyme usages in a mass-wise way (data not provided). Similarly as in the previous step, k_{cat} values of enzymes identified among top used in each condition were increased to reasonably higher values referenced in the literature. In the script the procedure was named round A. Later, in so called round B, the *topUsedEnzymes* function was applied to the same conditions again and k_{cat} values were modified until enzyme usage represented less than 1% of total protein pool, as provided in ecRhtoGEM/manualModifications. The final list of modified k_{cat} values included 27 enzymes, as summarized in [S6 Table](#). As automatic k_{cat} values were derived from studies that involved not the same organism and substrate, their values were often very low. For example, the k_{cat} value of fructose-bisphosphatase was increased from 0.002 s^{-1} to 127 s^{-1} , on the basis of specific enzyme activity for the same EC number.

During the *generate_protModels* pipeline, growth- and non-growth-associated energy requirements were fit using measured substrate uptake and gas rates from batch cultivations of *R. toruloides* obtained in this study, as provided in ecRhtoGEM/customGECKO/fermentation-Data. They were set from 124.4 to 140.0 mmol/gDCW and from 0 to 3.65 mmol/(gDCW/h). Coefficients in oxidative phosphorylation from rhto-GEM were not changed. Polymerization costs from the study in *S. cerevisiae* [60] were used, similarly as in *rhto-GEM*. Average enzyme saturation factor (sigma) was fit to physiological parameters (ecRhtoGEM/results/enhance-GEM_pipeline/sigma), and set at 0.35 in ecRhtoGEM/customGECKO/getModelParameters. Biomass composition was modified from rhto-GEM to include *R. toruloides* CCT 7815 protein content, lipid content and acyl chain profiles, as provided in ecRhtoGEM/customGECKO and ecRhtoGEM/data, respectively. The *scaleLipidProtein* and *scaleLipidsRhto* functions from GECKO Toolbox and SLIMER [61] were modified for the *generate_protModels* pipeline. To avoid the model to over constrain, automatic flexibilization was performed on concentrations of 7 (XNlim) to 25 (Gexp) enzymes, as listed in [S6 Table](#) (the old and new values are available at ecRhtoGEM/results/generate_prot_Models_pipeline/modifiedEnzymes.txt). An alternative approach to calculate the abundance of those enzymes for which no enzyme level had been measured was used as additional modification in addition to previously described modifications in the pipeline to handle the low protein levels observed in Nlim conditions. In this approach, we directly subtracted the measured enzyme concentration (Pmeasured) from the total enzyme concentration (enzymeConc) to obtain the unmeasured enzyme concentration (PpoolMass). Modifications to original approach, by which GECKO adjusts for the unmeasured enzyme concentration, are available from *generate_protModels* and *constrainEnzymes* functions at 'customGECKO' folder. These included sample specific f-factor calculation was moved before filtering proteomics data (*generate_protModels*). This ensured higher coverage, while not largely affecting f calculation. Total protein content (Ptot) calculation was rescaled by adding standard deviation and flexibilization because of too low measurement. Then f, which was calculated in the beginning of *generate_protModels*, and rescaled updated Ptot were used to calculate expected total enzyme concentration (enzymeConc), as provided in *constrainEnzymes*. This ensured higher coverage, likely critical in low total protein content biomass (in case of all Nlim phases). Other updates included rescaling of enzyme usage to prevent very low fluxes, as provided in *generate_protModels*.

Model calculations

Flux balance analysis was performed with the RAVEN toolbox using Gurobi solver (Gurobi Optimization Inc., Houston, Texas, United States). Flux variability was estimated with random sampling of the solution space with 2000 sampling iterations for each condition (ie., ec-model). For each sample, a random set of three reactions was given random weights and the sum of these were parsimoniously maximized to explore the constraint solution space [32], considering 1% variability from maximal growth rate and substrate uptake rate, 10% variability from predicted carbon dioxide production and oxygen consumption rate, 10% variability from measured by-product rates, 10% variability from protein pool, and 1% variability from NGAM, as specified in *analyze_ecRhtoGEM*. In glucose condition, simulated values were used to constrain gas exchange (carbon dioxide and oxygen) due to measurement problems in experimental values. In xylose condition, measured values were used to constrain the production of by-products xylitol and D-arabinitol. To allow the model to use either traditional or alternative xylose assimilation pathway, xylulokinase (XK) was not blocked, but eventually constrained with enzyme constraints from the protein pool. In acetate condition, measured values were used to constrain the production of citrate. Flux value was calculated as a median of 2000 sampling iterations. Flux variability was represented as SD divided by flux, multiplied by 100. Finally, fluxes were converted to base GEM formalism using *mapRxnsToOriginal* function from Domenzain et al. 2022 [30]. For the analysis, fluxes were normalized by dividing absolute flux with the specific substrate uptake rate to ensure the comparability among different conditions. Additional data analysis was performed on ATP, NADPH and NADH turnover extracted using the *getMetProduction* function from [7]. Yield was calculated as turnover (sum of fluxes) divided by the specific rate of substrate uptake. Apparent catalytic activities (k_{app} , s^{-1}) were calculated according to Eq (2).

$$k_{app} = \frac{\text{flux}}{E} \quad (2)$$

Where flux refers to median flux, mmol/gDCW/h, obtained from 2000 iterations of random sampling of the solution space [32] and E refers to mean protein concentration ($n = 2$), mmol/gDCW.

Supporting information

S1 Table. Physiological characterization parameters in *R. toruloides* CCT 7815 batch cultivations on three different carbon substrates—glucose (G) (63 g/L, C/N 68.6), xylose (X) (70g/L, C/N 80) and acetate (A) (20 g/L, C/N 80) at exponential growth (exp) and nitrogen limitation (Nlim) phases.
(XLSX)

S2 Table. Gene and metabolite names of *R. toruloides* selected for annotation in Figs 2–4 in main text.
(XLSX)

S3 Table. Using absolute proteomic data to calculate translation rate in *R. toruloides* batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases.
(XLSX)

S4 Table. Proteins whose concentration and the k_{cat} value were integrated in the enzyme-constrained models of *R. toruloides* representing batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp)

and nitrogen limitation (Nlim) phases. Proteins required EC numbers to allow the algorithm to query their k_{cat} values, therefore the existing and new EC numbers were provided to the input file (uniprot.tab) for the GECKO algorithm (columns J-K). In case of multiple EC numbers found for the same gene in rhto-GEM, EC numbers were combined. For further details on how the algorithm selected the k_{cat} values based on their EC numbers, see [17]. (XLSX)

S5 Table. Enzymatic reactions constrained with enzyme abundances in enzyme-constrained genome-scale models of *R. toruloides* for batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases.

(XLSX)

S6 Table. Enzymes with flexibilized concentrations and/or their k_{cat} values for the enzyme-constrained genome-scale metabolic models of *R. toruloides* for batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases. Flexibilization of the measured enzyme abundances was performed automatically by the algorithm in the GECKO Toolbox. Flexibilization of k_{cat} values was performed manually by changing the k_{cat} values retrieved automatically from BRENDa, based on suggestions by the algorithm. Detailed information on flexibilized protein concentrations at ecRhtoGEM repository /results/generate_protModel_s_pipeline. Detailed information on modified k_{cat} values at ecRhtoGEM repository /custom-GECKO/manualModifications.

(XLSX)

S1 Dataset. MS intensities (arbitrary unit) and absolute protein concentrations ($\mu\text{g/g_protein}$) in *R. toruloides* batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases. Absolute concentrations are calculated using total protein amount (TPA) quantification method of duplicate conditions. Normalization refers to 80% recovery of the sample injected applied to the sum of intensities.

(XLSX)

S2 Dataset. Absolute protein abundances ($\mu\text{g/g_protein}$) in *R. toruloides* batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases. Concentrations are calculated using total protein amount (TPA) quantification method. Pairs having adjusted p-value < 0.05 and log2 fold change (log2FC) > |1| of average of duplicate conditions were considered significantly differentially expressed. P value was adjusted for multiple comparisons (n = 3100) using Benjamini & Hochberg method [58]. Protein abundances were filtered by excluding instances, where standard deviation exceeds mean value of two replicates. Normalization refers to 80% recovery of the sample injected applied to the sum of intensities.

(XLSX)

S3 Dataset. Enzyme turnover numbers (k_{cat} , s^{-1}) and apparent catalytic activities (k_{app} , s^{-1}) of *R. toruloides* in batch cultivations on three different carbon sources—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases. k_{cat} values were retrieved from BRENDa using the GECKO Toolbox [17]. k_{app} values were obtained by dividing flux, mmol/gDCW/h , by protein abundance, mmol/gDCW . Flux refers to median from 2000 iterations of random sampling of the solution space [32].

Normalization of protein abundance refers to 80% recovery of sample injected.
(XLSX)

S4 Dataset. Flux predictions in *R. toruloides* batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A) at exponential growth (exp) and nitrogen limitation (Nlim) phases. Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values and are normalized by dividing flux with specific substrate uptake rate (representing % of carbon distribution). Fluxes are represented in non-ec model (base GEM) annotation by merging forward and reverse fluxes created by the GECKO formalism. Flux variability is SD divided by the flux value, multiplied by 100. Flux changes were compared using log2 fold change (log2FC).
(XLSX)

S5 Dataset. Flux predictions with acetate kinase added (t_0886) (phosphate transacetylase removed, t_0082) in *R. toruloides* batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A) at exponential growth (exp) and nitrogen limitation (Nlim) phases. Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values and are normalized by dividing flux with specific substrate uptake rate (representing % of carbon distribution). Fluxes are represented in non-ec model (base GEM) annotation by merging forward and reverse fluxes created by the GECKO formalism. Flux variability refers to SD divided by the flux value, multiplied by 100.
(XLSX)

S6 Dataset. Flux predictions with blocked phosphoketolase (t_0081) in *R. toruloides* batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A) at exponential growth (exp) and nitrogen limitation (Nlim) phases. Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values and are normalized by dividing flux with specific substrate uptake rate (representing % of carbon distribution). Fluxes are represented in non-ec model (base GEM) annotation by merging forward and reverse fluxes created by the GECKO formalism. Flux variability refers to SD divided by the flux value, multiplied by 100.
(XLSX)

S7 Dataset. Flux predictions with NAD/NADH as cofactor for DAD-2/LXR (t_0884) in *R. toruloides* batch cultivations on xylose- (X) based chemically defined medium at exponential growth (exp) and nitrogen limitation (Nlim) phases. Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values and are normalized by dividing flux with specific substrate uptake rate (representing % of carbon distribution). Fluxes are represented in non-ec model (base GEM) annotation by merging forward and reverse fluxes created by the GECKO formalism. Flux variability refers to SD divided by the flux value, multiplied by 100.
(XLSX)

S1 Fig. Growth curves of batch cultivation of *R. toruloides* CCT 7815 on three different carbon substrates at nitrogen limitation. (A) glucose (63 g/L, C/N 68.6), (B) xylose (70g/L, C/N 80) and (C) acetate (20 g/L, C/N 80). Arrows in red are used to denote sampling points for proteomics and protein content measurements. Average of duplicate experiments with SD in

extracellular metabolites concentration (g/L) and intracellular lipid content (g_lipid/gDCW) is illustrated. Curves represent a single measurement in bioreactor 2 (R2) in CO₂ (%), specific growth rate μ (h⁻¹) and biomass concentration (g/L), while for the rate calculations used for modelling duplicate conditions were used.

(TIF)

S2 Fig. D-arabinitol detection in supernatant obtained from *R. toruloides* batch cultivations in xylose-based chemically defined medium (70 g/L). Figures represent HPLC profiles of D-arabinitol during nitrogen limitation phase on xylose (XNlim) performed at 20°C. Column: Chiraldak; eluent: hexane-ethanol (70,30, v/v). Flow rate 0.3 mL/min; detection: refractive index.

(TIF)

S3 Fig. Venn diagrams of significantly differentially expressed proteins in *R. toruloides* during batch cultivations on three different carbon substrates—glucose (63 g/L), xylose (70 g/L) and acetate (20 g/L)—under nitrogen limitation conditions. (A) Comparison between exponential growth (exp) and nitrogen limitation (Nlim) phase. (B) Comparison among substrates during exp phase. Comparison was made using μ g/g of total protein. Pairs having adjusted p-value < 0.05 and log₂ fold change > |1| were considered significantly differentially expressed. P value was adjusted for multiple comparisons (n = 3100) according to Benjamini & Hochberg (1995).

(TIF)

S4 Fig. Proteome integration into enzyme-constrained models for *R. toruloides* in batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases. (A) Protein count as searched against the reference proteome database of *R. toruloides* strain NP11. (B) Mass-wise coverage of proteome in models (g_protein/g_DCW).

(TIF)

S5 Fig. Apparent enzyme catalytic activities, k_{app} , 1/s, of *R. toruloides* in batch cultivations on three different carbon sources—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases. k_{app} calculated using fluxes from flux balance analysis on enzyme-constrained models of *R. toruloides* and measured enzyme absolute abundances. Frequency of k_{app} values represented in log10 scale.

(TIF)

S6 Fig. Flux predictions in *R. toruloides* batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A) at exponential growth (exp) and nitrogen limitation (Nlim) phases. Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values and are normalized by dividing flux with specific substrate uptake rate (representing % of carbon distribution). PPP: pentose phosphate pathway; TCA cycle: tricarboxylic acid cycle. Gene names and corresponding metabolic reaction IDs are included in [S2 Table](#).

(TIF)

S7 Fig. Fluxes carrying NADPH in *R. toruloides* on glucose- (G) based chemically defined medium at exponential growth (exp) (A) and nitrogen limitation (Nlim) (B) phase (mmol/gDCW/h). Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values from flux sampling. Negative fluxes denote metabolite consumption,

positive fluxes denote metabolite production. Percentage (%) denotes the flux divided by NADPH turnover (sum of absolute fluxes involving NADPH). Gene names and corresponding metabolic reaction IDs are included in [S2 Table](#).
(TIF)

S8 Fig. Predicted ATP turnover (mmol/gDCW/h) (A) and ATP yield (mmol_ATP/mmol_substrate) (B) in *R. toruloides* on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—in a chemically defined medium at exponential growth (exp) and nitrogen limitation (Nlim) phases. ATP turnover is calculated as a sum of fluxes involving ATP. ATP yield is calculated as turnover divided by specific rate of substrate uptake. Fluxes are predicted using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Median flux values are used in calculations.

(TIF)

S9 Fig. Predicted NADH turnover (mmol/gDCW/h) (A) and NADH yield (mmol_NADH/mmol_substrate) (B) in *R. toruloides* on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—in a chemically defined medium at exponential growth (exp) and nitrogen limitation (Nlim) phases. NADH turnover is calculated as sum of absolute fluxes involving NADH. NADH yield is calculated as turnover divided by specific rate of substrate uptake. Fluxes are predicted using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Median flux values are used in calculations.

(TIF)

S10 Fig. Average absolute enzyme abundances (μg/g_protein) in *R. toruloides* batch cultivations on three different carbon substrates—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases. Absolute enzyme concentrations are calculated using total protein amount (TPA) quantification method. Results of duplicate experiments with SD are represented. PPP: pentose phosphate pathway; TCA cycle: tricarboxylic acid cycle. Full names of gene abbreviations are included in [S2 Table](#).
(TIF)

S11 Fig. Apparent enzyme catalytic activities, k_{app} , s⁻¹, of *R. toruloides* in batch cultivations on three different carbon sources—glucose (G), xylose (X) and acetate (A)—during exponential growth (exp) and nitrogen limitation (Nlim) phases. k_{app} calculated using fluxes from flux balance analysis on enzyme-constrained genome-scale models of *R. toruloides* and measured enzyme absolute abundances. PPP: pentose phosphate pathway; TCA cycle: tricarboxylic acid cycle. Full names of gene abbreviations are included in [S2 Table](#).
(TIF)

S12 Fig. Fluxes carrying NADPH in *R. toruloides* on xylose- (X) based chemically defined medium at exponential growth (exp) (A) and nitrogen limitation (Nlim) (B) phase (mmol/gDCW/h). Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values from flux sampling. DAD-2/LXR is considered NADP-dependent. Negative fluxes denote metabolite consumption, positive fluxes denote metabolite production. Gene names and corresponding metabolic reaction IDs are included in [S2 Table](#).
(TIF)

S13 Fig. Predicted NADPH turnover (mmol/gDCW/h) (A) and NADPH yield (mmol_NADPH/mmol_substrate) (B) in *R. toruloides* on three different carbon substrates—

glucose (G), xylose (X) and acetate (A)–in a chemically defined medium at exponential growth (exp) and nitrogen limitation (Nlim) phases. NADPH turnover is calculated as sum of absolute fluxes involving NADPH. NADPH yield is calculated as turnover divided by specific rate of substrate uptake. Fluxes are predicted using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Median flux values are used in calculations.

(TIF)

S14 Fig. Fluxes carrying NADH in *R. toruloides* on xylose- (X) based chemically defined medium at exponential (exp) (A) and nitrogen limitation (Nlim) (B) phase (mmol/gDCW/h). Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values from flux sampling. DAD-2/LXR is considered NADP-dependent. Negative flux denotes metabolite consumption, positive flux denotes metabolite production. Letters [m] and [c] denote compartments of cytoplasm and mitochondria. Gene names and corresponding metabolic reaction IDs are included in [S2 Table](#).

(TIF)

S15 Fig. Fluxes carrying ATP in *R. toruloides* on acetate-(A) based chemically-defined medium at exponential growth (exp) (A) and nitrogen limitation (Nlim) (B) phases (mmol/gDCW/h). Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values from flux sampling. Negative flux denotes metabolite consumption, positive flux denotes metabolite production. Letters [m] and [c] denote compartments of cytoplasm and mitochondria. Gene names and corresponding metabolic reaction IDs are included in [S2 Table](#).

(TIF)

S16 Fig. Fluxes carrying NADH in *R. toruloides* on acetate- (A) based chemically defined medium at exponential growth (exp) (A) and nitrogen limitation (Nlim) (B) phase (mmol/gDCW/h). Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values from flux sampling. Negative flux denotes metabolite consumption, positive flux denotes metabolite production. Letters [m] and [c] denote compartments of cytoplasm and mitochondria. Gene names and corresponding metabolic reaction IDs are included in [S2 Table](#).

(TIF)

S17 Fig. Fluxes carrying NADPH in *R. toruloides* on acetate- (A) based chemically defined medium at exponential growth (exp) (A) and nitrogen limitation (Nlim) (B) phase (mmol/gDCW/h). Fluxes are calculated using random sampling of the solution space with 2000 iterations (mmol/gDCW/h) on *R. toruloides* enzyme-constrained genome-scale models. Fluxes represent median values from flux sampling. Negative flux denotes metabolite consumption, positive flux denotes metabolite production. Gene names and corresponding metabolic reaction IDs are included in [S2 Table](#).

(TIF)

S18 Fig. D-arabinitol enantiomer detection using HPLC analysis. Figure represents retention times for arabinitol separation in Chiralpak column, at 20°C, hexane-ethanol (70:30, v/v) mixture.

(TIF)

Acknowledgments

The authors thank Laura Kibena for HPLC analysis, Proteomics Core Facility at University of Tartu for proteome quantification. Lipid analysis was carried out using the instrumentation at the Estonian Center of Analytical Chemistry (www.akki.ee).

Author Contributions

Conceptualization: Nemailla Bonturi, Petri-Jaan Lahtvee.

Data curation: Alina Rekēna.

Funding acquisition: Petri-Jaan Lahtvee.

Investigation: Alina Rekēna, Marina J. Pinheiro, Eliise Tammekivi, Petri-Jaan Lahtvee.

Methodology: Alina Rekēna, Eliise Tammekivi, Eduard J. Kerkhoven.

Project administration: Petri-Jaan Lahtvee.

Supervision: Nemailla Bonturi, Isma Belouah, Koit Herodes, Eduard J. Kerkhoven, Petri-Jaan Lahtvee.

Visualization: Alina Rekēna.

Writing – original draft: Alina Rekēna.

Writing – review & editing: Alina Rekēna, Marina J. Pinheiro, Nemailla Bonturi, Isma Belouah, Eliise Tammekivi, Koit Herodes, Eduard J. Kerkhoven, Petri-Jaan Lahtvee.

References

1. Pinheiro MJ, Bonturi N, Belouah I, Miranda EA, Lahtvee P-J. Xylose Metabolism and the Effect of Oxidative Stress on Lipid and Carotenoid Production in Rhodotorula toruloides: Insights for Future Biorefinery. *Front Bioeng Biotechnol*. 2020; 8: 1008. <https://doi.org/10.3389/fbioe.2020.01008> PMID: 32974324
2. Lopes HJS, Bonturi N, Miranda EA. Rhodotorula toruloides Single Cell Oil Production Using Eucalyptus urograndis Hemicellulose Hydrolysate as a Carbon Source. *Energies*. 2020; 13: 795. <https://doi.org/10.3390/en13040795>
3. Jagtap SS, Deewan A, Liu J-J, Walukiewicz HE, Yun EJ, Jin Y-S, et al. Integrating transcriptomic and metabolomic analysis of the oleaginous yeast Rhodosporidium toruloides IFO0880 during growth under different carbon sources. *Appl Microbiol Biotechnol*. 2021; 105: 7411–7425. <https://doi.org/10.1007/s00253-021-11549-8> PMID: 34491401
4. Chmielarz M, Blomqvist J, Sampels S, Sandgren M, Passoth V. Microbial lipid production from crude glycerol and hemicellulosic hydrolysate with oleaginous yeasts. *Biotechnol Biofuels*. 2021; 14: 65. <https://doi.org/10.1186/s13068-021-01916-y> PMID: 33712047
5. Monteiro de Oliveira P, Aborneva D, Bonturi N, Lahtvee P-J. Screening and Growth Characterization of Non-conventional Yeasts in a Hemicellulosic Hydrolysate. *Front Bioeng Biotechnol*. 2021; 9: 659472. <https://doi.org/10.3389/fbioe.2021.659472> PMID: 33996782
6. Brandenburg J, Blomqvist J, Shapaval V, Kohler A, Sampels S, Sandgren M, et al. Oleaginous yeasts respond differently to carbon sources present in lignocellulose hydrolysate. *Biotechnol Biofuels*. 2021; 14: 124. <https://doi.org/10.1186/s13068-021-01974-2> PMID: 34051838
7. Lopes HJS, Bonturi N, Kerkhoven EJ, Miranda EA, Lahtvee P-J. C/N ratio and carbon source-dependent lipid production profiling in Rhodotorula toruloides. *Appl Microbiol Biotechnol*. 2020; 104: 2639–2649. <https://doi.org/10.1007/s00253-020-10386-5> PMID: 31980919
8. Tiukova IA, Brandenburg J, Blomqvist J, Sampels S, Mikkelsen N, Skaugen M, et al. Proteome analysis of xylose metabolism in Rhodotorula toruloides during lipid production. *Biotechnol Biofuels*. 2019; 12: 137. <https://doi.org/10.1186/s13068-019-1478-8> PMID: 31171938
9. Kim J, Coradetti ST, Kim Y-M, Gao Y, Yaegashi J, Zucker JD, et al. Multi-Omics Driven Metabolic Network Reconstruction and Analysis of Lignocellulosic Carbon Utilization in Rhodosporidium toruloides. *Front Bioeng Biotechnol*. 2021; 8: 612832. <https://doi.org/10.3389/fbioe.2020.612832> PMID: 33585414

10. Zhu Z, Zhang S, Liu H, Shen H, Lin X, Yang F, et al. A multi-omic map of the lipid-producing yeast Rhodosporidium toruloides. *Nat Commun.* 2012; 3: 1112. <https://doi.org/10.1038/ncomms2112> PMID: 23047670
11. Dinh H V., Suthers PF, Chan SHJ, Shen Y, Xiao T, Deewan A, et al. A comprehensive genome-scale model for Rhodosporidium toruloides IFO0880 accounting for functional genomics and phenotypic data. *Metab Eng Commun.* 2019; 9: e00101. <https://doi.org/10.1016/j.mec.2019.e00101> PMID: 31720216
12. Ratledge C, Wynn JP. The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. *Adv Appl Microbiol.* 2002; 51: 1–51. [https://doi.org/10.1016/s0065-2164\(02\)51000-5](https://doi.org/10.1016/s0065-2164(02)51000-5) PMID: 12236054
13. Lian J, Zhao H. Recent advances in biosynthesis of fatty acids derived products in *Saccharomyces cerevisiae* via enhanced supply of precursor metabolites. *J Ind Microbiol Biotechnol.* 2015; 42: 437–451. <https://doi.org/10.1007/s10295-014-1518-0> PMID: 25306882
14. Evans CT, Ratledge C. Induction of xylulose-5-phosphate phosphoketolase in a variety of yeasts grown on d-xylose: the key to efficient xylose metabolism. *Arch Microbiol.* 1984; 139: 48–52. <https://doi.org/10.1007/BF00692711>
15. Tiukova IA, Prigent S, Nielsen J, Sandgren M, Kerkhoven EJ. Genome-scale model of *Rhodotorula toruloides* metabolism. *Biotechnol Bioeng.* 2019; 116: 3396–3408. <https://doi.org/10.1002/bit.27162> PMID: 31502665
16. Chen Y, Nielsen J. Mathematical modeling of proteome constraints within metabolism. *Curr Opin Syst Biol.* 2021; 25: 50–56. <https://doi.org/10.1016/j.coisb.2021.03.003>
17. Sánchez BJ, Zhang C, Nilsson A, Lahtvee P, Kerkhoven EJ, Nielsen J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol.* 2017; 13: 935. <https://doi.org/10.1525/msb.20167411> PMID: 28779005
18. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol.* 2013; 9: 693. <https://doi.org/10.1525/msb.2013.52> PMID: 24084808
19. Chen Y, Pelt-KleinJan E, Olst B, Douwenga S, Boeren S, Bachmann H, et al. Proteome constraints reveal targets for improving microbial fitness in nutrient-rich environments. *Mol Syst Biol.* 2021; 17. <https://doi.org/10.1525/msb.202010093> PMID: 33821549
20. Niebel B, Leupold S, Heinemann M. An upper limit on Gibbs energy dissipation governs cellular metabolism. *Nat Metab.* 2019; 1: 125–132. <https://doi.org/10.1038/s42255-018-0006-7> PMID: 32694810
21. Bonturi N, Crucello A, Viana AJC, Miranda EA. Microbial oil production in sugarcane bagasse hemicellulosic hydrolysate without nutrient supplementation by a Rhodosporidium toruloides adapted strain. *Process Biochem.* 2017; 57: 16–25. <https://doi.org/10.1016/j.procbio.2017.03.007>
22. Lopes JF, Gaspar EMSM. Simultaneous chromatographic separation of enantiomers, anomers and structural isomers of some biologically relevant monosaccharides. *J Chromatogr A.* 2008; 1188: 34–42. <https://doi.org/10.1016/j.chroma.2007.12.016> PMID: 18177879
23. Jagtap SS, Rao C V. Production of d-arabitol from d-xylose by the oleaginous yeast Rhodosporidium toruloides IFO0880. *Appl Microbiol Biotechnol.* 2018; 102: 143–151. <https://doi.org/10.1007/s00253-017-8581-1> PMID: 29127468
24. Haaber J, Cohn MT, Petersen A, Ingmer H. Simple method for correct enumeration of *Staphylococcus aureus*. *J Microbiol Methods.* 2016; 125: 58–63. <https://doi.org/10.1016/j.mimet.2016.04.004> PMID: 27080188
25. Koch AL. Turbidity measurements of bacterial cultures in some available commercial instruments. *Anal Biochem.* 1970; 38: 252–259. [https://doi.org/10.1016/0003-2697\(70\)90174-0](https://doi.org/10.1016/0003-2697(70)90174-0) PMID: 4920662
26. Sanchez BJ, Lahtvee P, Campbell K, Kasvandik S, Yu R, Domenzain I, et al. Benchmarking accuracy and precision of intensity-based absolute quantification of protein abundances in *Saccharomyces cerevisiae*. *Proteomics.* 2021; 21: 2000093. <https://doi.org/10.1002/pmic.202000093> PMID: 33452728
27. Metzl-Raz E, Kafri M, Yaakov G, Soifer I, Gurvich Y, Barkai N. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *Elife.* 2017; 6: e28034. <https://doi.org/10.7554/elife.28034> PMID: 28857745
28. Boehlke' And KW, Friesen JD. Cellular Content of Ribonucleic Acid and Protein in *Saccharomyces cerevisiae* as a Function of Exponential Growth Rate: Calculation of the Apparent Peptide Chain Elongation Rate. 1975; 121: 429–433. Available from: <https://journals.asm.org/journal/ib>.
29. Bonturi N, Pinheiro MJ, de Oliveira PM, Rusadze E, Eichinger T, Liudzütė G, et al. Development of a dedicated Golden Gate Assembly Platform (RtGGA) for *Rhodotorula toruloides*. *Metab Eng Commun.* 2022; 15: e00200. <https://doi.org/10.1016/j.mec.2022.e00200> PMID: 35662893

30. Domenzain I, Sánchez B, Anton M, Kerkhoven EJ, Millán-Oropeza A, Henry C, et al. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nat Commun.* 2022; 13: 3766. <https://doi.org/10.1038/s41467-022-31421-1> PMID: 35773252
31. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol.* 2010; 28: 245–248. <https://doi.org/10.1038/nbt.1614> PMID: 20212490
32. Bordel S, Agren R, Nielsen J. Sampling the Solution Space in Genome-Scale Metabolic Networks Reveals Transcriptional Regulation in Key Enzymes. *PLoS Comput Biol.* 2010; 6: e1000859. <https://doi.org/10.1371/journal.pcbi.1000859> PMID: 20657658
33. Bar-Even A, Noor E, Savir Y, Lieberman-Wolff D, Tawfik DS, et al. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry.* 2011; 50: 4402–4410. <https://doi.org/10.1021/bi2002289> PMID: 21506553
34. Wong B, Murray JS, Castellanos M, Croen KD. D-arabitol metabolism in *Candida albicans*: studies of the biosynthetic pathway and the gene that encodes NAD-dependent D-arabitol dehydrogenase. *J Bacteriol.* 1993; 175: 6314–6320. <https://doi.org/10.1128/jb.175.19.6314-6320.1993> PMID: 8407803
35. Verho R, Putkonen M, Londenborough J, Penttilä M, Richard P. A Novel NADH-linked L-Xylulose Reductase in the L-Arabinose Catabolic Pathway of Yeast. *J Biol Chem.* 2004; 279: 14746–14751. <https://doi.org/10.1074/jbc.M312533200> PMID: 14736891
36. Casal M, Paiva S, Queirós O, Soares-Silva I. Transport of carboxylic acids in yeasts. *FEMS Microbiol Rev.* 2008; 32: 974–994. <https://doi.org/10.1111/j.1574-6976.2008.00128.x> PMID: 18759742
37. Ilarionov A, Lahtvee P-J, Kumar R. Characterization of potassium and sodium salt stress in yeasts. *Appl Environ Microbiol.* 2021; (accepted manuscript). <https://doi.org/10.1128/AEM.03100-20>
38. Sonderegger M, Schümperli M, Sauer U. Metabolic Engineering of a Phosphoketolase Pathway for Pentose Catabolism in *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 2004; 70: 2892–2897. <https://doi.org/10.1128/AEM.70.5.2892-2897.2004>
39. Papini M, Nookaew I, Siewers V, Nielsen J. Physiological characterization of recombinant *Saccharomyces cerevisiae* expressing the *Aspergillus nidulans* phosphoketolase pathway: validation of activity through 13C-based metabolic flux analysis. *Appl Microbiol Biotechnol.* 2012; 95: 1001–1010. <https://doi.org/10.1007/s00253-012-3936-0> PMID: 22367611
40. Bergman A, Hellgren J, Moritz T, Siewers V, Nielsen J, Chen Y. Heterologous phosphoketolase expression redirects flux towards acetate, perturbs sugar phosphate pools and increases respiratory demand in *Saccharomyces cerevisiae*. *Microb Cell Fact.* 2019; 18: 25. <https://doi.org/10.1186/s12934-019-1072-6> PMID: 30709397
41. Cavero S, Vozza A, Del Arco A, Palmieri L, Villa A, Blanco E, et al. Identification and metabolic role of the mitochondrial aspartate-glutamate transporter in *Saccharomyces cerevisiae*. *Mol Microbiol.* 2003; 50: 1257–1269. <https://doi.org/10.1046/j.1365-2958.2003.03742.x> PMID: 14622413
42. Avéret N, Jobin M-L, Devin A, Rigoulet M. Proton pumping complex I increases growth yield in *Candida utilis*. *Biochim Biophys Acta—Bioenerg.* 2015; 1847: 1320–1326. <https://doi.org/10.1016/j.bbabi.2015.07.001> PMID: 26164102
43. Palmieri F, Agrimi G, Blanco E, Castegna A, Di Noia MA, Iacobazzi V, et al. Identification of mitochondrial carriers in *Saccharomyces cerevisiae* by transport assay of reconstituted recombinant proteins. *Biochim Biophys Acta—Bioenerg.* 2006; 1757: 1249–1262. <https://doi.org/10.1016/j.bbabi.2006.05.023> PMID: 16844075
44. Wang S, Chen H, Tang X, Zhang H, Hao G, Chen W, et al. The Role of Glyceraldehyde-3-Phosphate Dehydrogenases in NADPH Supply in the Oleaginous Filamentous Fungus *Mortierella alpina*. *Front Microbiol.* 2020; 11. <https://doi.org/10.3389/fmicb.2020.00818> PMID: 32411121
45. Verho R, Londenborough J, Penttilä M, Richard P. Engineering Redox Cofactor Regeneration for Improved Pentose Fermentation in *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 2003; 69: 5892–5897. <https://doi.org/10.1128/AEM.69.10.5892-5897.2003>
46. Xia J, Sánchez BJ, Chen Y, Campbell K, Kasvandik S, Nielsen J. Proteome allocations change linearly with the specific growth rate of *Saccharomyces cerevisiae* under glucose limitation. *Nat Commun.* 2022; 13: 2819. <https://doi.org/10.1038/s41467-022-30513-2> PMID: 35595797
47. Nomura M. Regulation of Ribosome Biosynthesis in *Escherichia coli* and *Saccharomyces cerevisiae*: Diversity and Common Principles. *J Bacteriol.* 1999; 181: 6857–6864. <https://doi.org/10.1128/JB.181.22.6857-6864.1999>
48. Massaiu I, Pasotti L, Sonnenschein N, Rama E, Cavaletti M, Magni P, et al. Integration of enzymatic data in *Bacillus subtilis* genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-γ-glutamic acid production strains. *Microb Cell Fact.* 2019; 18: 3. <https://doi.org/10.1186/s12934-018-1052-2> PMID: 30626384

49. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* (80-). 2021; 373: 871–876. <https://doi.org/10.1126/science.abb8754> PMID: 34282049
50. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021; 596: 583–589. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
51. Townshend RJL, Eismann S, Watkins AM, Rangan R, Karelina M, Das R, et al. Geometric deep learning of RNA structure. *Science* (80-). 2021; 373: 1047–1051. <https://doi.org/10.1126/science.abe5650> PMID: 34446608
52. Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, et al. Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nat Catal*. 2022; 5: 662–672. <https://doi.org/10.1038/s41929-022-00798-z>
53. Verduyn C, Postma E, Scheffers WA, Van Dijken JP. Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*. 1992; 8: 501–517. <https://doi.org/10.1002/yea.320080703> PMID: 1523884
54. Tammekivi E, Vahur S, Kekišev O, van der Werf ID, Toom L, Herodes K, et al. Comparison of derivatization methods for the quantitative gas chromatographic analysis of oils. *Anal Methods*. 2019; 11: 3514–3522. <https://doi.org/10.1039/C9AY00954J>
55. Tammekivi E, Vahur S, Vilbaste M, Leito I. Quantitative GC–MS Analysis of Artificially Aged Paints with Variable Pigment and Linseed Oil Ratios. *Molecules*. 2021; 26: 2218. <https://doi.org/10.3390/molecules26082218> PMID: 33921401
56. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc*. 2016; 11: 2301–2319. <https://doi.org/10.1038/nprot.2016.136> PMID: 27809316
57. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res*. 2019; 47: D442–D450. <https://doi.org/10.1093/nar/gky1106> PMID: 30395289
58. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995; 57: 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
59. Wang H, Marcišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, et al. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLOS Comput Biol*. 2018; 14: e1006541. <https://doi.org/10.1371/journal.pcbi.1006541> PMID: 30335785
60. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. *Genome Res*. 2003; 13: 244–253. <https://doi.org/10.1101/gr.234503> PMID: 12566402
61. Sánchez BJ, Li F, Kerkhoven EJ, Nielsen J. SLIMER: probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. *BMC Syst Biol*. 2019; 13: 4. <https://doi.org/10.1186/s12918-018-0673-8> PMID: 30634957

Publication III

Rekēna, A., Pals, K., Gavrilović, S., Lahtvee, P.-J. (2025). The role of ATP citrate lyase, phosphoketolase and malic enzyme in oleaginous *Rhodotorula toruloides*. *Applied Microbiology and Biotechnology*, 109, 77.



The role of ATP citrate lyase, phosphoketolase, and malic enzyme in oleaginous *Rhodotorula toruloides*

Alīna Rekēna¹ · Kristjan Pals¹ · Srđan Gavrilović¹ · Petri-Jaan Lahtvee¹

Received: 5 December 2024 / Revised: 21 February 2025 / Accepted: 11 March 2025
© The Author(s) 2025

Abstract

Rhodotorula toruloides is an oleaginous yeast recognized for its robustness and the production of high content of neutral lipids. Early biochemical studies have linked ATP citrate lyase (ACL), phosphoketolase (PK), and cytosolic malic enzyme (cMAE) with de novo lipid synthesis. In this study, we discovered that upon a CRISPR/Cas9-mediated knockout of the ACL gene, lipid content in *R. toruloides* IFO0880 decreased from 50 to 9% of its dry cell weight (DCW) in glucose medium and caused severe growth defects (reduced specific growth rate, changes in cell morphology). In xylose medium, the lipid content decreased from 43 to 38% of DCW. However, when grown on acetate as the sole carbon source, the lipid content decreased from 45 to 20% of DCW. Significant growth defects as a result of ACL knockout were observed on all substrates. In contrast, PK knockout resulted in no change in growth or lipid synthesis. Knocking out cMAE gene resulted in lipid increase of 2.9% of DCW and 23% increase in specific growth rate on glucose. In xylose or acetate medium, no change in lipid production as a result of cMAE gene knockout was observed. These results demonstrated that ACL plays a crucial role in lipid synthesis in *R. toruloides* IFO0880, as opposed to PK pathway or cMAE, whose presence in some conditions even disfavors lipid production. These results provided valuable information for future metabolic engineering of *R. toruloides*.

Key points

- *ACL is crucial for the fatty acid synthesis and growth in *R. toruloides* IFO0880.*
- *Lipid production and cell growth is are unchanged as a result of PK knockout.*
- *Cytosolic malic enzyme does not play a significant role in lipogenesis.*

Keywords Lipid synthesis · ATP citrate lyase · Phosphoketolase · Malic enzyme · *Rhodotorula toruloides*

Introduction

Oleaginous microorganisms are attractive for the production of biofuels and bio-based oleochemicals from cheap raw materials (Probst et al. 2016; Dourou et al. 2018; Saini et al. 2020; Chawla et al. 2022). Red oleaginous yeast *Rhodotorula toruloides* (*R. toruloides*) is a Basidiomycota fungi that stands out for high lipid content, cell densities, and robustness in consuming hemicellulosic hydrolyzates (Ageitos et al. 2011; Monteiro de Oliveira et al. 2021; Osorio-González et al. 2022b, 2022a); moreover, it is relatively mature in genetic engineering tools compared to other red

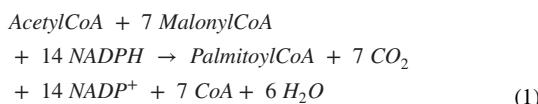
yeasts (Chattopadhyay et al. 2021). Biotechnological applications and prospects of *R. toruloides* as a cell factory are highly promising (Zhao et al. 2022; Yu and Shi 2023; Wu et al. 2023; Sunder et al. 2024). Genomic, transcriptomic, proteomics, and flux balance analyses have established a basic understanding of metabolic pathways and gene functions in *R. toruloides* (Zhu et al. 2012; Coradetti et al. 2018; Dinh et al. 2019; Tiukova et al. 2019a, 2019b; Pinheiro et al. 2020; Kim et al. 2021; Rekēna et al. 2023). Advanced genetic engineering tools have been developed in *R. toruloides* and are being constantly improved. IFO0880 is the best characterized haploid (mating type A2) strain natively producing high titers of neutral lipids up to 8 g/L, reaching lipid content up to 36% of the dry cell weight (DCW) (Zhang et al. 2016b). It is genetically different from another haploid strain NP11 (mating type A1), capable of accumulating similar lipid titers, but significantly higher lipid content (up to 54% of DCW) (Zhu et al. 2012; Zhang et al. 2016b, 2022).

✉ Petri-Jaan Lahtvee
lahtvee@taltech.ee

¹ Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia

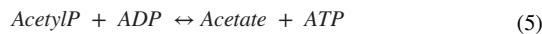
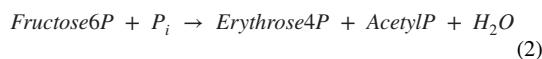
Successful targeted gene deletion with CRISPR has been demonstrated in both IFO0880 and NP11 strains (Otopual et al. 2019; Jiao et al. 2019; Schultz et al. 2019). The CRISPR/Cas9 offers a simplistic design for targeted (and multiplexed) genome editing (Chattopadhyay et al. 2021). The Cas9 enzyme introduces a double-stranded break in the targeted loci. Precise targeting of the Cas9 endonuclease can be achieved by changing a 20 nucleotide guide RNA (gRNA) sequence (Jinek et al. 2012). Due to lack of plasmid capable of replicating in *R. toruloides* to express CRISPR constructs, existing CRISPR tools in the species are based on a chromosomal integration strategy—stable genome integration of Cas9 and gRNA expression cassette in the chromosome (Otopual 2019). Several integration plasmids using a selectable drug marker cassette that encodes antibiotic resistance are available (Gong et al. 2024).

One of the most used strategies to enhance lipid production in oleaginous microorganisms is by cultivating cells in secondary nutrient limitation (Wang et al. 2018). An effective way to achieve it is by manipulating nitrogen availability in the culture medium (high C/N ratio) (Papanikolaou and Aggelis 2011; Lopes et al. 2020), but can also be done with manipulating phosphate (Wu et al. 2010) or sulfur concentration (Wu et al. 2011). The main reaction in biosynthesis of fatty acids is carried out by the multi-enzymatic complex of fatty acid synthetase (FAS) in the cytosol. FAS enzyme complex assembles acetyl-CoA and malonyl-CoA into coenzyme A activated fatty acids, such as palmitic acid, using the reducing power from NADPH (Eq. 1) (Tehlivets et al. 2007; Papanikolaou and Aggelis 2011). This is either followed by elongation and desaturation by dedicated enzymes in the endoplasmic reticulum or subsequently used for the acylation of glycerol backbone to synthesize neutral lipids (triacylglycerols, TAGs) (Tehlivets et al. 2007). Accordingly, FAS requires constant supply of acetyl-CoA, malonyl-CoA, and NADPH in cytosol (Beopoulos 2011).



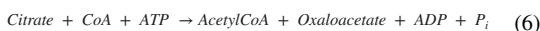
Phosphoketolase (PK) (EC 4.1.2.9) is an enzyme presumed to be one of the candidates to increase acetyl-CoA supply for de novo lipid synthesis. PK is a cytosolic enzyme that cleaves xylulose 5-phosphate or fructose 6-phosphate of the pentose phosphate pathway (PPP) to two-carbon compound acetyl-phosphate and glyceraldehyde 3-phosphate or erythrose 4-phosphate, respectively (Eq. 2 and Eq. 3) (Evans and Ratledge 1984). While acetyl-phosphate will be converted into acetate or acetyl-CoA, the latter two compounds can be recycled via glycolysis and the pentose phosphate pathway (PPP), respectively. On glucose, the involvement of PK in the lipid synthesis was

generally not supported by early biochemical studies with the wild type yeasts (Botham and Ratledge 1979; Boulton and Ratledge 1981). On xylose, there was a strong correlation between PK activity and higher biomass yields demonstrated in wild type *R. toruloides* CBS14 and 14 other yeasts (Evans and Ratledge 1984). Later, it was shown that with overexpression, the native or heterologous PK can enhance lipid production in various yeasts on hemicellulosic carbon sources (Xu et al. 2016; Niehus et al. 2018; Donzella et al. 2019; Kamineni et al. 2021). Based on stoichiometric genome-scale modeling, PPP has been demonstrated as the main glycolytic pathway in *R. toruloides*, while PK playing a crucial role in converting xylulose 5-phosphate into acetyl-phosphate and glyceraldehyde 3-phosphate (Lopes et al. 2020; Rekēna et al. 2023). Omics data show that PK is abundant on glucose, xylose, or acetate in mineral medium (Kim et al. 2021); furthermore, it was significantly upregulated on glucose during lipid accumulation (*p* value 0.043) (Rekēna et al. 2023). From the available genomic data, *R. toruloides* IFO0880 possesses a single PK gene, but it is not clear whether the conversion of acetyl-phosphate is catalyzed by phosphate transacetylase, also known as phosphate acetyltransferase (PTA) (EC 2.3.1.8) or acetate kinase (EC 2.7.2.1) (both reversible) (Eq. 4 and Eq. 5) (Dinh et al. 2019; Tiukova et al. 2019b). However, recently engineered NP11 strains overexpressing non-native PTA from bacterial *B. subtilis* demonstrated increased lipid production by up to 15%, increased glucose consumption, and cell mass (Yang et al. 2018).



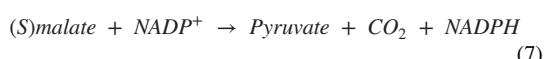
In contrast, ATP citrate lyase (ACL) (EC 2.3.3.8) was associated with lipid synthesis from the early biochemical studies, as its presence correlated to high lipid content in wild type yeasts (Boulton and Ratledge 1981). ACL is a cytosolic enzyme that cleaves cytosolic citrate to acetyl-CoA and oxaloacetate (OAA) (Eq. 6). According to the accepted mechanism for lipid synthesis, ACL activity is associated with the onset of nitrogen limitation that causes excess citrate being exported from the mitochondria to cytosol and used for lipid synthesis under nitrogen limitation (Botham and Ratledge 1979; Boulton and Ratledge 1981). Omics data show that ACL is abundant on glucose, xylose, and acetate in various *R. toruloides* strains, in most of the cases reported as upregulated on proteome level during lipid synthesis and

being more abundant compared to PK (Tiukova et al. 2019a, b; Kim et al. 2021; Rekēna et al. 2023). Also the metabolic engineering studies have shown that ACL plays an important role in lipid biosynthesis in *Ascomycota* yeasts (Liu et al. 2013; Zhang et al. 2014; Sato et al. 2021). In *R. toruloides*, ACL is encoded by a single gene (Zhu et al. 2012), while in ascomycetous yeasts and plants, it has two subunits (Nowrousian et al. 2000; Fatland et al. 2002). Recently, engineered IFO0880 strains overexpressing the ACL gene demonstrated increased fatty alcohol titers (Schultz 2022, Cao et al. 2022). Another possible route to cytosolic acetyl-CoA synthesis from pyruvate is via pyruvate decarboxylase (PDC)—acetaldehyde dehydrogenase—acetyl-CoA synthetase (ACS) path. Previous proteomics studies showed that the abundance of PDC and ACS was low compared to PK and ACL (Rekēna et al. 2023) and downregulated twofold during lipid accumulation (Tiukova et al. 2019a, b); therefore, this pathway was not investigated in this study.



In line with the lipid synthesis mechanism, it was suggested that ACL, FAS, and malic enzyme (MAE) (EC 1.1.1.40) could work together by forming a complex to facilitate fatty acid synthesis (Ratledge 2004; Beopoulos et al. 2011). Because ACL catalyzes the conversion of citrate into acetyl-CoA and OAA, the latter can be transformed into malate and further converted into pyruvate by MAE, simultaneously reducing NADP(+) into NADPH (Eq. 7); thus, MAE may also play an important role to supply of the reducing power for fatty acid synthesis. Several metabolic engineering studies confirmed the hypothesis of the role of MAE in filamentous fungi, like *Mucor circinelloides* (Wynn et al. 1997; Zhang et al. 2007), *Mortierella alpina* (Wynn et al. 2001), and *Aspergillus nidulans* (Wynn and Ratledge 1997), suggesting that malic enzyme was one of the rate-limiting steps for fatty acid synthesis. However in similar studies with the oleaginous ascomycetous yeast *Yarrowia lipolytica* and basidiomycetous yeast *Rhodotorula glutinis*, it did not perform the same way (Yoon et al. 1984; Zhang et al. 2013; Blazeck et al. 2014; Wasylenko et al. 2015; Dulermo et al. 2015; Qiao et al. 2017; Zhu et al. 2023). Namely, it resulted in a current understanding that oleaginous yeasts are different from filamentous fungi, and the primary sources of cytosolic NADPH are glucose 6-phosphate dehydrogenase (G6PD, also known as ZWF) and 6-phosphogluconate dehydrogenase of the PPP, and isocitrate dehydrogenase (Chawla et al. 2022; Sreeharsha and Mohan 2020; Yang et al. 2012; this study), an alternative route being via cytosolic malic enzyme only while growing on substrates that are not catabolized via glycolytic pathways. Both ZWF and MAE are abundant in *R. toruloides* in mineral medium on glucose, xylose, or acetate, but the results of their differential expression

during lipid synthesis are inconclusive among the studies (Tiukova et al. 2019a, b; Kim et al. 2021; Rekēna et al. 2023). The observed higher abundance of oxPPP enzymes suggests their primary role in NADPH regeneration, but it was not investigated further due to report on gene essentiality from a previous study (Coradetti et al. 2018). Moreover, the deletion of MAE would tell if the other pathway prevailed, as these are the two main candidate pathways for the regeneration of NADPH in *R. toruloides*. It should be noted that *Y. lipolytica* possesses only a mitochondrial MAE (mMAE), while *R. toruloides* holds both cytosolic and mitochondrial forms of MAE (Zhu et al. 2012; Coradetti et al. 2018). Engineered IFO0880 strains overexpressing native cytosolic MAE (cMAE) enzyme demonstrated a minor increase in lipid synthesis on glucose (Zhang et al. 2016a). A heterologous overexpression of MAE from *Y. lipolytica* did not increase the expression of fatty alcohols in *R. toruloides* IFO0880 (Schultz et al. 2022). The role and cofactor dependency of the *R. toruloides* mMAE form still remains unclear. mMAE was not investigated in this study.



Taken altogether, to clarify whether PK, ACL, and cMAE contributes to the lipid biosynthesis in *R. toruloides*, the latest CRISPR/Cas9 tool was used for a targeted single gene knockout of PK, ACL, and cMAE in a wild type IFO0880. Cell growth and lipid production in the knockout strains were examined. To understand the substrate-dependent differences, the knockouts were characterized in a chemically defined medium at nitrogen limitation (C/N 80) on glucose, xylose, and acetic acid as a sole carbon sources. The study revealed very different physiological responses between these genes and lipid synthesis, providing valuable insights in *R. toruloides* metabolism, building understanding relevant for the future metabolic engineering. ACL knockout strain demonstrated the most severe effect on the growth and lipid production on all the characterized substrates while ΔPK and ΔcMAE strains demonstrated less severe physiology differences in comparison to the wild type strain. This is the first study of these gene knockouts in *R. toruloides*.

Materials and methods

Strains, media, and conditions

Routine growth *R. toruloides* was performed in YPD medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L glucose).

For selection or maintenance of transformants, 200 $\mu\text{g}/\text{mL}$ G418 (Gibco) was added to the YPD medium.

Lipid content was measured in cells grown in a low-nitrogen mineral medium. Basal medium contained 3.0 g/L KH_2PO_4 , 0.5 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 15 mg/L EDTA, 4.5 mg/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.3 mg/L $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 1 mg/L $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 0.3 mg/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 4.5 mg/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 3 mg/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0.4 mg/L $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 1 mg/L H_3BO_3 , 0.1 mg/L KI, 0.05 mg/L biotin, 1 mg/L calcium pantothenate, 1 mg/L nicotinic acid, 25 mg/L inositol, 1 mg/L thiamine HCl, 1 mg/L pyridoxine HCl, and 0.2 mg/L para-aminobenzoic acid (Verduyn et al. 1992). Basal medium was supplemented with 20 g/L glucose (Roth), or 20 g/L xylose, or 10 g/L acetate (Chempur) as a sole carbon source. Ammonium sulfate $(\text{NH}_4)_2\text{SO}_4$ was used as a nitrogen source to obtain a C/N ratio of 80 (mol/mol); 0.55 g/L of $(\text{NH}_4)_2\text{SO}_4$ (Fisher Bioreagents) was added to the medium containing glucose or xylose, and 0.275 g/L $(\text{NH}_4)_2\text{SO}_4$ was added to the medium containing acetate. For adjusting the media containing glucose or xylose, 5.2 g/L K_2HPO_4 (Roth) was added (pH 5.8 to 8.0). The potassium phosphate buffer did not sustain the pH with 10 g/L of acetate in the medium; therefore, instead, it was adjusted to a starting pH 6 with 2 M KOH (Thermo Fisher Scientific).

Escherichia coli strain TOP10 (Thermo Fisher Scientific) was used for plasmid assembly and routine cloning. The strain and its derivatives were grown in Luria Broth (LB) medium at 37 °C and 200 rpm with 50 µg/mL kanamycin.

Unless differently stated, all other reagents were purchased from Sigma-Aldrich Co., St Louis, MO, USA.

DNA sequences

DNA sequences of *R. toruloides* IFO0880 strain were obtained from MycoCosm database (*R. toruloides* v4.0) (JGI) (Coradetti et al. 2018). Genes with accession numbers 9725 (ATP citrate lyase, ACL) and 13,382 (phospho-ketolase, PK) were retrieved and used to design the guide RNAs, as there were no other alternative genes with a similar sequence. Based on the annotation found from the two existing genome-scale models, gene with accession number 12761 was retrieved and used to design gRNAs to target the cytosolic, NADP-dependent malic enzyme (cMAE) (Tiukova et al. 2019b; Kim et al. 2021). The single gRNAs targeting the first exon 1 of ORF of the target gene (exon 5 in case of the ACL gene) were designed with the CCTop online tool using *Ganoderma lucidum* as the reference genome (Stemmer et al. 2015). 23-nucleotide target sequence ending with NGG (*Streptococcus pyogenes*) and set custom overhangs (forward strand 5' CGCA and reverse strand 5' AAC) was selected based on the CRISPRerater efficacy score (Labuhn et al. 2018) above 74 (high efficacy).

Oligonucleotides without the NGG 3-mer (20-nucleotide target sequence + overhangs) were synthesized by IDT (Integrated DNA Technologies, Leuven, the Netherlands).

Guide RNA cloning

The plasmid pPBO.202 for CRISPR/Cas9-mediated genome editing of *R. toruloides* IFO0880 was obtained from the JBEI Registry <https://registry.jbei.org/> (part ID JBEI223791). pPBO.202 contained constructs for a functional expression of the CRISPR/Cas9 system in *R. toruloides*. They were optimized by Otuopal and colleagues (Otuopal et al. 2019) to be as follows: (i) *E. coli* elements of ColE1 (KanR promoter, kanamycin resistance), (ii) a gRNA expression cassette with *R. toruloides* fusion 5S rRNA-tRNA^{Phc} promoter, 2 *Bsa*I sites, the *S. cerevisiae* SUP4 terminator, (iii) a codon optimized SpCas9 expression cassette with IFO0880 GPD1 promoter and NOS terminator, and (iv) *R. toruloides* G418 resistance cassette pTUB2-G418-tTUB2. Two oligos with a forward strand 5' CGCA and reverse strand 5' AAC overhang were annealed and subsequently cloned by digesting plasmid pPBO.202 with *Bsa*I (Thermo Fisher Scientific) and ligating in T4 Ligase buffer (Thermo Fisher Scientific) in a single pot reaction. Plasmid cloning was performed in *Escherichia coli* strain TOP10 (Thermo Fisher Scientific) according to the manufacturer's instructions, and bacterial plasmid DNA was purified using FavorPrep Mini Plasmid Kit (FAVORGEM, Ping Tung, Taiwan). Resulting plasmids pPBO.202-13382(3), pPBO.202-9725(1), and pPBO.202-12761(1) were checked by Sanger sequencing. Several different gRNAs per target gene were tested in sequential order if the previous gRNA candidate failed to result in the target site disruption (the number in the brackets next to each construct indicates the number of different gRNA test candidates that resulted in a functional CRISPR/Cas9 mediated target site disruption).

Yeast transformation

Yeast transformation was performed using lithium acetate/PEG-mediated chemical transformation method as previously described (Tsai et al. 2017; Bonturi et al. 2022). Briefly, cells from an overnight YPD culture were inoculated at OD_{600} 0.2 in a shake flask containing 50 mL of YPD (for up to 10 transformations). At OD_{600} 0.8, cells were harvested, washed in dH_2O , and eventually mixed with 240 µL PEG 4000 (Fisher Scientific), 36 µL 1 M LiAc pH 7.5 (Alfa Aesar), 24 µL dH_2O , 10 µL pre-boiled Salmon Sperm (10 mg/mL), and 0.1–10 µg purified circular plasmid DNA dissolved in 50 µL of water. The mixture was incubated at 30 °C shaking for 30 min, added with 34 µL of DMSO and heat shocked at 42 °C for 15 min. After removing the supernatant, collected cells were resuspended in 2 mL of YPD and incubated at 30 °C overnight for the recovery. Then, cells were collected, spread on YPD agar plates (agar 20 g/L) the selective antibiotic 200 µg/mL G418, and incubated at 30 °C for 2–4 days until the colonies appeared.

Isolation of gene knockouts

Multi-step verification was used to confirm target gene inactivation: (i) antibiotic selection, (ii) genomic DNA isolation using lithium acetate/SDS/heat lysis (Lööke et al. 2011) and Cas9-specific PCR on the genomic DNA, (iii) target gene PCR amplification and sequencing. PCR amplification at the target locus of PK was performed with primers 5' CCTCCCTCTCACTCTGCAC '3 (forward) and 5' CACGCTGTCCAGTCAAAGAA '3 (reverse); ACL—with primers 5' AGCTCCTCAAGCACGTCACT '3 (forward) and 5' GTAGACGACCGAAGCACCAC '3 (reverse), and MAE—with primers 5' ACTCGCTCTCCCTCTCTC '3 (forward) and 5' ACTCGGAAAACCACGGCTTC '3 (reverse). It was performed using high-fidelity Platinum SuperFi II DNA Polymerase Green PCR Master Mix (2X) (Thermo Fisher Scientific, Vilnius, Lithuania) according to manufacturer's instructions for a high GC content template. PCR amplification of the Cas9 coding sequence with primers 5' GGAGTCGCGGGACGCCAAC '3 (forward) and 5' ACACGTTGGCGTCCCGCGA '3 (reverse) was performed using DreamTaq Polymerase Green PCR Master Mix (2x) (Thermo Fisher Scientific, Vilnius, Lithuania). PCR DNA was purified using FavorPrep Mini Gel/PCR Kit (FAVORGEN, Ping Tung, Taiwan).

Physiological characterization

The seed culture from a fresh YPD agar plate was inoculated to a sterile 250-mL baffled Erlenmeyer flask filled with 25 mL of YPD. After 19 h at 30 °C and 200 rpm, sufficient amount of broth was transferred to a sterile 50 mL Falcon tube and pelleted (5000 g, 5'). After discard of supernatant, cell pellet was washed twice with a sterile dH₂O and used to inoculate 10 mL of low-nitrogen mineral medium at OD₆₀₀ of 0.5 in aerobic 50 mL Biosan RTS-8 Multi-channel falcon tube bioreactors (Biosan, Riga, Latvia). Calibration curves for OD quantification were prepared in a range between OD₆₀₀ 0 and 55 with 8 calibration points using *R. toruloides* shake flask culture in a chemically defined medium. Calibration curves for pH and dO measurement were set up according to the manufacturer's instructions. Real-time logging was set to every 30 min with RTS-8 proprietary software. Agitation speed was set to 2500 rpm, temperature 30 °C. Every 12 h after the cells had reached the exponential growth phase, except for the first sample after time point zero, samples to measure offline OD₆₀₀ and extracellular metabolites were withdrawn from the bioreactor tube in the laminar flow cabinet, transferred to a 2-mL Eppendorf tubes and centrifuged for 5' at 11,000 × g. The supernatant was collected and stored at –20 °C for further analysis. The first sample was collected 6 h after inoculation. All cultivations

were carried out in a batch regime; no feed was added throughout the process. All samples were collected from three independent replicates for each experimental condition, i.e., combination of strain and carbon source. Specific growth rate was calculated by fitting an exponential trend line to the real time OD data using Eq. 8 with MS Excel software.

$$OD(t) = OD_0 \times e^{\mu t} \quad (8)$$

where OD_0 is the initial optical density, t is elapsed time, μ is the specific growth rate, $OD(t)$ is the resulting optical density at time t .

Lipid analysis

At the end of the cultivation experiment, cell broth containing at least 12 mg of DCW was transferred from the falcon tube bioreactor to a 15 mL tube and separated by centrifugation (5000 g, 10', 4 °C). Cell pellets were frozen at –80 °C, lyophilized, and stored at –20 °C until further analysis. To quantify lipids, fatty acids were extracted and derivatized by using one-step method as described by Sukhija and Palmquist (Sukhija and Palmquist 1988), with internal standard heptadecanoic acid (17:0) solution in toluene (5 mg/mL). Fatty acids were analyzed using Agilent (Santa Clara, CA, USA) 6890 A gas chromatograph equipped with a flame ionization detector (GC-FID). The column was a quartz capillary column (100 m × 0.25 mm) with liquid phase CP-Sil 88, temperature programmed from 70 to 180 °C at 13 °C/min, held for 40 min, 180 to 225 °C at 5 °C/min, held for 15 min. Hydrogen was used as the carrier gas for GC and FID (FID flow rate 30 mL/min), air flow rate 300 mL/min. Fatty acids were identified by comparison of their retention times with the retention time of mixtures with known fatty acid methyl ester composition and concentration: Supelco 37 Component FAME Mix (Sigma-Aldrich Co., St Louis, MO, USA), Nu-Chek Prep CLC 603, and Nu-Chek Prep CLC 428 (Nu-Chek Prep Inc., Elysian, MN, USA). Fatty acid composition was presented as a gram of individual fatty acids per 100 g of total fatty acids (the same as relative %). Lipid content was calculated as the weight sum of individual fatty acids in 100 g of sample, divided by 0.9 to provide TAG equivalent known from earlier biochemical studies, as described by Sukhija and Palmquist (1988), and expressed as gram lipids per gram dry cell weight (DCW). Non-lipid cell mass was calculated after subtraction of intracellular lipids from cell mass.

Analytical methods

Extracellular metabolites in cultivation broth were measured using HPLC (LC-2050C, Shimazu, Kyoto, Japan) equipped with a refractive index detector and a variable wavelength

detector. Prior analysis, all samples were thawed, filtered (0.22 µm), and diluted in an appropriate diluent as indicated below. Metabolites were analyzed using the Aminex HPX-87H 300×7.8 mm column (Bio-Rad, Hercules, CA, USA) at 45 °C, with 5 mM H₂SO₄ as the mobile phase with isocratic elution at 0.6 mL/min. The concentrations of acetate, glucose, glycerol, xylose, citrate, xylitol, and D-arabitol were quantified by refractive index detector based on the calibration curves prepared in the range between 0.625 and 20 g/L. No unknown peaks by variable wavelength detector set at 210 nm were detected. It should be noted that with the HPX-87H, it is not possible to determine the chirality of arabitol, but we have determined it in our previous publication (Rekēna et al. 2023). To adjust the sample volume, samples containing glucose or xylose were diluted to 5X in ultrapure H₂O (Milli-Q Ultrapure Water System, Merck, Darmstadt, Germany). To adjust the pH and the volume, samples containing acetate were diluted to 5X in 10 mM mobile phase to reduce the pH below 8, the maximum range of the column. Biomass optical density data were calibrated by gravimetric dry cellular mass measurements. The linearly fitted calibration coefficient was 0.3 for cells under the exponential growth and 0.26 for cells under nitrogen limitation. At the end of cultivation, 3 µL of culture were transferred to a glass slide for microscopy. Cells from an undiluted culture were visualized by a bright-field microscope under 100× magnification CX21 (Olympus, Tokyo, Japan). All images were acquired using Apple iPhone 13 mini (Apple Inc., Cupertino, CA, USA).

Statistical analysis

Statistical analysis of physiological parameters evaluation was performed using GraphPad Prism 9.5.1 (GraphPad Software Inc., San Diego, CA, USA). Statistical significance was calculated using one-way ANOVA (Analysis of variance) at 0.05 significance level. *p*-values were adjusted (apval) for multiplicity following Dunnett multiple comparison testing against the wild type IFO0880.

Results

Sequencing of the genomic DNA confirmed gene knockouts introduced by CRISPR/Cas9

Gene knockouts were introduced into the *R. toruloides* IFO0880 genome by using a one-step CRISPR/Cas9 strategy. The coding sequence of the Cas9 and appropriate gRNAs were randomly integrated into the genome through use of a single plasmid vector pPBO.202 (JBEI Registry part ID JBEI223791) with the G418 selection marker using lithium acetate/PEG-mediated chemical transformation (see

Table 1 List of gRNAs used in this study. Number in the brackets refers to number of different gRNA needed to be tested in order to successfully generate the gene knockout

gRNA	Target gene	Target sequence (5'-3')
9725(1)	ACL	AGTACGTCGTCGGTCCCAAG
13382(3)	PK	GATGCAGAGGAAGTTGACCA
12761(1)	cMAE	CCCTCCCAGCCCCCTTCAAGG

“Materials and Methods”). A set of single guide RNAs targeting PK, ACL, or cMAE were designed to separately target and disrupt the first exon of the targeted genes (Table 1). Gene knockouts by CRISPR/Cas9 were confirmed by sequencing of the genomic DNA around the gRNA-targeted cut site. Mutations at the site targeted by the gRNAs were identified by aligning sequences against the wild type *R. toruloides* IFO0880 reference genome (Fig. 1a). In case of ACL knockout, there was a visible phenotypic difference (Fig. 1b). Other knockouts did not display a phenotypic difference. The gene knockouts by CRISPR/Cas9 were presumably generated by sequence of Cas9 endonuclease and error prone NHEJ DNA repair resulting in frameshift mutation in wild type *R. toruloides* strain. It has been reported that the frequency of HR in *R. toruloides* is low (Koh et al. 2014). List of strains used in this study is provided in Table 2; number of clones screened is reported in Supplementary Table S1.

During the strain characterization step, growth curves (in the next section) of wild type cells and cells harboring gene editing vector did not show difference in growth rates. No significant differences in growth rates were observed between different mutation variants of cMAE and PK during initial screening. This shows that cassette did not cause detrimental fitness effects, similarly as reported by Otoupal and colleagues (Otoupal et al. 2019). These results demonstrated that it is possible to achieve the disruption of central carbon metabolism genes with a method that relies on a stable genome integration of CRISPR constructs in the genome using a selectable drug marker.

ACL knockout decreases lipid synthesis on glucose

To understand the role of PK, ACL, and cMAE in the lipid synthesis of *R. toruloides* IFO0880 strain, we compared the lipid content produced by the knockout strains versus the wild type IFO0880. Cells were induced for lipid production using low-nitrogen mineral medium containing glucose, xylose, or acetate as a solo carbon source. *R. toruloides* is an obligate aerobe. For the best mixing and aeration support, high-throughput falcon tube bioreactors equipped with a reverse-Spin® technology RTS-8 (Biosan, Riga, Latvia) instead of shaker flasks or traditional stirred tank bioreactors

were used for strain characterization. This system was not equipped with the option to maintain desired pH and dissolved oxygen settings, but provided for a real time non-invasive tracking of OD, pH, and dissolved oxygen data. We used real-time dissolved oxygen monitoring to precisely track the end of the experiment. To account for the pH drop happening with *R. toruloides* in an uncontrolled pH environment (Tingajeva 2024), K₂HPO₄ was added in the starting media at pH 6.5 to an amount sufficient to buffer significant pH changes (see “Materials and Methods”).

On glucose, online OD₆₀₀ data showed no difference among the strains in the lag phase prior to the exponential growth phase (Fig. 2a), but the maximum specific growth rate of the ACL knockout was significantly decreased by 41% as compared to the wild type strain (Fig. 2b, Supplementary Table S2, *apval* < 0.002). The growth of the ACL mutant continued until 18 h, but then the growth stopped and final DCW did not reach that of the wild type strain (Fig. 2c). Sugar analysis confirmed that the Δ ACL stopped consuming glucose after 18 h (Fig. 2d). The pH remained within the optimal range throughout the cultivation (Supplementary

Fig. S1). The dissolved oxygen curves showing when the end of the experiment was declared are available from Supplementary Fig. S2.

For lipid quantification, we used gas chromatographic (GC) determination (see “Materials and Methods”). In the wild type strain, lipid content reached $50.18 \pm 0.37\%$ (g lipid per g of DCW) after 54 h of growth on low-nitrogen mineral media (Table 3). The ACL knockout strain showed a significant decrease in lipid content by 81% compared to the wild type strain, reaching only $9.43 \pm 0.039\%$ (g lipid/g DCW) (Fig. 2e, *apval* < 0.0001). In case of the ACL knockout, the difference in non-lipid DCW (g/L) compared to the wild type strain was 58% (*apval* < 0.0001) (Supplementary Fig. S3a). These results indicated that ACL was not only essentially involved in *R. toruloides* fatty acid synthesis, similarly as previously studied in *Y. lipolytica* and *L. starkeyi* (Dulermo et al. 2015; Liu et al. 2013; Sato et al. 2021), but also involved in some other essential processes that had a complementary effect on cellular physiology, whereas knocking out the PK gene showed no change in specific growth rate or lipid content, suggesting PK had no major

Fig. 1 Examples of targeted gene disruption using CRISPR/Cas9. Only one example shown per targeted gene. **a** Partial sequencing of phosphoketolase (PK), ATP citrate lyase (ACL), and cytosolic malic enzyme (cMAE) of one edited colony near the cut site in the targeted genes. **b** Phenotype comparison of wild type *R. toruloides* IFO0880 versus Δ ACL, Δ PK, and Δ cMAE strains after 9 days growth on YPD agar

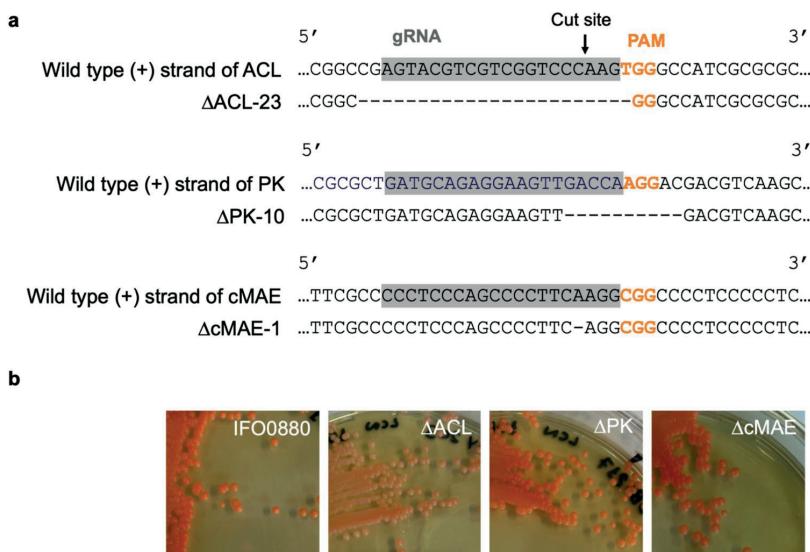


Table 2 List of strains used in this study

Strain name	Genotype	Parent strain	Antibiotic	Mutation	Source/reference
IFO0880 (NBRC 0880)	<i>Rhodotorula toruloides</i> strain IFO0880 (now NBRC 0880), mating type A2	N/A	N/A	N/A	NBRC collection
Δ ACL	pPBO.202-9725(1)	IFO0880	G418	23 bp deletion	This study
Δ PK	pPBO.202-13382(3)	IFO0880	G418	10 bp deletion	This study
Δ cMAE	pPBO.202-12761(1)	IFO0880	G418	1 bp deletion	This study

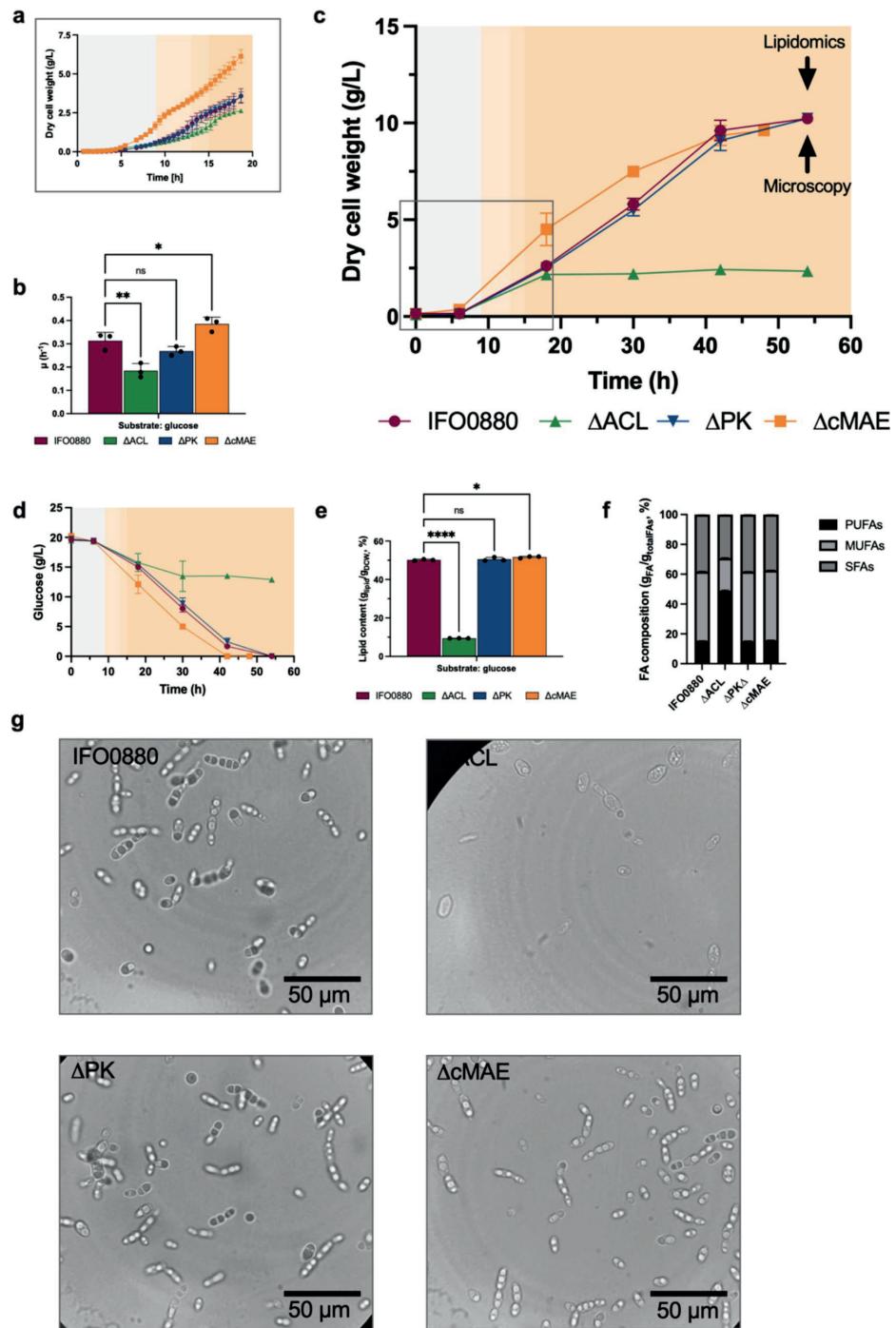


Fig. 2 Physiological parameters obtained in glucose-based pH-adjusted low-nitrogen mineral medium (C/N ratio 80) (Verduyn et al. 1992). Data obtained from cultivating *R. toruloides* wild type IFO0880 (circle), ATP citrate lyase knockout (Δ ACL, pyramid), phosphoketolase knockout (Δ PK, inverted pyramid), and cytosolic malic enzyme knockout (Δ cMAE, square) strains in falcon tube bioreactors with non-invasive OD, dO_2 , and pH sensors. **a** A snapshot of the online OD_{600} data displayed as the dry cell weight (g/L), **b** maximum specific growth rate fitted from the exponential growth phase OD_{600} data (h^{-1}), **c** dry cell weight (g/L), **d** substrate concentration (g/L), **e** lipid content (g_{lipid}/g_{DCW} %); cells harvested at the end of cultivation, **f** relative fatty acid composition ($g_{FA}/g_{totalFAs}$ %); cells harvested at the end of cultivation, **g** morphology comparison of wild type IFO0880 versus knockout strains visualized by bright-field microscope, cells harvested at the end of cultivation. Panels **a** and **c** background colors denote the switch in growth phases based on dO_2 curves available from Supplementary Fig. S2, presumably pointing to the onset of the nitrogen limitation phase. *FAs* fatty acids, *SFAs* saturated fatty acids, *MUFAs* monounsaturated fatty acids, *PUFAs* polyunsaturated fatty acids. Error bars are calculated as a standard deviation from three biological experiments. Asterisks denote statistical significance (ANOVA Dunnett's multiple comparison test against the wild type IFO0880 strain, adjusted p value * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). ns is used to denote changes that are statistically nonsignificant

effect on growth or lipid synthesis under the studied environmental conditions (Fig. 2a–f). In contrast, the specific growth rate of Δ cMAE was increased by 23% compared to the wild type strain, reaching $0.39 \pm 0.030 h^{-1}$ (Fig. 2b, p val 0.04). Lipid content of Δ cMAE increased by 2.9% of DCW compared to the wild type strain (p val 0.03), but no significant changes in the fatty acid profile were observed (Fig. 2e). These results indicated that cMAE in *R. toruloides* is not essential in lipid biosynthesis under the studied environmental conditions. Previously, it was reported in oleaginous *Y. lipolytica* that the deletion of mMAE in the wild type strain has no effect on lipid content or fatty acid profiles (Blazeck et al. 2014; Dulermo et al. 2015; Zhu et al. 2023). However, only the deletion of cMAE was studied in the present work.

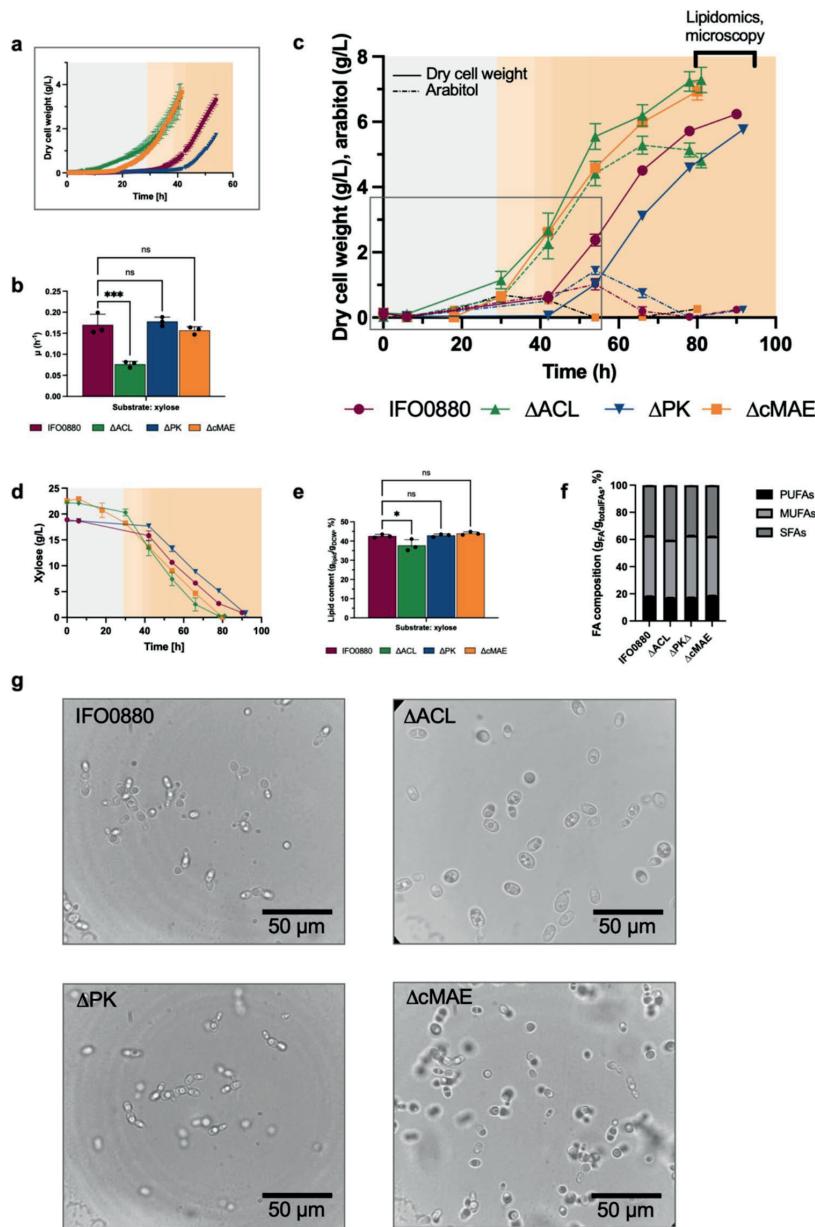
The fatty acid composition of the wild type *R. toruloides* IFO0880 was C16:0 22.67% and C18:1 45.62% ($g_{FA}/g_{totalFA}$) (Table 3), closely similar to as previously reported (Ratledge and Wynn 2002). In fact, our results were closely similar to the *R. glutinis* strain, which has more recently been designated as *R. toruloides*, and according to Zhang et al. may even be IFO0880 (Zhang et al. 2016b). The knockout of the ACL caused changes in the fatty acid composition. We observed a relative increase in polyunsaturated fatty acids (C18:2 and C18:3) of 34% of total fatty acids when compared to the wild type strain ($g_{FA}/g_{totalFA}$) (Fig. 2f). In the ACL knockout, the fraction of oleic acid (C18:1) decreased significantly compared to the wild type strain and constituted only 19.47% (p val < 0.0001), while linoleic acid (C18:2) increased significantly compared to the wild type and became the largest fraction with 29.46% (p val < 0.0001), followed by 17.78% of alpha-linoleic acid (C18:3) (Table 3, p val < 0.0001). Similar observations were reported in *Y. lipolytica* (Dulermo et al. 2015). It should be noted, however, that when calculated per DCW, the total quantities of all measured fractions decreased compared to the wild type strain (Supplementary Table S3).

Yeasts and filamentous fungi undergo morphological differentiations when triggered by extracellular stimuli such as nutrient limitation. Therefore, we examined cells harvested at the end of the experiment under the microscope. We observed that the majority of strains, including the wild type, had adopted linear chain morphology (Fig. 2g), notably different from their typical oval-shaped unicellular form (Supplementary Fig. S4). Similar morphological appearance, also known as pseudohyphae formation, has been observed at slow growth rates below $0.1 h^{-1}$ in glucose-limited chemostats of the non-conventional yeast *Komagataella phaffii* (Rebnegger et al. 2014) and *S. cerevisiae* under the nitrogen limitation (Gimeno et al. 1992). However, as a result

Table 3 Bioreactor data and fatty acid composition ($g_{FA}/g_{totalFA}$, %) measured by GC-FID for strains grown in the glucose-based (20 g/L) pH-adjusted chemically defined low-nitrogen medium (C/N of 80). C16:0 palmitic acid, C18:0 stearic acid, C18:1(c9) oleic acid,

C18:2(n6) linoleic acid, C18:3(n3) alpha-linolenic acid, and *FAs* fatty acids. Mean and standard deviation are calculated from three biological replicates

Strain	Dry cell weight (g/L)	Lipid content (g_{lipid}/g_{DCW} , %)	Major fatty acid residues ($g_{FA}/g_{totalFAs}$, %)						Harvest time (h)
			C16:0	C18:0	C18:1(c9)	C18:2(n6)	C18:3(n3)	Other FAs	
IFO0880 (NBRC 0880)	10.23 ± 0.15	50.18 ± 0.37	22.67 ± 0.14	12.62 ± 0.06	45.62 ± 0.11	11.38 ± 0.14	3.58 ± 0.01	4.13	54
Δ PK	10.23 ± 0.15	50.52 ± 0.94	22.70 ± 0.10	12.64 ± 0.12	45.73 ± 0.14	11.29 ± 0.06	3.50 ± 0.01	4.14	54
Δ ACL	2.34 ± 0.065	9.43 ± 0.039	16.25 ± 0.03	9.42 ± 0.09	19.47 ± 0.49	29.46 ± 0.28	17.78 ± 0.16	7.61	54
Δ cMAE	9.62 ± 0.26	51.64 ± 0.50	22.80 ± 0.10	11.72 ± 0.07	46.04 ± 0.12	11.49 ± 0.12	3.90 ± 0.06	4.04	48



of the knockout of the ACL gene, cells did not form these filament-like structures (Fig. 2g), instead their morphology looked very similar to that during the exponential growth phase (Supplementary Fig. S4), in a good agreement with the above analysis that the loss of ACL indeed deprived cells from consuming carbon under nitrogen limitation.

Xylose as a carbon source modifies the cellular response to ACL loss

R. toruloides can consume xylose as a sole carbon source (Pinheiro et al. 2020). DCW of 18 (g/L) on xylose as a carbon source was comparable to the one grown on glucose, 22 (g/L), but the DCW yield on substrate ($g_{DCW}/g_{substrate}$) even

Fig. 3 Physiological parameters obtained in xylose-based pH-adjusted low-nitrogen mineral medium (C/N ratio 80) (Verduyn et al. 1992). Data obtained from cultivating *R. toruloides* wild type IFO0880 (circle), ATP citrate lyase knockout (Δ ACL, pyramid), phosphoketolase knockout (Δ PK, inverted pyramid), and cytosolic malic enzyme knockout (Δ cMAE, square) strains in falcon tube bioreactors with non-invasive OD, dO_2 , and pH sensors. **a** A snapshot of the online OD_{600} data displayed as the dry cell weight (g/L), **b** maximum specific growth rate fitted from the exponential growth phase OD_{600} data (h^{-1}), **c** dry cell weight and arabitol (g/L), **d** substrate concentration (g/L), **e** lipid content (g_{lipid}/g_{DCW} , %); cells harvested at the end of cultivation, **f** relative fatty acid composition ($g_{FA}/g_{totalFAs}$, %); cells harvested at the end of cultivation, **g** morphology comparison of wild type IFO0880 versus knockout strains visualized by bright-field microscope; cells harvested at the end of cultivation. Panels **a** and **c** background colors denote the switch in growth phases based on dO_2 curves available from Supplementary Fig. S2, presumably pointing to the onset of the nitrogen limitation phase. *FAs* fatty acids, *SFAs* saturated fatty acids, *MUFAs* monounsaturated fatty acids, *PUFAs* polyunsaturated fatty acids. Error bars are calculated as standard deviation from three biological experiments. Asterisks denote statistical significance (ANOVA Dunnett's multiple comparison test against the wild type IFO0880 strain, adjusted p value $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$). ns is used to denote changes that are statistically nonsignificant

outperformed glucose (Pinheiro et al. 2020). Xylose is a relevant substrate of *R. toruloides* that has been less explored; therefore, we cultivated Δ PK, Δ ACL, and Δ cMAE strains in the low-nitrogen pH-adjusted chemically defined medium, but this time containing xylose as a sole carbon source. All strains demonstrated slightly different lag phases, but the ACL knockout exhibited the shortest lag phase and the lowest specific growth rate compared to the wild type strain (Fig. 3a, Supplementary Table S2). The specific growth rate (μ) was 45% of the wild type strain (Fig. 3b, $apval < 0.001$). Nevertheless, ACL knockout reached even a significantly higher final DCW titer as compared to the wild type strain (Fig. 3c, Supplementary Fig. S3, b, $apval < 0.01$) and consumed xylose until depletion (Fig. 3d). The non-lipid DCW

titors of the ACL knockout were higher compared to the wild type strain ($apval < 0.01$), suggesting that ACL was involved in other cellular processes apart from lipid synthesis (Supplementary Fig. S3b). Lipid content of the wild type strain was $42.61 \pm 0.84\%$ (g_{lipid}/g_{DCW}) after 90 h of growth on low-nitrogen mineral medium (Table 4), but the ACL knockout gave significantly lower, $37.71 \pm 3.01\%$ (g_{lipid}/g_{DCW}) (Fig. 3e, $apval 0.02$), with no major changes in the fatty acid profiles (Fig. 3f, Table 4). Linear chain pseudohyphal growth was unchanged compared to the glucose medium. Chain-like pseudohyphal growth was observed for all strains, except for the ACL knockout (Fig. 3g). The knockout of the cMAE gene did not have an effect on growth or lipid synthesis on medium containing xylose as a carbon source (Fig. 3a–f).

Altogether, these results indicated that *R. toruloides* IFO0880 is less reliant on ACL than assumed, which suggests an alternative source of cytosolic acetyl-CoA on xylose as sole carbon source. It cannot be ruled out that the PK replaced ACL for the supply of acetyl-CoA upon the ACL knockout, but on the other hand, the knockout of the PK gene alone did not give any change in the specific growth rate or lipid content (Fig. 3a–f). We speculate that one of the options for the cytosolic acetyl-CoA synthesis could be the pyruvate-acetaldehyde-acetate pathway, also known as “pyruvate dehydrogenase bypass.” Despite low protein expression levels of ACS on glucose and xylose, it was previously reported higher abundant on xylose during the exponential growth (p value 0.038) and lipid accumulation phase (p value 0.49) compared to glucose-grown cells (Rekēna et al. 2023).

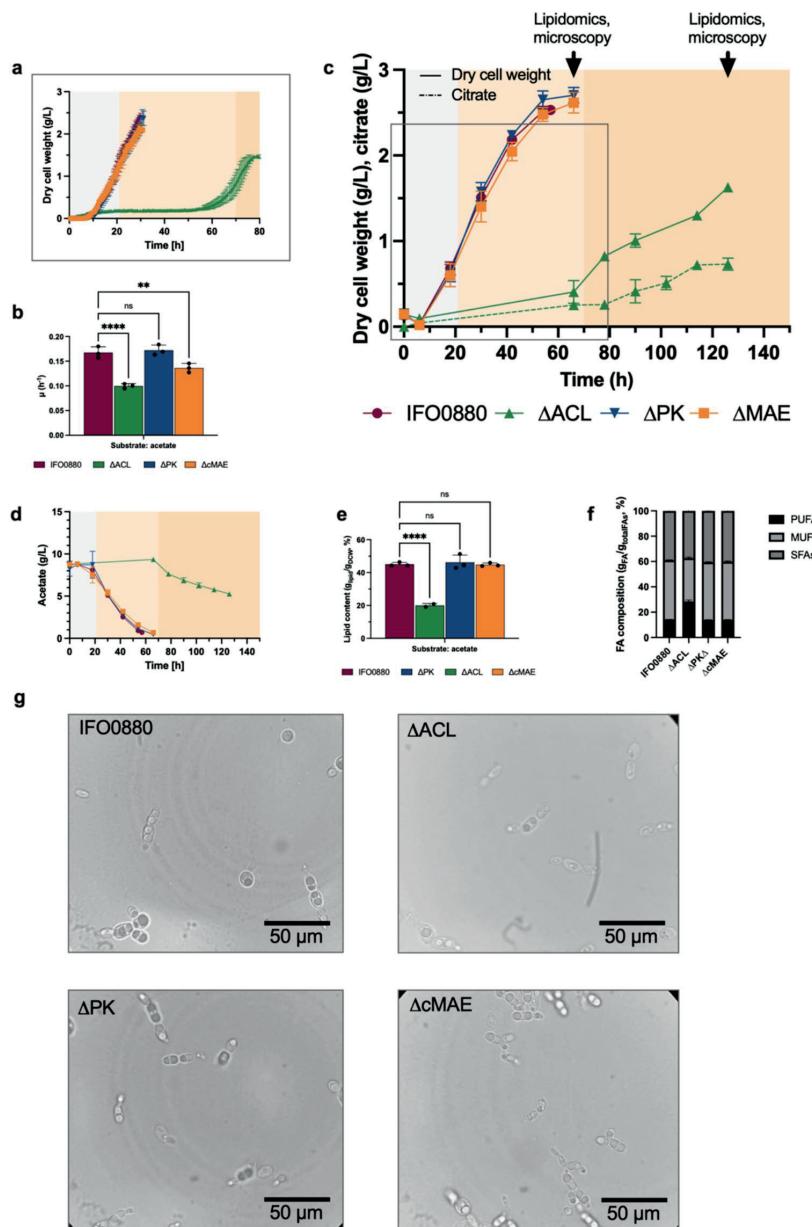
ACL loss demonstrates negative effects on cell growth and lipid synthesis on acetate

Previous reports have shown differences in growth response on acetate caused by loss of ACL. In pathogenic

Table 4 Bioreactor data and fatty acid composition ($g_{FA}/g_{totalFAs}$, %) measured by GC-FID for strains grown in the xylose-based (20 g/L) pH-adjusted chemically defined low-nitrogen medium (C/N of 80). *C16:0* palmitic acid, *C18:0* stearic acid, *C18:1(c9)* oleic acid,

C18:2(n6) linoleic acid, *C18:3(n3)* alpha-linolenic acid, *FAs* fatty acids. Mean and standard deviation are calculated from three biological replicates

Strain	Dry cell weight (g/L)	Lipid content (g_{lipid}/g_{DCW} , %)	Major fatty acid residues ($g_{FA}/g_{totalFAs}$, %)						Harvest time (h)
			C16:0	C18:0	C18:1(c9)	C18:2(n6)	C18:3(n3)	Other FAs	
IFO0880 (NBRC 0880)	6.24 ± 0.13	42.61 ± 0.84	21.80 ± 0.03	11.42 ± 0.03	43.52 ± 0.06	14.65 ± 0.17	3.30 ± 0.03	5.31	90
Δ PK	5.76 ± 0.075	42.94 ± 0.73	21.28 ± 0.19	11.61 ± 0.19	44.44 ± 0.24	13.70 ± 0.11	3.40 ± 0.03	5.57	91
Δ ACL	7.28 ± 0.39	37.71 ± 3.01	28.27 ± 0.09	9.47 ± 0.16	41.05 ± 0.05	14.67 ± 0.18	2.18 ± 0.07	4.35	81
Δ cMAE	6.93 ± 0.27	43.94 ± 0.82	22.21 ± 0.11	11.36 ± 0.06	42.62 ± 0.15	14.85 ± 0.18	3.51 ± 0.02	5.45	80



basidiomycetous yeast *C. neoformans*, the knockout of the ACL1 gene caused growth defects on glucose, but no effect during growth on acetate (Griffiths et al. 2012). In oleaginous ascomycetous yeast *L. starkeyi*, a knockout of both ACL1 and ACL2 gene resulted in restored growth on acetate, but TAG productivity was lower than that of the control using acetate with glucose (Sato et al. 2021). In

accordance with these reports, we cultivated Δ PK, Δ ACL strains, and IFO0880 in low-nitrogen chemically defined medium with acetate as a sole carbon source. The initial medium pH was adjusted to 6 by adding KOH and was not maintained. During the cultivation, the pH in all cultures rose equally, as the cells started growing, and reached pH 9 at the end of cultivation (Supplementary

Fig. 4 Physiological parameters obtained in acetate-based pH 7-adjusted low-nitrogen mineral medium (C/N ratio 80) (Verduyn et al. 1992). Data obtained from cultivating *R. toruloides* wild type IFO0880 (circle), ATP citrate lyase knockout (Δ ACL, pyramid), phosphoketolase knockout (Δ PK, inverted pyramid), and cytosolic malic enzyme knockout (Δ cMAE, square) strains in falcon tube bioreactors with non-invasive OD, dO_2 , and pH sensors. **a** A snapshot of the online OD₆₀₀ data displayed as the dry cell weight (g/L), **b** maximum specific growth rate fitted from the exponential growth phase OD₆₀₀ data (h⁻¹), **c** dry cell weight and citrate (g/L), **d** substrate concentration (g/L), **e** lipid content (g_{lipid}/g_{DCW}, %); cells harvested at the end of cultivation, **f** relative fatty acid composition (g_{FA}/g_{totalFA}, %); cells harvested at the end of cultivation, **g** morphology comparison of wild type IFO0880 versus knockout strains visualized by bright-field microscope; cells harvested at the end of cultivation. Panels **a** and **c** background colors denote the switch in growth phases based on dO_2 curves available from Supplementary Fig. S2, presumably pointing to the onset of the nitrogen limitation phase. FAs fatty acids. SFAs saturated fatty acids, MUFA monounsaturated fatty acids, PUFAs polyunsaturated fatty acids. Error bars are calculated as a standard deviation from three biological experiments. Asterisks denote statistical significance (ANOVA Dunnett's multiple comparison test against the wild type IFO0880 strain, adjusted p value * p <0.05, ** p <0.01, *** p <0.001, **** p <0.0001). ns is used to denote changes that are statistically nonsignificant

Fig. S1). The growth of Δ ACL was extremely delayed as compared to the other strains (Fig. 4a), and the maximum specific growth rate was reduced by 40% in comparison to wild type strain (Fig. 4b, $apval$ 0.009), falling to 0.10 ± 0.0047 h⁻¹ (Supplementary Table S2). The ACL knockout started growing only at 60 h and 6 h after that started to produce slight amounts of citrate, a substrate for ACL (Fig. 4c). Although the analysis showed a consistent, slow consumption of acetate after 66 h, Δ ACL cells stopped respiring before all the acetate was depleted around 126 h (Fig. 4d, Supplementary Fig. S1). Other strains consumed the substrate completely with no visible disturbance to the growth rates. These results demonstrated that *R. toruloides* responded similarly to ACL loss compared to glucose medium, and acetate did not restore the growth defects caused by ACL loss. It was in contrast to findings by Hynes and Murray, in which the authors concluded that external sources of cytoplasmic acetyl-CoA result in repression of ACL (Hynes and Murray 2010). The loss of ACL caused 56% reduction in lipids (g_{lipid}/g_{DCW}) (Fig. 4e, $apval$ <0.0001) and alterations in fatty acid profiles (Fig. 4f). On acetate, ACL loss did not result in significant changes to non-lipid DCW (Supplementary Fig. S3c), indicating that the effect was directly on lipid synthesis and not beyond as in case of the glucose-grown cells. The majority fatty acid fraction, C18:1, decreased relatively from 46 to 33% (g_{FA}/g_{totalFA}), while C18:2 and C18:3 increased relatively to 20% and 8%, respectively, 20% of C18:2 and 8% of C18:3 (g_{FA}/g_{totalFA}) (Table 5). These results were similar to the *L. starkeyi* study, in which the TAG productivity of the ACL knockout

strain was significantly lower compared to the reference strain on medium containing glucose and acetate (Sato et al. 2021). It also indicated that acetate must be converted into citrate for its assimilation, which is possible if mitochondrial malate-oxoglutarate and oxoglutarate/citrate shuttling takes place.

cMAE knockout gave a 19% decrease in the specific growth rate, falling to 0.14 ± 0.0091 h⁻¹ ($apval$ 0.009), even though it did not affect the lipid content (Table 5). These results suggested that cMAE is not involved in lipogenesis on acetate. Consistent with the above results, the knockout of the PK gene caused no change in growth or lipid synthesis (Fig. 4a-f).

Growth on acetate induced differences in cell morphology in this experiment. Cells formed less chain-like structures compared to glucose or xylose-grown cells, even though they were under nutrient starvation (Fig. 4g).

Discussion

R. toruloides is an emerging oleaginous cell factory for the production of fats, oils, and oleochemicals. The metabolic mechanisms of lipid production in *R. toruloides* are less understood compared to other oleaginous fungi *M. circinelloides* or *M. alpina*, oleaginous yeast *Y. lipolytica*, or model yeast *S. cerevisiae*. Due to the development of gene editing tools (Otopual et al. 2019; Liu et al. 2019; Jiao et al. 2019; Schultz et al. 2019), the possibilities for functional genomics studies are significantly improved. In this study, we applied CRISPR/Cas9 genome editing system to elucidate the role of PK pathway, ACL, and cMAE in lipid biosynthesis of *R. toruloides* IFO0880 by creating gene knockouts and characterizing them on alternative carbon sources.

The results of this study demonstrated that ACL is crucial for not only the fatty acid synthesis but also for growth in *R. toruloides*. Loss of ACL reduced lipid content by 81% on xylose and 56% on acetate (Figs. 2, 3, and 4). It also significantly reduced the specific growth rate on all carbon sources. Although previous studies have reported growth defects as a result of ACL inactivation in *A. niger*, *A. nidulans*, *S. macrospora*, and *Y. lipolytica* (Nowrousian et al. 1999; Hynes and Murray 2010; Chen et al. 2014; Dulermo et al. 2015), the present study revealed a difference to *Y. lipolytica*. ACL knockout resulted in a growth arrest on glucose and on acetate once nitrogen is depleted, as demonstrated by a significantly lower cell density compared to the reference strain by the end of experiment (see Figs. 2 and 4). On glucose, it could not have been caused by the pH changes, because it remained within the optimal range (pH 5.5–7.0) throughout cultivation (Supplementary Fig. S1). Currently, there is not enough evidence to provide a single explanation

Table 5 Bioreactor data and fatty acid composition ($g_{FA}/g_{totalFAs}$, %) measured by GC-FID for strains grown in the acetate-based (20 g/L) chemically defined low-nitrogen medium (C/N of 80). C16:0 palmitic

Strain	Dry cell weight (g/L)	Lipid content (g_{lipid}/g_{DCW} , %)	Major fatty acid residues ($g_{FA}/g_{totalFAs}$, %)						Harvest time (h)
			C16:0	C18:0	C18:1(c9)	C18:2(n6)	C18:3(n3)	Other FAs	
IFO0880 (NBRC 0880)	2.53 ± 0.03	45.07 ± 1.15	17.30 ± 0.26	18.35 ± 0.12	46.04 ± 0.31	9.83 ± 0.02	4.07 ± 0.08	4.41	54
ΔPK	2.70 ± 0.09	46.23 ± 4.45	18.53 ± 0.37	18.64 ± 0.19	44.81 ± 0.47	9.40 ± 0.13	4.14 ± 0.09	4.48	66
ΔACL	1.63 ± 0.03	20.02 ± 1.18	17.27 ± 0.04	15.69 ± 0.37	33.06 ± 0.91	19.69 ± 0.83	8.10 ± 0.39	6.21	66
ΔcMAE	2.62 ± 0.12	44.79 ± 1.05	18.25 ± 0.45	18.48 ± 0.07	45.09 ± 0.65	9.49 ± 0.04	4.19 ± 0.07	4.49	126

to our observations. Lowered acetyl-CoA in cytoplasm impairs the synthesis of signaling molecules and secondary metabolites leading to developmental effects (Hynes and Murray 2010). In yeast, acetyl-CoA is localized in four different compartments, and nuclear acetyl-CoA is required for histone acetylation (Takahashi et al. 2006; Pietrocola et al. 2015). Recently, the role of ACL in supplying acetyl-CoA for histone acetylation to promote proliferation was demonstrated in myeloid cells (Rhee et al. 2019). Another consequence of the ACL knockout is likely impaired citrate conversion to cytosolic acetyl CoA that probably leads to citrate oxidation in mitochondria and an increased flux through the TCA cycle increasing the production of substrates for the electron transport chain, which was confirmed by proving elevated mitochondrial membrane potential between control and ACL knockdown cells in hematopoietic murine cells (Board and Newsholme 1996; Bauer et al. 2005). It is in line with the detected extracellular citrate on acetate (Fig. 4c). In this study, we also observed an effect on cell morphology during nitrogen limitation. In *S. cerevisiae*, it has been investigated that pseudohyphal growth is regulated by two different signaling pathways, the MAP kinase cascade and cAMP-dependent pathway (Gancedo 2001). In mammalian cells, ACL is hierarchically regulated through cAMP-dependent phosphorylation (Pierce et al. 1981; Pant et al. 2023). The observed morphological changes, growth arrest during nitrogen starvation, and the drop in lipid production that were all associated with the ACL loss suggest a complex regulatory network regulation of ACL in *R. toruloides*. The comparison of lipid content to ACL loss in glucose, xylose, and acetate media indicated very interesting carbon-source dependent differences in metabolic regulation and the existence of alternative routes to cytosolic acetyl-CoA. To our best knowledge, this is the first report characterizing ACL knockout in yeast in the xylose medium. While these results are more in agreement with what was reported earlier in *Y. lipolytica* on glucose, further studies are required to understand why in *R. toruloides* the knockout responses on xylose

and glucose were different. On a preliminary level, we speculated that upon knockout of ACL on xylose, the acetyl-CoA synthetase (ACS) could be used for the synthesis of cytosolic acetyl-CoA resulting in higher lipid content, while in glucose and acetate media the main role was staying with the ACL. It is known that ACS is transcriptionally regulated (Chen et al. 2012) and the only source of cytosolic acetyl-CoA (Pronk et al. 1996) in *S. cerevisiae*. Large differences in signaling pathways under the consumption of xylose have already been reported, but mainly in *S. cerevisiae*. Assimilation of xylose is weakly sensed by the intracellular branch of the cAMP/PKA pathway in yeast (Brink et al. 2021).

In this study, we showed that the lipid production was not changed as a result of the PK knockout (Figs. 2, 3, and 4). Unintuitively, our result also did not support the evidence from early biochemical studies on xylose as a carbon source. Compared to the alternative biosynthetic pathways for cytosolic acetyl-CoA production, the PK route bypasses the decarboxylation step of pyruvate into CO_2 and acetaldehyde (catalyzed by pyruvate decarboxylase) and the ATP expenditure for the activation of acetate by ACS or citrate by ACL, thus potentially increasing the acetyl-CoA yield. We thought it was the reason why previously stoichiometric genome-scale modeling predicted the PK route for the utilization of glucose or xylose in *R. toruloides* (Lopes et al. 2020; Rekēna et al. 2023) being in the agreement with previous enzyme abundance studies (Kim et al. 2021; Rekēna et al. 2023). But our results did not support this hypothesis on either of the substrates. Based on existing genetic engineering studies, we speculated that the metabolic flux can be rerouted through PK in a combination with silencing the expression of other enzymes in a close proximity. In *Y. lipolytica*, expression of PK worked to improve lipid production in phosphofructokinase (PFK) deletion background (Kamineni et al. 2021). Introduction of PK pathway was demonstrated in TKT and TAL deletion background to produce 3HP in *S. cerevisiae* (Hellgren et al. 2020). On the other hand, *R. toruloides* PK K_m value would potentially explain why the PK is not critical

for lipid synthesis in the wild type strain. Alternatively, the bottleneck could be PTA or acetate kinase (ACK). Overexpression of non-native PTA improved lipid titers in *R. toruloides* NP11 by 15.1% (Yang et al. 2018), but the role of ACK in *R. toruloides* still needs to be studied. In non-oleaginous *A. niger*, which has a native PK pathway, heterologous PK pathway expression increased target product synthesis, while the deletion of ACK decreased it (Liu et al. 2023). Also, it is worth noting that in the enzyme database BRENDA, the PTA reaction equation denotes acetyl-CoA as a substrate and acetyl-phosphate as a product. The thermodynamic feasibility of the pathway is unclear.

Finally, the results of this study suggested that cMAE plays no significant role in lipogenesis in *R. toruloides* IFO0880 and its presence has an adverse effect when growing on glucose. It was in good agreement with the early biochemical studies with *R. glutinis* (Yoon et al. 1984). Compared to the catabolic pathways of glucose or xylose, acetate assimilation requires less enzymatic conversions to synthesize the substrate of malic enzyme and, perhaps, the reason why stoichiometric genome-scale models have predicted the use of cMAE for NADPH regeneration upon assimilation of acetate (Lopes et al. 2020; Rekēna et al. 2023). Our results confirmed that cytosolic malic enzyme is not critical for lipogenesis, suggesting that alternative routes for NADPH regeneration can be used, likely enzymes from the oxidative pentose phosphate pathway. Recent genetic engineering studies are available only on glucose. Lipogenesis induction upon a knockout was reported by Dulermo et al., when a mMAE in *Y. lipolytica* was deleted (Dulermo et al. 2015); although *Y. lipolytica* does not have a cMAE. Nevertheless, there is a slight difference that in wild type background, such an effect was not observed, but in an engineered strain for lipid production (“Obese”), the adverse effect was observed. Perhaps we can say that according to our evidence, a wild type background is an even stronger argument for the cMAE non-essentiality in lipogenesis. But it is worth noting that overexpression of native cMAE in *R. toruloides* IFO0880 (same strain) increased lipid titers by 28% (Zhang et al. 2016a). Authors noted that this was a relatively minor increase compared to other enzymes that were introduced, but it shows that cMAE can play a role in lipogenesis of *R. toruloides* under certain conditions (e.g., overexpression).

In conclusion, we showed that ACL is essential for lipid synthesis and cell growth in *R. toruloides*. The cellular response to ACL loss was affected by the carbon source present in cultivation media. The results also demonstrated that cytosolic malic enzyme is not critically involved in lipid synthesis in *R. toruloides* and in some conditions is even disfavorable. The characterization of the PK knockout suggested that it is not critical for the lipid synthesis. This work

is useful for future metabolic engineering strategies of *R. toruloides*.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00253-025-13454-w>.

Acknowledgements We thank Dr. Peter Otoopal for providing us the pPBO.202 plasmid. We also acknowledge Olesia Duda for the fatty acid analysis.

Author contribution AR and PJL conceptualized and designed the study. AR and KP conducted experiments. SG contributed to genetic engineering methodology. AR analyzed data and wrote the manuscript. All authors read and approved the manuscript.

Funding This study was funded by the Estonian Research Council Grant PRG1011 and European Union's HORIZON Teaming for Excellence program under grant agreement 101060066.

Data availability The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Ageitos JM, Vallejo JA, Veiga-Crespo P, Villa TG (2011) Oily yeasts as oleaginous cell factories. *Appl Microbiol Biotechnol* 90:1219–1227. <https://doi.org/10.1007/s00253-011-3200-z>
- Bauer DE, Hatzivassiliou G, Zhao F, Andreadis C, Thompson CB (2005) ATP citrate lyase is an important component of cell growth and transformation. *Oncogene* 24:6314–6322. <https://doi.org/10.1038/sj.onc.1208773>
- Beopoulos A, Nicaud J-M, Gaillardin C (2011) An overview of lipid metabolism in yeasts and its impact on biotechnological processes. *Appl Microbiol Biotechnol* 90:1193–1206. <https://doi.org/10.1007/s00253-011-3212-8>
- Blazeck J, Hill A, Liu L, Knight R, Miller J, Pan A, Otoopal P, Alper HS (2014) Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nat Commun* 5:3131. <https://doi.org/10.1038/ncomms4131>

Board M, Newsholme E (1996) Hydroxycitrate causes altered pyruvate metabolism by tumorigenic cells. *IUBMB Life* 40:1047–1056. <https://doi.org/10.1080/15216549600201683>

Bonturi N, Pinheiro MJ, De Oliveira PM, Rusadze E, Eichinger T, Liudziutė G, De Biaggi JS, Brauer A, Remm M, Miranda EA, Ledesma-Amaro R, Lahtvee P-J (2022) Development of a dedicated Golden Gate Assembly Platform (RtGGA) for *Rhodotorula toruloides*. *Metab Eng Commun* 15:e00200. <https://doi.org/10.1016/j.mec.2022.e00200>

Botham PA, Ratledge C (1979) A biochemical explanation for lipid accumulation in *Candida 107* and other oleaginous microorganisms. *J Gen Microbiol* 114:361–375. <https://doi.org/10.1099/00221287-114-2-361>

Boulton CA, Ratledge C (1981) Correlation of lipid accumulation in yeasts with possession of ATP: citrate lyase. *Microbiology* 127:169–176. <https://doi.org/10.1099/00221287-127-1-169>

Brink DP, Borgström C, Persson VC, Ofuji Osiro K, Gorwa-Grauslund MF (2021) D-xylose sensing in *Saccharomyces cerevisiae*: insights from D-glucose signaling and native D-xylose utilizers. *Int J Mol Sci* 22:12410. <https://doi.org/10.3390/ijms22212410>

Cao M, Tran VG, Qin J, Olson A, Mishra S, Schultz JC, Huang C, Xie D, Zhao H (2022) Metabolic engineering of oleaginous yeast rhodotorula toruloides for overproduction of triacetic acid lactone. *Biotechnol Bioeng* 119:2529–2540. <https://doi.org/10.1002/bit.28159>

Chattopadhyay A, Mitra M, Maiti MK (2021) Recent advances in lipid metabolic engineering of oleaginous yeasts. *Biotechnol Adv* 53:107722. <https://doi.org/10.1016/j.biotechadv.2021.107722>

Chawla K, Kaur S, Kaur R, Bhunia RK (2022) Metabolic engineering of oleaginous yeasts to enhance single cell oil production. *J Food Process Eng* 45:e13634. <https://doi.org/10.1111/jfpe.13634>

Chen Y, Siewers V, Nielsen J (2012) Profiling of cytosolic and peroxisomal acetyl-CoA metabolism in *Saccharomyces cerevisiae*. *PLoS ONE* 7:e42475. <https://doi.org/10.1371/journal.pone.0042475>

Chen H, He X, Geng H, Liu H (2014) Physiological characterization of ATP-citrate lyase in *Aspergillus niger*. *J Ind Microbiol Biotechnol* 41:721–731. <https://doi.org/10.1007/s10295-014-1418-3>

Coradetti ST, Pinel D, Geisenle GM, Ito M, Mondo SJ, Reilly MC, Cheng Y-F, Bauer S, Grigoriev IV, Gladden JM, Simmons BA, Brem RB, Arkin AP, Skerker JM (2018) Functional genomics of lipid metabolism in the oleaginous yeast *Rhodosporidium toruloides*. *eLife* 7:e32110. <https://doi.org/10.7554/eLife.32110>

Dinh HV, Suthers PF, Chan SHJ, Shen Y, Xiao T, Deewan A, Jagtap SS, Zhao H, Rao CV, Rabinowitz JD, Maranas CD (2019) A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data. *Metab Eng Commun* 9:e00101. <https://doi.org/10.1016/j.mec.2019.e00101>

Donzella S, Cucchetti D, Capusoni C, Rizzi A, Galafassi S, Chiara G, Compagno C (2019) Engineering cytoplasmic acetyl-CoA synthesis decouples lipid production from nitrogen starvation in the oleaginous yeast *Rhodosporidium azoricum*. *Microb Cell Factories* 18:199. <https://doi.org/10.1186/s12934-019-1250-6>

Dourou M, Aggeli D, Papanikolaou S, Aggelis G (2018) Critical steps in carbon metabolism affecting lipid accumulation and their regulation in oleaginous microorganisms. *Appl Microbiol Biotechnol* 102:2509–2523. <https://doi.org/10.1007/s00253-018-8813-z>

Dulermo T, Lazar Z, Dulermo R, Rakicka M, Haddouche R, Nicaud J-M (2015) Analysis of ATP-citrate lyase and malic enzyme mutants of *Yarrowia lipolytica* points out the importance of mannitol metabolism in fatty acid synthesis. *Biochim Biophys Acta BBA - Mol Cell Biol Lipids* 1851:1107–1117. <https://doi.org/10.1016/j.bbapli.2015.04.007>

Evans CT, Ratledge C (1984) Induction of xylulose-5-phosphate phosphotolase in a variety of yeasts grown on d-xylose: the key to efficient xylose metabolism. *Arch Microbiol* 139:48–52. <https://doi.org/10.1007/BF00692711>

Fatland BL, Ke J, Anderson MD, Mentzen WI, Cui LW, Allred CC, Johnston JL, Nikolau BJ, Wurtele ES (2002) Molecular characterization of a heteromeric ATP-citrate lyase that generates cytosolic acetyl-coenzyme A in an arabiopsis. *Plant Physiol* 130:740–756. <https://doi.org/10.1104/pp.008110>

Gancedo JM (2001) Control of pseudohyphophore formation in *Saccharomyces cerevisiae*. *FEMS Microbiol Rev* 25:107–123. <https://doi.org/10.1111/j.1574-6976.2001.tb00573.x>

Gimeno CJ, Ljungdahl PO, Styles CA, Fink GR (1992) Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell* 68:1077–1090. [https://doi.org/10.1016/0092-8674\(92\)90079-R](https://doi.org/10.1016/0092-8674(92)90079-R)

Gong G, Wu B, Liu L, Li J, He M (2024) Engineering oleaginous red yeasts as versatile chassis for the production of oleochemicals and valuable compounds: current advances and perspectives. *Biotechnol Adv* 108432. <https://doi.org/10.1016/j.biotechadv.2024.108432>

Griffiths EJ, Hu G, Fries B, Caza M, Wang J, Gsponer J, Gates-Hollingsworth MA, Kozel TR, De Repentigny L, Kronstad JW (2012) A defect in ATP-citrate lyase links acetyl-CoA production, virulence factor elaboration and virulence in *Cryptococcus neoformans*. *Mol Microbiol* 86:1404–1423. <https://doi.org/10.1111/mmi.12065>

Hellgren J, Godina A, Nielsen J, Siewers V (2020) Promiscuous phosphotolase and metabolic rewiring enables novel non-oxidative glycolysis in yeast for high-yield production of acetyl-CoA derived products. *Metab Eng* 62:150–160. <https://doi.org/10.1016/j.ymben.2020.09.003>

Hynes MJ, Murray SL (2010) ATP-citrate lyase is required for production of cytosolic acetyl coenzyme A and development in *Aspergillus nidulans*. *Eukaryot Cell* 9:1039–1048. <https://doi.org/10.1128/EC.00080-10>

Jiao X, Zhang Y, Liu X, Zhang Q, Zhang S, Zhao ZK (2019) Developing a CRISPR/Cas9 system for genome editing in the basidiomycetous yeast *Rhodosporidium toruloides*. *Biotechnol J* 14:1900036. <https://doi.org/10.1002/biot.201900036>

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821. <https://doi.org/10.1126/science.1225829>

Kamieni A, Consiglio AL, MacEwen K, Chen S, Chifamba G, Shaw AJ, Tsakraklides V (2021) Increasing lipid yield in *Yarrowia lipolytica* through phosphotolase and phosphotransacetylase expression in a phosphofructokinase deletion strain. *Biotechnol Biofuels* 14:113. <https://doi.org/10.1186/s13068-021-01962-6>

Kim J, Coradetti ST, Kim Y-M, Gao Y, Yaegashi J, Zucker JD, Munoz N, Zink EM, Burnum-Johnson KE, Baker SE, Simmons BA, Skerker JM, Gladden JM, Magnuson JK (2021) Multi-omics driven metabolic network reconstruction and analysis of lignocellulosic carbon utilization in *Rhodosporidium toruloides*. *Front Bioeng Biotechnol* 8:612832. <https://doi.org/10.3389/fbioe.2020.612832>

Koh CMJ, Liu Y, Du Moehninsi, M, Ji L (2014) Molecular characterization of KU70 and KU80 homologues and exploitation of a KU70-deficient mutant for improving gene deletion frequency in *Rhodosporidium toruloides*. *BMC Microbiol* 14:50. <https://doi.org/10.1186/1471-2180-14-50>

Labuhn M, Adams FF, Ng M, Knoess S, Schambach A, Charpentier EM, Schwarzer A, Mateo JL, Klusmann J-H, Heckl D (2018) Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic Acids Res* 46:1375–1385. <https://doi.org/10.1093/nar/gkx1268>

Liu X-Y, Chi Z, Liu G-L, Madzak C, Chi Z-M (2013) Both decrease in ACL1 gene expression and increase in ICL1 gene expression in marine-derived yeast *Yarrowia lipolytica* expressing INU1

gene enhance citric acid production from inulin. *Mar Biotechnol* 15:26–36. <https://doi.org/10.1007/s10126-012-9452-5>

Liu X, Zhang Y, Liu H, Jiao X, Zhang Q, Zhang S, Zhao ZK (2019) RNA interference in the oleaginous yeast *Rhodosporidium toruloides*. *FEMS Yeast Res* 19:foz031. <https://doi.org/10.1093/femsyr/foz031>

Liu J, Zhang S, Li W, Wang G, Xie Z, Cao W, Gao W, Liu H (2023) Engineering a phosphoketolase pathway to supplement cytosolic acetyl-CoA in *Aspergillus niger* enables a significant increase in citric acid production. *J Fungi* 9:504. <https://doi.org/10.3390/jof9050504>

Lööke M, Kristjuhan K, Kristjuhan A (2011) Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques* 50:325–328. <https://doi.org/10.2144/000113672>

Lopes HJS, Bonturi N, Kerkhoven EJ, Miranda EA, Lahtvee P-J (2020) C/N ratio and carbon source-dependent lipid production profiling in *Rhodotorula toruloides*. *Appl Microbiol Biotechnol* 104:2639–2649. <https://doi.org/10.1007/s00253-020-10386-5>

Monteiro de Oliveira P, Aborneva D, Bonturi N, Lahtvee P-J (2021) Screening and growth characterization of non-conventional yeasts in a hemicellulosic hydrolysate. *Front Bioeng Biotechnol* 9. <https://doi.org/10.3389/fbioe.2021.659472>

Niehus X, Crutz-Lec Coq A-M, Sandoval G, Nicaud J-M, Ledesma-Amaro R (2018) Engineering *Yarrowia lipolytica* to enhance lipid production from lignocellulosic materials. *Biotechnol Biofuels* 11:11. <https://doi.org/10.1186/s13068-018-1010-6>

Nowrouzian M, Masloff S, Pöggeler S, Kück U (1999) Cell differentiation during sexual development of the fungus *Sordaria macrospora* requires ATP citrate lyase activity. *Mol Cell Biol* 19:450–460. <https://doi.org/10.1128/MCB.19.1.450>

Nowrouzian M, Kück U, Loser K, Weltring K-M (2000) The fungal ac11 and ac12 genes encode two polypeptides with homology to the N- and C-terminal parts of the animal ATP citrate lyase polypeptide. *Curr Genet* 37:189–193. <https://doi.org/10.1007/s002940050518>

Osorio-González CS, Saini R, Hegde K, Brar SK, Avalos Ramirez A (2022a) Furfural degradation and its effect on *Rhodosporidium toruloides* 1588 during microbial growth and lipid accumulation. *Biore sour Technol* 359:127496. <https://doi.org/10.1016/j.biortech.2022.127496>

Osorio-González CS, Saini R, Hegde K, Brar SK, Lefebvre A, Avalos-Ramírez A (2022b) Inhibitor degradation by *Rhodosporidium toruloides* NRRL 1588 using undetoxified wood hydrolysate as a culture media. *Biomass Bioenergy* 160:106419. <https://doi.org/10.1016/j.biombioe.2022.106419>

Otoupal PB, Ito M, Arkin AP, Magnuson JK, Gladden JM, Skerker JM (2019) Multiplexed CRISPR-Cas9-based genome editing of *Rhodosporidium toruloides*. *mSphere* 4:e00099–19. <https://doi.org/10.1128/mSphere.00099-19>

Pant A, Brahim Belhaouari D, Dsouza L, Yang Z (2023) Upregulation of ATP citrate lyase phosphorylation and neutral lipid synthesis through viral growth factor signaling during vaccinia virus infection

Papanikolaou S, Aggelis G (2011) Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production. *Eur J Lipid Sci Technol* 113:1031–1051. <https://doi.org/10.1002/ejlt.201100014>

Pierce MW, Palmer JL, Keutmann HT, Avruch J (1981) ATP-citrate lyase. Structure of a tryptic peptide containing the phosphorylation site directed by glucagon and the cAMP-dependent protein kinase. *J Biol Chem* 256:8867–8870. [https://doi.org/10.1016/S0021-9258\(19\)52474-7](https://doi.org/10.1016/S0021-9258(19)52474-7)

Pietrocola F, Galluzzi L, Bravo-San Pedro JM, Madeo F, Kroemer G (2015) Acetyl coenzyme A: a central metabolite and second messenger. *Cell Metab* 21:805–821. <https://doi.org/10.1016/j.cmet.2015.05.014>

Pinheiro MJ, Bonturi N, Belouah I, Miranda EA, Lahtvee P-J (2020) Xylose metabolism and the effect of oxidative stress on lipid and carotenoid production in *Rhodotorula toruloides*: insights for future biorefinery. *Front Bioeng Biotechnol* 8:1008. <https://doi.org/10.3389/fbioe.2020.01008>

Probst KV, Schulte LR, Durrett TP, Rezac ME, Vadlani PV (2016) Oleaginous yeast: a value-added platform for renewable oils. *Crit Rev Biotechnol* 36:942–955. <https://doi.org/10.3109/0738551.2015.1064855>

Pronk JT, Yde Steensma H, Van Dijken JP (1996) Pyruvate metabolism in *Saccharomyces cerevisiae*. *Yeast* 12:1607–1633. [https://doi.org/10.1002/\(SICI\)1097-0061\(199612\)12:16%3c1607::AID-YEA70%3e3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0061(199612)12:16%3c1607::AID-YEA70%3e3.0.CO;2-4)

Qiao K, Wasylenko TM, Zhou K, Xu P, Stephanopoulos G (2017) Lipid production in *Yarrowia lipolytica* is maximized by engineering cytosolic redox metabolism. *Nat Biotechnol* 35:173–177. <https://doi.org/10.1038/nbt.3763>

Ratledge C (2004) Fatty acid biosynthesis in microorganisms being used for single cell oil production. *Biochimie* 86:807–815. <https://doi.org/10.1016/j.biochi.2004.09.017>

Ratledge C, Wynn JP (2002) The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. In: *Advances in Applied Microbiology*. Elsevier, pp 1–52

Rebnegger C, Graf AB, Valli M, Steiger MG, Gasser B, Maurer M, Mattanovich D (2014) In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. *Bio-technol J* 9:511–525. <https://doi.org/10.1002/biot.201300334>

Rekēna A, Pinheiro MJ, Bonturi N, Belouah I, Tammekivi E, Herodes K, Kerkhoven EJ, Lahtvee P-J (2023) Genome-scale metabolic modeling reveals metabolic trade-offs associated with lipid production in *Rhodotorula toruloides*. *PLOS Comput Biol* 19:e1011009. <https://doi.org/10.1371/journal.pcbi.1011009>

Rhee J, Solomon LA, DeKoter RP (2019) A role for ATP citrate lyase in cell cycle regulation during myeloid differentiation. *Blood Cells Mol Dis* 76:82–90. <https://doi.org/10.1016/j.bcmd.2019.02.006>

Saini R, Hegde K, Brar SK, Vezina P (2020) Advanced biofuel production and road to commercialization: an insight into bioconversion potential of *Rhodosporidium* sp. *Biomass Bioenergy* 132:105439. <https://doi.org/10.1016/j.biombioe.2019.105439>

Sato R, Ara S, Yamazaki H, Ishiya K, Aburatani S, Takaku H (2021) Citrate-mediated Acyl-CoA synthesis is required for the promotion of growth and triacylglycerol production in oleaginous yeast *Lipomyces starkeyi*. *Microorganisms* 9:1693. <https://doi.org/10.3390/microorganisms9081693>

Schultz JC, Cao M, Zhao H (2019) Development of a CRISPR/Cas9 system for high efficiency multiplexed gene deletion in *Rhodosporidium toruloides*. *Biotechnol Bioeng* 116:2103–2109. <https://doi.org/10.1002/bit.27001>

Schultz JC, Mishra S, Gaither E, Mejia A, Dinh H, Maranas C, Zhao H (2022) Metabolic engineering of *Rhodotorula toruloides* IFO0880 improves C16 and C18 fatty alcohol production from synthetic media. *Microb Cell Factories* 21:26. <https://doi.org/10.1186/s12934-022-01750-3>

Sreeharsha RV, Mohan SV (2020) Obscure yet promising oleaginous yeasts for fuel and chemical production. *Trends Biotechnol* 38:873–887. <https://doi.org/10.1016/j.tibtech.2020.02.004>

Stemmer M, Thümberger T, Del Sol KM, Wittbrodt J, Mateo JL (2015) CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE* 10:e0124633. <https://doi.org/10.1371/journal.pone.0124633>

Sukhija PS, Palmquist DL (1988) Rapid method for determination of total fatty acid content and composition of feedstuffs and feces. *J Agric Food Chem* 36:1202–1206. <https://doi.org/10.1021/jf00084a019>

Sunder S, Gupta A, Kataria R, Ruhal R (2024) Potential of *Rhodosporidium toruloides* for fatty acids production using lignocellulose biomass. *Appl Biochem Biotechnol* 196:2881–2900. <https://doi.org/10.1007/s12010-023-04681-w>

Takahashi H, McCaffery JM, Irizarry RA, Boeke JD (2006) Nucleocytosolic acetyl-coenzyme A synthetase is required for histone acetylation and global transcription. *Mol Cell* 23:207–217. <https://doi.org/10.1016/j.molcel.2006.05.040>

Tehlivets O, Scheuringer K, Kohlwein SD (2007) Fatty acid synthesis and elongation in yeast. *Biochim Biophys Acta BBA - Mol Cell Biol Lipids* 1771:255–270. <https://doi.org/10.1016/j.bbalip.2006.07.004>

Tingajeva O (2024) *Rhodotorula toruloides*' exopolysaccharides: production, optimization and characterization. Tallinn University of Technology, Thesis

Tiukova IA, Brandenburg J, Blomqvist J, Sampels S, Mikkelsen N, Skaugen M, Arntzen MØ, Nielsen J, Sandgren M, Kerkhoven EJ (2019a) Proteome analysis of xylose metabolism in *Rhodotorula toruloides* during lipid production. *Biotechnol Biofuels* 12:137. <https://doi.org/10.1186/s13068-019-1478-8>

Tiukova IA, Prigent S, Nielsen J, Sandgren M, Kerkhoven EJ (2019b) Genome-scale model of *Rhodotorula toruloides* metabolism. *Biotechnol Bioeng* 116:3396–3408. <https://doi.org/10.1002/bit.27162>

Tsai Y-Y, Ohashi T, Kanazawa T, Polburee P, Misaki R, Limtong S, Fujiyama K (2017) Development of a sufficient and effective procedure for transformation of an oleaginous yeast, *Rhodotorula toruloides* DMKU3-TK16. *Curr Genet* 63:359–371. <https://doi.org/10.1007/s00294-016-0629-8>

Verduyn C, Postma E, Scheffers WA, Van Dijken JP (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8:501–517. <https://doi.org/10.1002/yea.320080703>

Wang Y, Zhang S, Zhu Z, Shen H, Lin X, Jin X, Jiao X, Zhao ZK (2018) Systems analysis of phosphate-limitation-induced lipid accumulation by the oleaginous yeast *Rhodotorula toruloides*. *Biotechnol Biofuels* 11:148. <https://doi.org/10.1186/s13068-018-1134-8>

Wasylewko TM, Ahn WS, Stephanopoulos G (2015) The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in *Yarrowia lipolytica*. *Metab Eng* 30:27–39. <https://doi.org/10.1016/j.ymben.2015.02.007>

Wu S, Hu C, Jin G, Zhao X, Zhao ZK (2010) Phosphate-limitation mediated lipid production by *Rhodotorula toruloides*. *Bioreactor Technol* 101:6124–6129. <https://doi.org/10.1016/j.bioreacttech.2010.02.111>

Wu S, Zhao X, Shen H, Wang Q, Zhao ZK (2011) Microbial lipid production by *Rhodotorula toruloides* under sulfate-limited conditions. *Bioresour Technol* 102:1803–1807. <https://doi.org/10.1016/j.bioreacttech.2010.09.033>

Wu C-C, Honda K, Kazuhito F (2023) Current advances in alteration of fatty acid profile in *Rhodotorula toruloides*: a mini-review. *World J Microbiol Biotechnol* 39:234. <https://doi.org/10.1007/s11274-023-03595-3>

Wynn JP, Ratledge C (1997) Malic enzyme is a major source of NADPH for lipid accumulation by *Aspergillus Nidulans*. *Microbiology* 143:253–257. <https://doi.org/10.1099/00221287-143-1-253>

Wynn JP, Kendrick A, Ratledge C (1997) Sesamol as an inhibitor of growth and lipid metabolism in *Mucor circinelloides* via its action on malic enzyme. *Lipids* 32:605–610. <https://doi.org/10.1007/s11745-997-0077-1>

Wynn JP, Ratledge C, Hamid AA, Li Y (2001) Biochemical events leading to the diversion of carbon into storage lipids in the oleaginous fungi *Mucor circinelloides* and *Mortierella alpina*. *Microbiology* 147:2857–2864. <https://doi.org/10.1099/00221287-147-10-2857>

Xu P, Qiao K, Ahn WS, Stephanopoulos G (2016) Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proc Natl Acad Sci* 113:10848–10853. <https://doi.org/10.1073/pnas.1607295113>

Yang F, Zhang S, Zhou YJ, Zhu Z, Lin X, Zhao ZK (2012) Characterization of the mitochondrial NAD+-dependent isocitrate dehydrogenase of the oleaginous yeast *Rhodotorula toruloides*. *Appl Microbiol Biotechnol* 94:1095–1105. <https://doi.org/10.1007/s00253-011-3820-3>

Yang X, Sun W, Shen H, Zhang S, Jiao X, Zhao ZK (2018) Expression of phosphotransacetylase in *Rhodotorula toruloides* leading to improved cell growth and lipid production. *RSC Adv* 8:24673–24678. <https://doi.org/10.1039/C8RA03028F>

Yoon SH, Park JS, Rhee JS (1984) Production of NADPH for Lipogenesis in Oleaginous Yeast *Rhodotorula glutinis* 12:247–251

Yu Y, Shi S (2023) Development and perspective of *Rhodotorula toruloides* as an efficient cell factory. *J Agric Food Chem* 71:1802–1819. <https://doi.org/10.1021/acs.jafc.2c07361>

Zhang Y, Adams IP, Colin R (2007) Malic enzyme: the controlling activity for lipid production? Overexpression of malic enzyme in *Mucor circinelloides* leads to a 2.5-fold increase in lipid accumulation. *Microbiology* 153:2013–2025. <https://doi.org/10.1099/mic.0.2006/002683-0>

Zhang H, Zhang L, Chen H, Chen YQ, Ratledge C, Song Y, Chen W (2013) Regulatory properties of malic enzyme in the oleaginous yeast, *Yarrowia lipolytica*, and its non-involvement in lipid accumulation. *Biotechnol Lett* 35:2091–2098. <https://doi.org/10.1007/s10529-013-1302-7>

Zhang H, Zhang L, Chen H, Chen YQ, Chen W, Song Y, Ratledge C (2014) Enhanced lipid accumulation in the yeast *Yarrowia lipolytica* by over-expression of ATP: citrate lyase from *Mus musculus*. *J Biotechnol* 192:78–84. <https://doi.org/10.1016/j.biote.2014.10.004>

Zhang S, Ito M, Skerker JM, Arkin AP, Rao CV (2016a) Metabolic engineering of the oleaginous yeast *Rhodotorula toruloides* IF00880 for lipid overproduction during high-density fermentation. *Appl Microbiol Biotechnol* 100:9393–9405. <https://doi.org/10.1007/s00253-016-7815-y>

Zhang S, Skerker JM, Rutter CD, Maurer MJ, Arkin AP, Rao CV (2016b) Engineering *Rhodotorula toruloides* for increased lipid production. *Biotechnol Bioeng* 113:1056–1066. <https://doi.org/10.1002/bit.25864>

Zhang Y, Kamal R, Li Q, Yu X, Wang Q, Zhao ZK (2022) Comparative fatty acid compositional profiles of *Rhodotorula toruloides* haploid and diploid strains under various storage conditions. *Fermentation* 8:467. <https://doi.org/10.3390/fermentation8090467>

Zhao Y, Song B, Li J, Zhang J (2022) *Rhodotorula toruloides*: an ideal microbial cell factory to produce oleochemicals, carotenoids, and other products. *World J Microbiol Biotechnol* 38:13. <https://doi.org/10.1007/s11274-021-03201-4>

Zhu Z, Zhang S, Liu H, Shen H, Lin X, Yang F, Zhou YJ, Jin G, Ye M, Zou H, Zhao ZK (2012) A multi-omic map of the lipid-producing yeast *Rhodotorula toruloides*. *Nat Commun* 3:1112. <https://doi.org/10.1038/ncomms2112>

Zhu J, Gu Y, Yan Y, Ma J, Sun X, Xu P (2023) Knocking out central metabolism genes to identify new targets and alternating substrates to improve lipid synthesis in *Y. lipolytica*. *Front Bioeng Biotechnol* 11:1098116. <https://doi.org/10.3389/fbioe.2023.1098116>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Curriculum vitae

Personal data

Name Alīna Reķēna

Contact data

Website <https://linktr.ee/alinarekena>

Education

2021–2025 PhD – Science (Chemistry and Biotechnology), Tallinn University of Technology, Tallinn, Estonia

2019–2021 MSc – Bioengineering, University of Tartu, Tartu, Estonia

2015–2019 BSc Eng – Chemical Engineering, Riga Technical University, Riga, Latvia

Language competence

Latvian C2 Native

English C1 Fluent

Estonian A2 Elementary

Professional employment

2022–2025 Junior Research Fellow, Tallinn University of Technology, Tallinn, Estonia

2017–2019 Undergraduate Research Assistant, Riga Technical University, Riga, Latvia

Short study visits

July – Aug 2025 Yeast Genetics and Genomics Course, Cold Spring Harbor Laboratory, NY, USA

April 2024 RNA-Seq Data Analysis Course, BioSB NL, Breda, the Netherlands

September 2022 Metabolite Profiling Approaches in Nonconventional Yeast Training School, Royal Holloway University of London and Imperial College London, London, UK

September 2022 Agri-food Wastes and By-products Valorization: Functional Product Development Course, Estonian University of Life Sciences, Tartu, Estonia

Aug – Oct 2021 Short Term Scientific Mission, University of Groningen, Groningen, the Netherlands

Honors and awards

2023–2025 Two times IMES Early career scientist conference travel grant

2022–2025 Three times FEMS Early career scientist conference travel grant

2025 FEBS Early career scientist conference travel grant

2021 COST Yeast4Bio Action short term scientific mission mobility scholarship

2019 DORA Plus Estonian government scholarship

2016–2018 Latvian government excellence scholarship

Teaching experience and supervision

Spring 2025	Laboratory training for Biotechnology course at Tallinn University of Technology
Spring 2024,	
Spring 2025	Two times guest lecturer at System Biology course at Tallinn University of Technology
Autumn 2023,	
Autumn 2025	Two times guest lecturer at High-Throughput Biology and Biostatistics course at Tallinn University of Technology
Autumn 2023	Guest lecturer at Biorefinery systems course at Tallinn University of Technology
2023–2024	Supervision of Bachelor's thesis – "Phosphoketolase gene knockout by CRISPR/Cas9 in unconventional yeast <i>Rhodotorula toruloides</i> " – Kristjan Pals
2023–2024	Supervision of Bachelor's thesis – "Comparison of genome-scale metabolic models for investigating lipogenesis metabolism in <i>Rhodotorula toruloides</i> " – Maive Hanni

Conference presentations:

Oral presentations:

- I **A.Rekēna**, P.-J.Lahtvee. Chemostats and omics: tools to improve lipid-producing yeasts. Oral presentation at 6th Congress of Baltic Microbiologists (6CBM), October 1-3, Riga, Latvia, 2025.
- II **A.Rekēna**, S.Gavrilović, P.-J.Lahtvee. Metabolic regulation of acetyl-CoA and NADPH production pathways in *Rhodotorula toruloides*. Lightning talk at 16th Metabolic Engineering Conference (ME16), June 15-19, Copenhagen, Denmark, 2025.
- III **A.Rekēna**, M.Pinheiro, K.Pals, N.Bonturi, I.Belouah, S.Gavrilović, E.Kerkhoven, P.-J.Lahtvee. Quantitative systems biology analysis of lipid-producing yeast *Rhodotorula toruloides* cultivation. Oral presentation at FEBS3+ Conference "Biochemistry at a Fore Line", April 23-25, Vilnius, Lithuania, 2025.
- IV **A.Rekēna**. Designing microorganisms for the transition to bioeconomy. Oral presentation at XV Taltech School of Science Conference, November 29, Tallinn, Estonia, 2023.
- V **A.Rekēna**, M.Pinheiro, N.Bonturi, I.Belouah, E.Tammekivi, K.Herodes, E.J.Kerkhoven, P.-J.Lahtvee. Genome-scale modelling-based discovery of metabolic strategies in oleaginous yeast *Rhodotorula toruloides*. Oral presentation at 8th Conference on Physiology of Yeasts and Filamentous Fungi (PYFF8), June 5-8, Cork, Ireland, 2023.
- VI **A.Rekēna**, M.Pinheiro, N.Bonturi, I.Belouah, E.Tammekivi, K.Herodes, E.J.Kerkhoven, P.-J.Lahtvee. Metabolic network analysis of *Rhodotorula toruloides* using stoichiometric models and proteomics analysis. Oral presentation at FEBS3+ Conference of Estonian, Latvian and Lithuanian Biochemical Societies, June 15-17, Tallinn, Estonia, 2022.

VII **A.Rekēna**, M.Pinheiro, N.Bonturi, I.Belouah, E.Tammekivi, K.Herodes, E.Kerkhoven, P.-J.Lahtvee. Metabolic flux and proteome analysis of lipid production in *Rhodotorula toruloides*. Oral presentation at 15th Yeast Lipid Conference (YLC15), June 1-3, Gothenburg, Sweden, 2022.

Poster presentations:

VIII **A.Rekēna**, S.Gavrilovic, P.-J.Lahtvee. The design-build-test-learn cycle for investigating lipid synthesis on oleaginous *Rhodotorula toruloides*. Flash poster presentation at 7th Applied Synthetic Biology in Europe (ASBE7), November 6-8, Brno, Czech Republic, 2024.

IX **A.Rekēna**, E.N.Smith, M.Heinemann, P.-J.Lahtvee. Thermodynamics-based model-driven discovery of metabolic strategies in oleaginous yeast *Rhodotorula toruloides*. Poster presentation at 15th Metabolic Engineering Conference (ME15), June 11-15, Singapore, 2023.

X **A.Rekēna**, M.Pinheiro, N.Bonturi, I.Belouah, E.Tammekivi, K.Herodes, E.Kerkhoven, P.-J.Lahtvee. Enzyme-constrained genome-scale metabolic models of *Rhodotorula toruloides*. Poster presentation at 2nd VH online Yeast Conference (VHOYC), September 14, 2021.

Elulookirjeldus

Isikuandmed

Nimi Alīna Rekēna

Kontaktandmed

Kodulehekülg <https://linktr.ee/alinarekena>

Hariduskäik

2021–2025	PhD – Loodusteadused (Keemia ja biotehnoloogia), Tallinna Tehnikaülikool, Tallinn, Eesti
2019–2021	MSc – Biotehnika, Tartu Ülikool, Tartu, Eesti
2015–2019	BSc Eng – Keemiatehnika, Riia Tehnikaülikool, Riia, Läti

Keelteoskus

Läti keel	C2 emakeel
Ingliste keel	C1
Eesti keel	A2

Teenistuskäik

2022–2025	Doktorant-nooremteadur, Tallinna Tehnikaülikool, Tallinn, Eesti
2017–2019	Bakalaureuseõppje teadusassistent, Riia Tehnikaülikool, Riia, Läti

Lühiajalised õppetöövõtted

juuli – aug 2025	Pärmi geneetika ja genoomika kursus, Cold Spring Harbor Laboratory, New York, Ameerika Ühendriigid
aprill 2024	RNA-Seq andmete analüüsni kursus, BioSB NL, Breda, Holland
september 2022	Metaboliitide profileerimise meetodid mitte-konventsionaalsetes pärmidest – koolituskursus, Royal Holloway Londoni Ülikool ja Imperial College London, Suurbritannia
september 2022	Põllumajandusliku tooraine jätkide ja kõrvvalsaaduste väärindamine: funktsionaalsete toodete arendamine, Eesti Maaülikool, Tartu, Eesti
aug – okt 2021	Lühiajaline teadusmissioon, Groningeni Ülikool, Holland

Teaduspreemiad ja tunnustused

2023–2025	Kaks korda IMES noorteadlaste konverentsitoetust
2022–2025	Kolm korda FEMS noorteadlaste konverentsitoetust
2025	FEBS noorteadlaste konverentsitoetus
2021	COST Yeast4Bio võrgustiku lühiajalise teadusmissiooni mobiilsusstipendium
2019	DORA Plus Eesti valitsuse stipendium
2016–2018	Läti valitsuse tippstipendium (excellence scholarship)

Õpetamis- ja juhendamiskogemus

Kevad 2025	Biotehnoloogia kursuse laboratoorne koolitus Tallinna Tehnikaülikoolis
Kevad 2024,	
Kevad 2025	Kaks korda külalislektor süsteembioloogia kursusel Tallinna Tehnikaülikoolis

Sügis 2023,	
Sügis 2025	Kaks korda külalislektor kõrge läbilaskevõimiga bioloogia ja biostatistika kursusel Tallinna Tehnikaülikoolis
Sügis 2023	Külalislektor biorefinineerimissüsteemide kursusel Tallinna Tehnikaülikoolis
2023–2024	Bakalaureusetöö juhendamine – “Fosfoketolaasi geeni väljalülitamine CRISPR/Cas9 meetodil mitte-konventsionaalses pärmis <i>Rhodotorula toruloides</i> ” – Kristjan Pals
2023–2024	Bakalaureusetöö juhendamine – “Genoomimastaapsete metabolismimudelite võrdlus lipiidide sünteesi uurimiseks pärmis <i>Rhodotorula toruloides</i> ” – Maive Hanni

ISSN 2585-6901 (PDF)
ISBN 978-9916-80-454-4 (PDF)